INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY
CAMPUS ESTADO DE MÉXICO

PROTOCOLOS DE AUTENTIFICACIÓN SEGURA PARA COMUNICACIÓNES EN AMBIENTES PCS

TESIS QUE PRESENTA

DAVID HIGUERA ROSALES

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

MCCR 95, ITESM-CEM

JULIO, 2002

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY
CAMPUS ESTADO DE MÉXICO

PROTOCOLOS DE AUTENTIFICACIÓN SEGURA PARA COMUNICACIÓNES EN AMBIENTES PCS

TESIS QUE PRESENTA

DAVID HIGUERA ROSALES

PARA OBTENER EL GRADO DE MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

Asesor:	Dr. Gustavo Santana Torrellas	
Comité de Tesis:	Dr. José de Jesús Vázquez Gómez Dr. Luis A. Trejo Rodríguez	
Jurado:	Dr. José de Jesús Vázquez Gómez	Presidente
	Dr. Luis A. Trejo Rodríguez	Secretario
	Dr. Gustavo Santana Torrellas	Vocal

Atizapán de Zaragoza, Edo. de México, Julio de 2002

PRÓLOGO

Hoy en día la seguridad es uno de los más importantes requerimientos para la extensa aceptación de los sistemas de comunicaciones personales y la autentificación resulta ser un procedimiento esencial para asegurar que el servicio esta siendo utilizado apropiadamente.

En este trabajo de tesis, se examinan varios protocolos relativos al ambiente móvil, dando mayor énfasis a los más representativos y bien conocidos para los servicios de comunicación móvil, a partir de varios aspectos, incluyendo técnicas empleadas, arquitectura de sistemas, colocación de la confianza, eficiencia, vulnerabilidades de seguridad, aplicabilidad de la implementación, etc.

Con el desarrollo de las técnicas de software y hardware, los usos pensados para datos, voz, imagen, o sus combinaciones pueden todos ser incorporados dentro de los ambientes inalámbricos actualmente.

Debido a que surgen nuevos servicios, los requerimientos para la seguridad serán diferentes dependiendo de los usos o aplicación. La compatibilidad con servicios e interoperabilidad existentes entre diversos proveedores de servicio debe también ser considerada. Todo esto, complica el diseño de un protocolo apropiado de autentificación, lo cual es un motivo fuerte para introducirnos en este mundo de los protocolos de autentificación seguros para ambientes PCS.

Organización del documento de tesis

En resumen, este trabajo de tesis está estructurado en cuatro capítulos y tres anexos.

En el capítulo 1, se da una introducción al tema, se plantean las inquietudes del problema de la seguridad de los ambientes móviles y se describe de forma cualitativa el estado actual de ambiente en cuestión.

En el capítulo 2, se dan los fundamentos y un marco de referencia conceptual y metodológico efectivo de los protocolos de autentificación existentes en las comunicaciones inalámbricas actuales. Por lo que se analizan de manera detallada, las distintas propuestas y estándares más representativos en los últimos años, así como los distintos ambientes de oportunidad de aplicación y forma en que se relacionan con el trabajo de esta tesis.

El capítulo 3, contiene los protocolos que hemos propuesto en el desarrollo de esta tesis, y que han sido publicados en conferencias y congresos internacionales [ver 1, 2, 3, 4 del capitulo 3].

En él capitulo 4, se da una descripción general, de algunos estándares que actualmente están funcionando para proporcionar servicios para los PCS. Y se plantean propuestas consideradas como áreas de oportunidad para la aplicación de nuestros protocolos de autentificación propuestos en este trabajo de tesis.

Finalmente, en el Anexo A, se encuentra una descripción de la lógica de BAN así como otra lógica simplificada, las cuales fueron utilizadas para revisar la seguridad de los protocolos de interés de esta tesis. El anexo B, contiene una descripción general de las curvas elipticas y el anexo C contempla varios aspectos que complementan el estudio de los sistemas de comunicaciones inalámbricas en términos generales.

Cabe señalar, que cada capítulo incluye por separado la bibliografia y referencias utilizadas. Así mismo, se proporciona una lista de figuras y tablas también por capítulos. Y las abreviaturas se dan en forma general, y muchas de estas ultimas son definidas dentro del texto de cada capítulo

CONTENIDO

CAPITULO 4

LISTA DE FIGURAS

CAPITULO 1

CAPITULO 2

CAPITULO 3

CAPITULO 4

LISTA DE TABLAS

CAPITULO 2

CAPITULO 3

CAPITULO 4

Glosario de Términos

ACL: Access Control List (Lista de Control de Acceso LCA)
AMPS: Advanced Mobile Phone System o norma TIA-533.
API: Application Programming Interface (Interfaz de Programación de Aplicación)
ATM: Asynchronous Transfer Mode
CA: Certificate Authority (Autoridad Certificadora AC)
CDMA: Code Division Multiple Access (Acceso Múltiple por División en el Código)
CSD: Circuit Switched Data (Conmutación de Circuitos de Datos)
(Sistema Global para Comunicaciones Móviles)
DECT: Digital Enhanced Cordless Telephony
DSSS; Direct Sequence Spread Spectrum
DTMF: Dual tone MultiFrequency
eCommerce Electronic Comerce: (Comercio Electrónico- Transacciones comerciales llevadas a cabo en redes de información)
eStore Electronic Store: (Tienda electrónica- Un sitio de web donde uno puede realizar transacciones de comercio electrónico)
FDMA: Frequency Division Multiple Access
FM: Frequency Modulation
FSK: Frequency Shift Key
GSM: Global System for Mobile Communication
HFSS: Frecuency Hopping Spread Spectrum
IP: Internet Protocol (Protocolo de Internet)
IR: Infrared
ISDN: Integrated Services Digital Network
ITU: International Telecommunications Union
LAN: Local Access Network
LMCS: Local Multipoint Communications System
LMDS: Local Multipoint Distribution Service
MAC: Medium Access Control (Control de Acceso al Medio)
MC-DSSS: MultiCode Direct Sequence Spread Spectrum
MDG: Mobile Data Gateway (Pasarela de Datos Móviles)
ME: Mobile Entity, Mobile Equipment (Entidad Móvil, Equipo Móvil)
MeT: Mobile Electronic Transactions Initiative (Iniciativa de transacciones Electrónicas Móviles)
MSISDN number: Mobile Subscriber Integrated Services Digital Network Number (Número de Red Digital de Servicios Integrados de Subscriptor Móvil)
NIU: Network Interface Unit
NMS: Network Management System
PC : Personal Computer
PCMCIA: Personal Computer Memory Card International Association
PDA: Personal Digital Assistant
PIN: Personal Identification Number (Número de Identificación Personal)
PKI: Public Key Infrastructure (Infraestructura de Llave Pública)
PMP: Point to MultiPoint
POTS: Plain Old Telephone System
PPP: Point-to-Point Protocol (Protocolo Punto-a-Punto)

PRA: Primary Rate Access
PRI: Primary Rate ISDN
QoS: Quality of Service
RFC: Request For Comments
RSC: Radio Station Central
RST: Radio Station Terminal
RSCW: Radio Station Central Wireless
SAR: Segmentation and Reassembly (Segmentación y Re-ensamblado)
SIM: Subscriber Identity Module (Módulo de Identidad de Subscriptor)
SPKI: Simple Public Key Infrastructure (Infraestructura de Llave Pública Simple)
SSL: Secure Sockets Layer (Capa de Sockets Segura)
TCP/IP: Transmission Control Protocol / Internet Protocol
TDM: Time Division Multiplexing
TIA: Telecommunications Industry Association
URI: Universal/Uniform Resource Identifier (Identificador Universal/Uniforme de Recursos)
WAE: Wireless Application Environment (Entorno Inalámbrico de Aplicación)
WAP: Wireless Application Protocol (Protocolo de Aplicación Inalámbrica)
WBS: Wireless Base Station
WDP: Wireless Datagram Protocol (Protocolo Inalámbrico de Datagramas)
WIM: Wireless Identity Module (Módulo de Identificación Inalámbrico)
WLAN: Wireless LAN
WLL: Wireless Local Loop
WNT: Wireless Network Termination
WSP: Wireless Session Protocol (Protocolo Inalámbrico de Sesión)
WTLS: Wireless Transport Layer Security (Capa de Seguridad de Transporte Inalámbrico)
WTP: Wireless Transaction Protocol (Protocolo Inalámbrico de Transacciones)
W-OFDM: Wideband Orthogonal Frequency Division Multiplexing
XBS: Exchange Base Station

CAPÍTULO 1

1. INTRODUCCIÓN

El desarrollo de los dispositivos portátiles y de computación móvil con construcción de interfaces de radio de alta velocidad tiene un gran impacto en la industria de las comunicaciones. Un número grande de usuarios móviles equipados con comunicadores que habilitan el Internet inalámbrico requieren acceso a la web basándose en servicios en cualquier lugar y en cualquier momento. La habilidad de ubicuidad de accesos a Internet inalámbrico quizás podrá superar la popularidad de la telefonía celular y cambiar la forma en que nos comunicamos. Este ambiente da lugar a demandas significativas de soluciones de movilidad de generaciones próximas y las ya existentes.

Los años recientes han sufrido un desarrollo rápido de la tecnología de comunicaciones móviles. El principio del celular (ver anexo C) nos permite un uso eficiente de los recursos de radio limitados y ayuda a soportar poblaciones de subscriptores grandes. Los avances en microelectrónica, por otro lado, han hecho de los teléfonos celulares y computo portátil una comodidad. El número creciente de usuarios de teléfono celular sugiere que la movilidad pronto sé este convirtiendo en la norma en comunicaciones, en vez de la excepción. Mientras que el estado del arte de los sistemas móviles celulares todavía se optimizan para la comunicación de voz, estos soportan una incrementada variedad de servicios de datos. Las iniciativas recientes para aumentar el Internet con el soporte de la movilidad indican el interés de aumento en servicios de datos móviles.

Las tecnologías futuras para el soporte de acceso a Internet inalámbrico deberán estar influenciadas tanto de tecnología de Internet como sistemas de telefonía celular. Soluciones escalables y flexibles son requeridas para que se puedan adaptar a amplios rangos de ambientes. A los usuarios se les debe de poder ofrecer movilidad sin problemas a través de los sistemas heterogéneos que necesitan colaborar recíprocamente y cooperar para proporcionar el mejor servicio disponible.

Con la introducción de las redes de comunicaciones móviles, surgen nuevas necesidades de seguridad propias de estos ambientes, debido a la falta de mecanismos de protección, semejantes a los tradicionales en las topologías fijas. Por ello, la necesidad de contar con mecanismos de disponibilidad, confidencialidad e integridad eficientes y acordes a dicho ambiente. Lo cual sugiere nuevos problemas y violaciones relativos a la seguridad de los sistemas de comunicaciones móviles, tales como: La prevención de los accesos ilegales en las funciones de seguridad, las intercepciones no autorizadas de datos del usuario, lo cual puede provocar la perdida de la confidencialidad de la identidad del usuario[1], así como, los distintos proveedores de servicio y la carencia de interoperabilidad entre estos.

En principio, algunas de las características de seguridad se pueden alcanzar con los protocolos de autentificación.

Por lo que todo lo anterior, además de otras amenazas y violaciones, determina la necesidad de contar con mecanismos de autentificación en situaciones de movilidad y de diversidad de los dominios. Ofreciendo con ello, mediante estos mecanismos una garantía en las diferentes transacciones que realizan.

El enmascaramiento y escuchar sin autorización son algunas de las amenazas importantes para la seguridad de las comunicaciones inalámbricas. Para proporcionar una protección apropiada para la comunicación inalámbrica, el contenido de la comunicación debe ser codificado y la autentificación mutua debe existir entre el suscriptor y la red.

La autentificación de usuario móvil puede ser implementada mediante el empleo de diferentes esquemas de seguridad, mismos que utilizan procedimientos criptológicos, tales como:

- Encripción de llave secreta
- Encripción de llave pública
- Firma digital.

Cuando se emplean mecanismos de llave secreta, la autentificación es implementada mediante algoritmos simétricos, obteniéndose como resultado alta velocidad de autentificación. Por otra parte, cuando son utilizados mecanismos de llave pública, la autentificación es implementada mediante algoritmos asimétricos, los cuales demandan una mayor cantidad de cálculos. En consecuencia, la velocidad de autentificación es lenta, pero se puede obtener una alta seguridad en comparación con los algoritmos simétricos. La firma digital se puede implementar utilizando cualesquiera de los mecanismos antes señalados, siendo necesario, en ambos casos, realizar cuando menos tres pasos; generación, firma y verificación.

[1] En este caso, el uso ilegitimo de los servicios de red pueden resultar en acusaciones falsas para usuarios legítimos

En este trabajo de tesis se tratan los problemas del análisis, diseño, desarrollo, implementación y verificación de protocolos de autentificación en sistemas de comunicaciones móviles inalámbricos.

Estos protocolos de autentificación forman parte de un conjunto de diseños alternativos, orientados a determinar altos niveles de confiabilidad en condiciones de movilidad inter-dominios de los usuarios de Sistemas o Redes de Comunicaciones Personales (PCS / PCN)[2]

Los protocolos propuestos en esta tesis fueron diseñados como una extensión de los semejantes de Beller-Chang-Yacobi. [1,2]

En la primer propuesta [3] presentamos un protocolo de autentificación de usuario móvil en un ambiente que considera "inter-dominios" con algunas ventajas para la autentificación de los usuarios empleando una modificación del esquema de firma digital con llave pública, que emplea certificados y curvas elípticas, como una alternativa conveniente para evaluar la seguridad En una segunda propuesta [4] se presenta una variación del protocolo mencionado anteriormente, en la cual se considera la situación de autoridades de certificación distribuidas. Mientras que en un tercer planteamiento [5] describimos un protocolo de usuario móvil empleando autentificación mutua y criptografía de llave pública distribuida. Este ultimo protocolo se piensa para los usuarios con alto grado de movilidad; donde el procedimiento de la autentificación, no emplearía los certificados debido a la movilidad experimentada por el usuario. Además, se propone incorporar el primer protocolo en una propuesta de aplicación para un protocolo de autentificación de usuario móvil para ATM Inalámbrico[6]

1.1 ESTADO ACTUAL DE LOS ESTÁNDARES DE COMUNICACIÓN PERSONAL

El primer sistema de telefonía celular analógico fue desarrollado por los laboratorios Bell, posteriormente estandarizado por la TIA (Telecommunications Industry Association) y conocido como AMPS (Advanced Mobile Phone System) o norma TIA-533. Este sistema utiliza la modulación FM (Frequency Modulation) para la transmisión de voz, y la modulación FSK (Frequency Shift Key) para la señalización. La tecnología utilizada para compartir un mismo espectro es llamada FDMA (Frequency Division Multiple Access), y es la que utiliza el sistema AMPS.

Siendo el sistema AMPS diseñado para el tráfico de señales analógicas de audio, este no resulta apropiado para la transmisión de datos. Para adecuarlo, Debe ser utilizado un módem con las características requeridas por la red telefónica celular.

El Propio proceso de "hand off", ocasiona una interrupción momentánea de la señal. Estos intervalos de interrupción pueden ser del orden de 100 milisegundos, generando errores en la transmisión. Un aspecto relevante de estos sistemas analógicos de FM de banda ancha, es que no tienen espacio para mas abonados, por lo que son reemplazados por sistemas de FM de banda angosta (NAMPS.).

---

[2] En ingles definido como; Personal Communication Systems y Personal Communication Networks

A pesar de todo esto, el sistema celular utilizado en Norte América ha sido el AMPS, y para poder introducirlo, la FCC asignó el ancho de banda (806-890 Mhz), antes utilizado para los canales de televisión 70-83. Además permitió la competencia entre dos compañías dentro de un área, donde una compañía era alámbrica y otra es inalámbrica. De esta manera cada usuario tiene la opción de poder subscribirse a ambas si así lo desea.

Sistema D-AMPS, definido también como norma IS-54, fue el primer estándar de telefonía celular digital americano. El sistema D-AMPS, opera en el mismo espectro utilizado por los sistemas AMPS, conservando los 30 Khz. de ancho de banda de canal. Un factor esencial en la transición de los sistemas AMPS hacia la tecnología de acceso TDMA (Time Division Multiple Access), es el aprovechamiento de la estructura ya instalada y ampliamente utilizada, por eso el D-AMPS define el concepto "dual mode", significando que provee una operación para ambos sistemas analógico y digital.

El sistema D-AMPS, utiliza multiplexación FDM para canales de RF de la frecuencia de telefonía celular igual al AMPS, y multiplexación TDM dentro de cada canal RF. La multiplexación TDM implica técnicas de transmisión digital en los canales RF, lo que equivale a multiplicar la capacidad del sistema por un factor de tres con relación al sistema AMPS.

En el anexo C, se puede encontrar mas información respecto a las actuales técnicas de acceso y estándares, así como el uso del espectro. Mientras que en la figura 1.3, se puede ver un panorama general del universo inalámbrico.

1.1.1 GSM (GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS)

Fue desarrollado en Europa por ETSI (Instituto de Estándares de Telecomunicaciones Europeo) para proveer un único estándar que ofreciera una mayor capacidad de usuarios, mejor calidad, facilidades para la transmisión de voz y datos, conexión a sistemas de correo electrónico, envío de mensajes cortos, así como, el servicio de "roaming" continental para cubrir todo el territorio Europeo. Soporta igualmente prestaciones adicionales, como: Servicios y cualidades especificas de los dispositivos de comunicación.

En el tema de la seguridad ofrece novedades importantes, como el uso de tarjeta de usuario para autentificación de llamadas (SIM), y el encriptado, que facilita confidencialidad e imposibilidad de utilización de terminales robadas, mediante la asignación previa de un número de serie a cada estación móvil.

GSM, esta basado en la tecnología TDMA. Dos bandas de frecuencia son definidas para GSM, una de 890 a 915 Mhz. Para transmisión de la unidad móvil, y otra de 935 a 960 Mhz. Para la transmisión de la estación base (ver anexo C y E, para mayor información.)

El sistema GSM esta compuesto del Subsistema de estación Base (BSS), el Subsistema de Conmutación y Red (NSS) y el Subsistema de mantenimiento y Operación (OSS) [7]. La arquitectura de red es presentada en la figura 1.1 El BSS consiste de la Estación Base de Transmisión-recepción (BTS) y los Controladores de Estación Base (BSC) y está a cargo de las trayectorias de transmisión de abastecimiento y el manejo entre las estaciones móviles (MS) y los Centros de Conmutación Móvil (CCM o MSC) los cuáles son los bloques de edificio primarios

del NSS. El NSS incluye funciones de manejo de localidad y conmutación. En particular, abarca las funciones del registro de localización Local (HLR) y registro de localización del visitante (VLR), las cuales representan la base de datos de manejo de localidad de GSM. El NSS también es responsable de interfazar redes externas tales como la red de telefonía pública. Finalmente el OSS provee de funciones para interacciones de manejo de red a todas las entidades mencionadas anteriormente.

Aunque GSM fue diseñado principalmente para soportar servicios de voz, puede ofrecer varios servicios de datos. Los usuarios de GSM pueden enviar y recibir mensajes alfanuméricos cortos empleando el servicio de mensajes cortos (SMS.) Recientemente, la importancia de proporcionar servicios de paquetes de datos para usuarios de celular móvil ha crecido debido al creciente rol de las redes basadas en IP. Como respuesta a esta demanda, GSM evoluciona a GPRS (General Packet Radio Service) [8]. Similar a la extensión de CDPD para AMPS, utilizada en Estados Unidos, GPRS reutiliza las interfases de radio existentes para la transmisión de paquetes de datos.

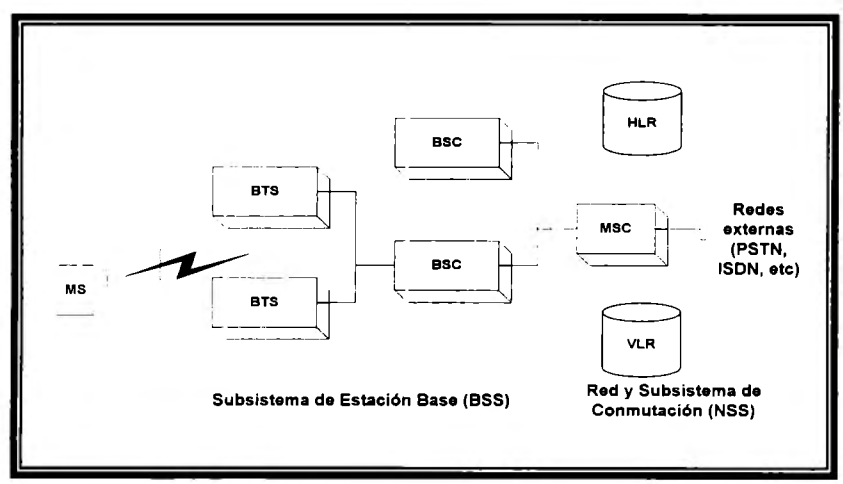


Fig. 1.1 Arquitectura general de GSM

1.1.2 SISTEMAS DE TERCERA GENERACIÓN

La tercera generación de sistemas celulares provee de una variedad de servicios para usuarios móviles en cualquier lugar y en cualquier momento. El concepto referido como Telecomunicaciones Móviles Internacionales 2000 (IMT-2000) incluye acceso de alta calidad para Internet y para ISDN-B. La estandarización de estos nuevos sistemas es realizada principalmente por la ITU-T SG11 en un nivel internacional y por el ETSI en Europa [9]. Para poder ganar amplia aceptación, la iniciativa europea para IMT-2000, llamada Sistema de Telecomunicaciones Móviles Universales (UMTS) incluye la evolución suave a partir de sistemas móviles de segunda generación, particularmente el sistema GSM. En la figura 1.2 se ilustra la arquitectura de UMTS. [10]

GPRS es un nuevo conjunto de servicios desarrollado por el ETSI, los cuales son añadidos a los actuales que posee GSM, básicamente añade conmutación de paquetes de datos a todos los niveles de la red GSM [11]. GPRS ofrece funciones de autentificación, control de accesos, confidencialidad de la identidad del usuario y confidencialidad de la información [12]. Los

algoritmos empleados en el proceso de Autentificación son los mismos que los de GSM (A3 y A8), mientras que el algoritmo utilizado para el cifrado de los datos de usuario ha sido modificado debido a la naturaleza del tráfico de GPRS, dicho algoritmo denominado GPRS A5 fue definido en SAGE (Security Advisor Group of Experts) del ETSI y no se encuentra disponible de forma pública. A GPRS se le denomina como la generación 2.5, ya que es el paso intermedio a los nuevos sistemas de 3ª generación. UMTS es el sistema de telefonía móvil de 3ª generación el cual se dice que será capaz de alcanzar velocidades entre 384 kbps para entornos de redes de banda ancha y 2.0 Mbps para entornos locales. Respecto a los mecanismos de seguridad del sistema UMTS estos se encuentran en la fase de desarrollo, han sido propuestos diferentes mecanismos para proporcionar autenticación, confidencialidad y generación de claves [10].

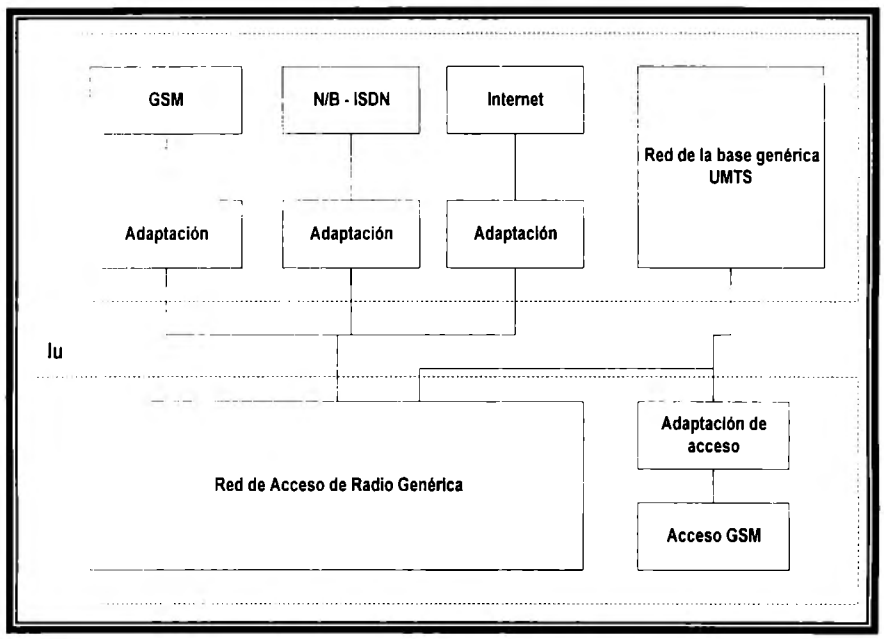


Fig. 1.2 Arquitectura de UMTS

1.1.3 PROPUESTAS DE LA MOVILIDAD DEL INTERNET

Independientemente de las iniciativas que surgieron para agregar servicios de datos a las redes de comunicación móvil celular, recientemente ha aparecido una gran cantidad de propuestas para aumentar el uso de Internet con soporte de movilidad de host[3].

Una dificultad básica que presentan los protocolos en el soporte de tal movilidad de host, es que debe hacer frente, a que la dirección de host en el Protocolo de Internet (IP) tiene significado dual. El primero, como un único identificador de host debe ser mantenido constante sin importar movilidad. El segundo, en su papel como indicador de la localización debe cambiar mientras que los hosts cambian la localidad. Éstos son los requisitos competentes que los protocolos móviles

[3] Palabra que nos identifica a las entidades direccionables en el nivel de red. Un host es distinguido por el hardware de soporte fundamentalmente.

del host deben resolver eficientemente. Un problema fundamental a resolver es por lo tanto la separación de esos dos roles mientras que el mapeo actualizado de los identificadores del host para información de la localización se hace disponible. Se ha mostrado en [13] que la mayoría de las soluciones propuestas pueden ser vistas como casos especiales de una arquitectura llamada "two tires addressing" donde un host móvil es lógicamente asociado con dos direcciones IP, eso es, su dirección local que sirve como un identificador de host que no cambia y una dirección que refleja su punto de unión al Internet. Estas arquitecturas generales comprenden tres componentes fundamentales:

- Un directorio de la localización, que representa una base de datos que contiene la mayoría de los mapas de actualización entre los dos espacios de dirección.
- La traducción del identificador de host a la dirección destino actual en cada paquete es presentada por Agentes de Traducción de Dirección.
- El componente final de la arquitectura generalizada es el agente de reenvío que realiza la traducción inversa para asegurarse de que los paquetes que llegan al host móvil tengan su dirección local constante en el campo de destino.

Una descripción de las propuestas de movilidad de Internet fundamentadas en este concepto general se proporciona en [13].

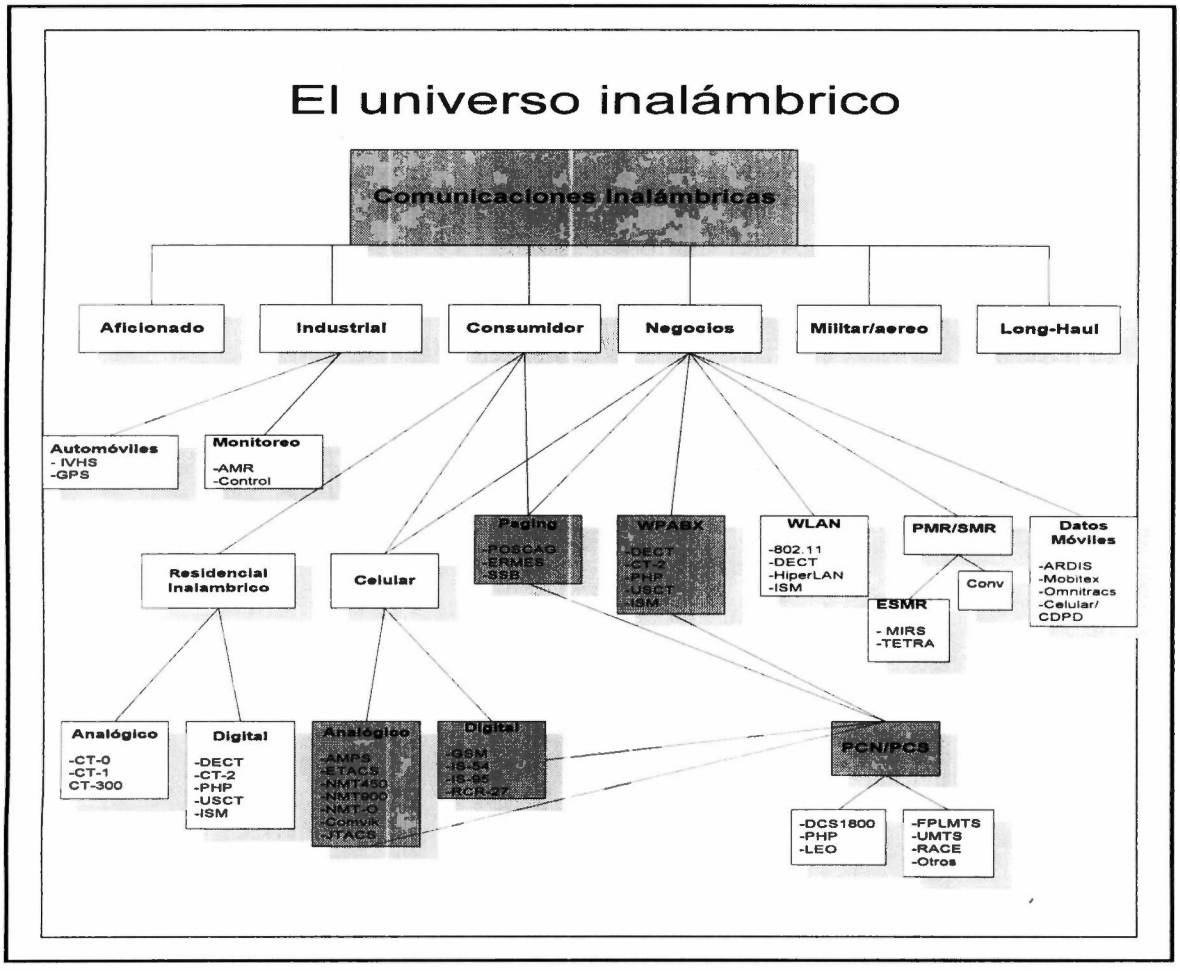


Fig. 1.3 El universo inalámbrico

1.2 SISTEMAS DE COMUNICACIONES PERSONALES (PCS) y REDES DE COMUNICACIONES PERSONALES (PCN)

El acelerado crecimiento tenido en los últimos años en los sistemas de comunicaciones inalámbricas se debe principalmente a las ventajas ofrecidas respecto a la movilidad de los usuarios. Lo cual a dado lugar a la generación de nuevos conceptos; tales como: PCS y PCN[4], con los cuales se pretende englobar todo lo relacionado con los aspectos de comunicación móvil de los años venideros. En la figura 1.3, se puede ver un panorama general del universo de las comunicaciones inalámbricas, así como las ligas directas e indirectas a los sistemas PCS y PCN.

1.2.1 DEFINICIÓN

De acuerdo con la "US Federal Communications Commisions (FCC)" PCS se ha definido como:

" El sistema por el cual todo usuario puede intercambiar información con cualquier otro, en cualquier momento, en cualquier lugar, a través de cualquier dispositivo, utilizando un sencillo Numero Personal de Telecomunicación NPT o TPN" [5]

1.2.2 OBJETIVOS DE LAS PCS Y PCN

Ofrecer servicios de multimedia de alta calidad, ambientes múltiples, múltiples tipos de usuarios, capacidad de "roaming" global, Número Personal de Telecomunicación, alta capacidad, dispositivos de comunicación universal y servicios de seguridad, para poder englobar el concepto como sistemas que sean multi-servicios, multi-ambiente y multi-operador.

Observándolo desde el punto de vista de los usuarios, se puede decir que los PCS y PCN son sistemas de servicios que integran múltiples redes de servicio tales como: PSTN(Public Switched Telephone Network) , ISDN (Integrated Services Digital Network), CS (Cordless System), Sistema Satelital, Sistema Móvil Terrestre y PBX inalámbrico (Private Branch Exchange) las cuales son operadas por diversos proveedores. Pero, para el usuario debe ser transparente la transferencia entre distintas redes.

Una revisión mas detallada de la tecnología de PCS se puede encontrar en el anexo C.

---

[4] La definición de PCS y PCN ha causado polémica entre los Europeos y los Estados Unidos de Norte América, siendo definido PCS en Estados Unidos como los servicios que brinda PCN, es decir; Servicios de Comunicaciones Personales. Mientras que para los Europeos PCN es una red de PCS, esto es, PCN se refiere a todos los sistemas y medios de comunicaciones personales.

[5] Esta definición fue tomada literalmente de lo que dice la FCC " The system by which every user can exchange information with anyone at anytime, in any place, through any type of device, using a single personal telecommunications number PTN"

1.2.3 CARACTERÍSTICAS SOBRESALIENTES

Servicios de multimedia de alta calidad. Se presume ofrecer servicios de voz y datos de alta calidad, video de alta definición, más aún, la contraparte de los recursos disponibles en ISDN también podrán estar disponibles en ambientes móviles con la misma definición y calidad.

Ambientes y tipos de usuarios múltiples. Se dice que deben proporcionar servicios a usuarios con diferentes necesidades de calidad, además de no importar el ambiente al cual pertenezcan, es decir, domestico, gubernamental, militar, educativo, comercial o financiero.

Capacidad de "roaming" global. Se debe contar con esta capacidad, es decir, un usuario debe poder pasar entre distintos puntos geográficos y entre distintas redes sin cambiar de sistema. Más aún, este sistema debe ser global o mundial. Y la movilidad a distintas velocidades debe también ser un punto importante dentro de las características, sin afectarse la calidad del servicio.

Número Personal de Telecomunicación (PTN) Esta es la base de la movilidad del sistema. Los usuarios deben de poder tener acceso al sistema con un único número de identificación personal sin que importe el tipo de servicio requerido o la ubicación.

Alta capacidad. En este mercado será necesario un sistema de alta capacidad, ya que la demanda para potencial se estima para una conexión por usuario.
Dispositivos de comunicación universal. Se debe emplear un dispositivo de fácil empleo, compacto, y que pueda dar disponibilidad a todos los servicios del sistema, a pesar de sus dificultades de diseño.

Servicios de seguridad. ***Por las características de este sistema, se requiere del empleo de tecnologías y técnicas de protección, autentificación, mas avanzadas y efectivas que las ya existentes en las redes fijas.***

Siendo este ultimo nuestro punto central de interés en este trabajo de tesis.

1.3. CONCLUSIÓN DEL CAPITULO

Este capitulo básicamente nos un panorama muy general de los sistemas de comunicación móvil (MCS), de los sistemas de comunicación personal, así como de las redes de comunicación personal. Ambientes propios de nuestro interés para poder proporcionar propuestas en el ámbito de la seguridad necesaria para estos sistemas, específicamente los protocolos de autentificación.

Los sistemas de comunicaciones móviles digitales actuales son generalmente expresados a partir de la segunda generación en distinción a los sistemas analógicos de la primera generación. Las investigaciones actuales principalmente se están enfocando a los sistemas emergentes de tercera generación, los cuales son caracterizados por un ancho de banda amplio y servicios de datos integrados. Ambiente propicio para los PCS y PCN.

En los sistemas de segunda generación la seguridad ha sido aplicada solamente al enlace de radio. Las redes móviles de la tercera generación pueden requerir alta seguridad en el enlace de radio

pero además seguridad "end-to-end" entre los usuarios móviles y sus contrapartes de comunicación podrían ser deseables. A continuación se puede resumir justamente los requisitos para un protocolo de seguridad que protege el enlace de radio.

- Confidencialidad del enlace de radio entre la estación base y el móvil
- Autentificación mutua entre la estación base y el móvil
- Confidencialidad de la identidad de la estación móvil
- Simplicidad computacional o baja complejidad computacional del protocolo con respecto a los requerimientos en la estación móvil

Los servicios de seguridad end-to-end requeridos corresponderán a qué se requiere para asegurar la aplicación particular. Éstos incluirán típicamente confidencialidad e integridad de los datos del usuario, y muchos también incluyen la no-repudiación para aplicaciones tales como comercio electrónico. En esta tesis básicamente nos dedicamos a generar una serie de propuestas de protocolos de autentificación segura en el camino de poder alcanzar enlaces end-to-end. A través del enlace de inalámbrico.

REFERENCIAS DEL CAPÍTULO

[1] M. J.Beller, L. F. Chang, and Y. Yacobi, "Privacy and Authentication on a portable Communications System," IEEE J. on Selected Areas in Communications, Vol. 11, No. 6, pp. 821-829, August 1993.

[2] M. J.Beller, L. F. Chang, and Y. Yacobi, " Security for Personal Communications Services: Public-Key vs. Private-Key Approaches," in Proceedings of third IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'92), pp. 26-31, IEEE Press, 1992.

[3] Gustavo A. Santana. T, David Higuera R, "A Mobile User Authentication Protocol for Personal Communication Networks", paper accepted for presentation at the IASTED (International Conference on Wireless and Optical Communications (WOC 2001)). June 27 to June 29, 2001, in Banff, Canada.

[4] Gustavo A. Santana. T, David Higuera R., "Protocolo de Autentificación de Usuario Móvil empleando Criptografía de Llave Pública Distribuida" X Congreso Internacional de Computo CIC 2001, 12-16 Noviembre de 2001, Ciudad de México, D.F.

[5] Gustavo A. Santana. T, David Higuera R, "A Mobile User Authentication Protocol Using Mutual Authentication for Personal Communication Networks", paper accepted for presentation at the IASTED (International Conference on Intelligent Systems and Control (ISC 2001)). November 19-22, 2001 Tampa, Florida, USA.

[6] Gustavo A. Santana. T, Arturo Torres D., David Higuera R., " Protocolo de Autentificación de Usuario Móvil para ATM Inalámbrico" X Congreso Internacional de Computo CIC 2001, 12-16 Noviembre de 2001, Ciudad de México, D.F.

[7] M. Rahnema, Overview of the GSM System and Protocol Architecture, IEEE Communications Magazine, April 1993.

[8] PAÚL, D. *Introducción de GPRS en redes GSM*. 19 de Abril de 1999. URL: http://www.telecomid.com/litxers/1/367.pdf.

[9] ETSI, ETS 300 175-7, October 1992.

[10] European Telecommunications Standards Institute, Universal Personal Telecommunications, ETSI NA7 WP1, November 1992.

[11] *La conmutación de paquetes llega a GSM GPRS*. N° 86. 1 de Enero de 1995. URL: http://www.idg.es/comunicaciones/mainart.asp?artid=10870

[12] KARI, H. H. *GPRS security issues*. 12 de Febrero de 1999. URL: http://www.cs.hut.fi/~hhk/GPRS/ps/

CAPÍTULO 2

2. LOS PROTOCOLOS DE AUTENTIFICACIÓN EN LAS PCNs Y PCSs.

El trabajo realizado en este capítulo está fundamentado en la preocupación de contar con un marco de referencia conceptual y metodológico efectivo de los protocolos de autentificación existentes en las comunicaciones inalámbricas.

Motivo por el cual analizamos de manera detallada, las distintas propuestas y estándares más representativos en los últimos años, así como los distintos ambientes de oportunidad de aplicación y forma en que se relacionan con el trabajo de esta tesis.

2.1 INTRODUCCIÓN: REVISIÓN DE LOS PROTOCOLOS DE AUTENTIFICACIÓN PARA SISTEMAS DISTRIBUIDOS

Un sistema distribuido es susceptible a una gran variedad de amenazas de seguridad montadas por intrusos. Un sistema distribuido es una colección de hosts interconectados por una red. Posee algunos problemas complejos de seguridad [1]. Una preocupación fundamental es la autentificación de entidades locales y remotas en el sistema. En un sistema distribuido, los hosts se comunican por el envío y recepción de mensajes sobre la red. Varios recursos (como archivos e impresoras) distribuidos entre los hosts son compartidos a través de la red en forma de servicios de red proporcionados por servidores. Los procesos individuales (clientes) que desean el acceso a los recursos solicitan servicio directo a los servidores apropiados. Aparte de tales computos del cliente-servidor, hay muchas otras razones por que tener un sistema distribuido. Por ejemplo, una tarea puede ser dividida en sub-tareas que sean ejecutadas concurrentemente en diferentes hosts.

Un sistema distribuido es susceptible a una variedad de amenazas montadas por intrusos así como usuarios legítimos del sistema. De hecho, los usuarios legítimos son adversarios muy poderosos, ya que ellos poseen información interna que no es usualmente disponible para un intruso (excepto después de una penetración acertada de un host.)

En este sentido se identifican dos tipos generales de amenazas:

El primer tipo es, compromiso del host, referente a la suversión de hosts individuales en un sistema. Varios grados de subversión son posibles, al extenderse del caso relativamente benigno de corromper la información del estado del proceso, al caso extremo del control total asumido de un host. Amenazas de compromisos de host pueden ser contrarrestadas por una combinación de técnicas de hardware (como modos de protección de procesador) y técnicas de software (como monitores de referencia.). Ya que estas técnicas no son nuestro principal motivo de estudio, se hace referencia a los lectores interesados a Denning [2] para una completa información de seguridad en sistemas de computo. Aquí, se asume que cada host implementa un monitor de referencia que puede ser confiable para segregar procesos correctamente.

El segundo tipo, compromiso de comunicación, incluye amenazas asociadas con comunicaciones de mensajes. Se pueden subdividir estas dentro de:

(CC1) (Eavesdropping) Escucha de mensajes transmitidos sobre enlaces de red para extraer información en conversaciones privadas.

(CC2) Modificación arbitraria, inserción, y cancelado de mensajes transmitidos sobre enlaces de red para confundir a un receptor dentro de mensajes fabricados de aceptación.

(CC3) reenvío de mensajes viejos (una combinación de (CC1) y (CC2).)

(CC1) es una amenaza pasiva, mientras (CC2) y (CC3) son activas. Una amenaza pasiva no afecta el sistema que esta siendo amenazado, mientras que una amenaza activa sí. Por lo que, las amenazas pasivas son inherentemente indetectables por el sistema y solamente pueden ser evitadas utilizando medidas preventivas. Las amenazas activas, por el otro lado, son combatidas por una combinación de técnicas de prevención, detección, y recuperación. No se consideran en este momento las amenazas de "análisis de trafico" y "negación del servicio" por que ellas son más relevantes para la seguridad general de un sistema distribuido que para este grupo restringido de autentificación.

Por lo tanto, algunos requerimientos de seguridad básicos pueden ser formulados. Por ejemplo, secrecia e integridad son dos requerimientos comunes para la comunicación segura. La secrecia especifica que un mensaje puede ser leído solamente por sus receptores destinados, mientras la integridad especifica que cada mensaje es recibido exactamente como fue enviado, o una discrepancia es detectada.

Un criptosistema fuerte puede proporcionar un alto nivel de confianza para la secrecia e integridad (ver "criptografía básica" sección 2.1.1 [3,4,5,6]) Mas aún, un mensaje encriptado no proporciona información en cuanto al mensaje original, de aquí que garantice la secrecia, además si es adulterado, no sería desencriptado dentro de un mensaje legal, por lo que garantiza la integridad.

El reenvío de mensajes viejos puede ser prevenido por el uso de "nonces[6]" o "time stamps[7]" Un nonce es información que esta garantizada reciente, esto es, no ha aparecido o sido usada antes. Por lo que, un reenvío que contiene algunas funciones de un envío de nonce recientemente debería ser considerado a tiempo, por que el reenvío debió haber sido generado solamente después de que el nonce fue enviado.

Números aleatorios perfectos son buenos candidatos de nonce; sin embargo, su eficacia es dependiente sobre la aleatoriedad que sea prácticamente realizable. El uso de time stamps requiere por lo menos alguna ligera sincronización de todos los relojes locales, y por lo tanto, su eficacia es también algo restringida.

2.1.1 CRIPTOGRAFÍA BÁSICA

Un criptosistema viene con un procedimiento para encripción y otro para desencripción. Una descripción formal de un criptosistema incluye las especificaciones para mensaje, llave, espacios de texto cifrado, y funciones de encripción y desencripción.

Existen dos amplias clases de criptosistemas, simétricos y asimétricos [3]. En el primero las llaves de encripción y desencripción son las mismas y deben mantenerse en secreto. En el segundo, la llave de encripción difiere de la llave de desencripción, y la llave de desencripción es mantenida secreta. La llave de encripción, sin embargo puede ser publica, consecuentemente, es importante que nadie pueda determinar la llave de desencripción desde la llave de encripción. Los criptosistemas simétricos y asimétricos son también referenciados como criptosistema de llave compartida y y criptosistema de llave pública respectivamente.

El conocimiento de la llave de encripción nos permite encriptar mensajes arbitrariamente desde el espacio del mensaje, donde el conocimiento de la llave de desencripción nos permite recuperar el mensaje de su forma encriptada. Así, las funciones de encripción y desencripción satisfacen la siguiente relación:

M es el espacio del mensaje, K_E x K_D es el sistema del par de llaves de encripción-desencripción:

$$\forall m \in M : \forall (k, k^{-1}) \in K_E \times K_D : \{\{m\}_k\}_{k^{-1}} = m \quad \text{.....(1)}$$

donde $\{X\}_Y$ denota la operación de encripción sobre el mensaje X si Y es una llave de encripción y la operación de desencripción sobre X si Y es una llave de desencripción. (En el caso del criptosistema simétrico con llaves de encripción y desencripción idénticas, la operación debera ser clara desde el contexto).
Dos criptosistemas han sido ampliamente utilizados son: el Estándar de Encripción de Datos (DES) [4] sistema simétrico, y RSA [5] sistema asimétrico. En RSA, el par de llaves de encripción desencripción satisfacen las siguientes propiedades conmutativas [6]:

---

[6] Los "Nonces" son números aleatorios comúnmente utilizados en criptosistemas de seguridad

[7] Los "time stamps" son señales de tiempo de vida (valores de un reloj local) también muy utilizados en criptosistemas.

$$\forall m \in M : \forall (k, k^{-1}) \in K_E \times K_D : \{\{m\}_{k^{-1}}\}_k = m \quad \text{.....(2)}$$

Rinden una capacidad de la firma. Esto es, suponiendo que k y k^{-1} son llaves asimétricas de David, luego $\{m\}_{k^{-1}}$ puede ser usada como firma de David sobre m, puesto que habría podido ser producida solamente por David, el principal[8] únicamente conoce k^{-1}. De acuerdo con la ecuación 2, la firma de David es verificable por cualquier principal con conocimiento de k, La llave publica de David. Cabe hacer notar que en la ec.2, los roles de k y k^{-1} se invierten, específicamente, k^{-1} es utilizada como llave de encripción mientras que las funciones de k son como llave de desencripción, Para evitar confusión con los roles más típicos para k y k^{-1} como es ejemplificado en la ec. 1, nos referimos a la encripción de k^{-1} como operación de firma.
Ya que, en la practica, los criptosistemas simétricos pueden operar mucho más rápido que los criptosistemas asimétricos, los criptosistemas simétricos pueden ser utilizados tanto para inicializaciones y transferencia de datos actuales, mientras que los criptosistemas asimétricos son típicamente utilizados únicamente para inicialización-funciones de control.

2.1.2 ¿QUÉ NECESITA DE ESQUEMAS Y PROTOCOLOS DE AUTENTIFICACIÓN?

En términos simples, la autentificación es, identificación más verificación. La identificación es el proceso por el que una entidad demanda una cierta identidad, mientras que la verificación es el proceso por el que esa demanda es validada. Así, la corrección de una excesiva confianza en autentificación en el procedimiento de verificación es empleada.

Hay tres tipos centrales de autentificación en sistemas de computo distribuido:

(A1), Autentificación del contenido del mensaje. Verifica que el contenido de un mensaje recibido es el mismo que cuando este fue enviado.
(A2), Autentificación de origen del mensaje. Verifica que el transmisor de un mensaje recibido es el mismo guardado en el campo del transmisor de un mensaje.
(A3), Autentificación de identidad general. Se cerciora de que la identidad de un principal es como es demandada.

(A1) Es comúnmente manejada marcando con un Código de Autentificación del Mensaje (MAC) dependiente de una llave sobre un mensaje antes de que este sea enviado. La integridad del mensaje puede ser confirmada en la recepción calculando el MAC y comparando este con el otro que viene agregado.

(A2) es un sub-caso de (A3). Una autentificación de identidad general completa, usualmente resulta en una creencia tenida por el principal de autentificación (el verificador) que el principal autentificado (el demandante) posee la identidad solicitada. Por tanto, las acciones de demanda subsecuentes son atribuidas para la identidad solicitada; por ejemplo, la autentificación de identidad general es necesaria tanto para autorización como para funciones del estado de cuenta.

---

[8]Las entidades en un sistema distribuido que pueden ser distintamente identificadas son referenciadas colectivamente como principales.

En un ambiente donde ambos, host y compromisos de comunicación puedan ocurrir, los principales deberían adoptar una actitud sospechosa mutuamente. En consecuencia, la autentificación mutua, por la que ambos principales comunicándose verifican la identidad de cada uno, en vez de la autentificación de un solo sentido, en donde solamente un principal verifica la identidad del otro principal es usualmente requerida.

En un ambiente de computo distribuido, la autentificación es llevada a cabo utilizando un protocolo que involucra intercambios de mensajes. Hacemos referencia a estos protocolos como protocolos de autentificación.

La mayoría de los sistemas existentes, utilizan solamente medidas de autentificación muy primitivas o ninguna. Por ejemplo:

- El procedimiento de "login" muy generalizado requiere que los usuarios den su clave secreta (password) en respuesta a una señal del sistema. Los usuarios son luego autentificados en un solo sentido por la verificación de la clave secreta contra una tabla guardada internamente. Sin embargo, ningún mecanismo les permite a los usuarios autentificar al sistema. Este diseño es aceptable solamente cuando el sistema es confiable, o la probabilidad de compromiso es baja.

- En una interacción típica cliente-servidor, el servidor para aceptar una solicitud del cliente tiene que creer que: (1) el host residente del cliente tiene autentificado correctamente al cliente y (2) la identidad suministrada en la solicitud actualmente corresponde al cliente. Tal confianza es valida solamente si los hosts del sistema son confiables y sus canales de comunicación son seguros.

Estas medidas son formalmente inadecuadas por que la noción de confianza, en sistemas distribuidos es entendida pobremente. Una explicación formal satisfactoria de confianza tiene aun que ser propuesta, avalada y refinada por las partes involucradas. En segundo lugar, la proliferación de una gran escala de sistemas distribuidos abarcando dominios de administración múltiple ha producido relaciones de confianza extremadamente complejas.

En un sistema de computo distribuido, las entidades que requieren identificación son los hosts, usuarios, y procesos como se comenta en [7]. Ellos de este modo constituyen los principales involucrados en una autentificación, los cuales se describen a continuación.

Hosts. Estos son entidades direccionables en el nivel de red. Un host es distinguido por el hardware de soporte fundamentalmente. Por ejemplo, el host H corriendo en una estación de trabajo A puede ser movido hacia la estación de trabajo B por la presentación de la secuencia "bootstrap" para H en B. Un host es generalmente identificado por su nombre (por ejemplo, un nombre de dominio) o su dirección de red (por ejemplo, una dirección de internet), mientras que un hardware de host particular es generalmente identificado por su numero serial asignado de fabrica (por ejemplo, una estación de trabajo en "ethernet" puede ser identificada por la dirección única de esta tarjeta de adaptación de "ethernet".).

Usuario. Estas entidades son finalmente responsables por todas las actividades del sistema. En otras palabras, los usuarios inician y son tomados en cuenta para todas las actividades del sistema. La mayoría de los controles de acceso y funciones de estados de cuentas son basadas en usuarios. Dentro de los usuarios típicos estan incluidos los humanos, así como las cuentas

mantenidas en la base de datos. Cabe destacar, que los usuarios son considerados para estar fuera de los límites del sistema.

Procesos. El sistema crea procesos dentro de la frontera del sistema para representar a los usuarios. Un proceso solicita y consume recursos en nombre de sus únicos usuarios asociados. Los procesos caen en dos clases:
Cliente y servidor, los procesos de cliente son los consumidores, quienes obtienen servicios de los procesos del servidor, quienes son proveedores de servicio. Un proceso particular puede actuar como ambos, un cliente y un servidor. Por ejemplo, los servidores de impresión son generalmente creados por el usuario raíz y actúan como servidores para solicitudes de impresión por otros procesos. Sin embargo, ellos actúan como clientes cuando solicitan archivos desde el servidor de archivos.

2.1.3 INTERCAMBIOS DE AUTENTIFICACIÓN

Podemos identificar los siguientes tipos importantes de intercambio de autentificación en un sistema distribuido.

Host-host. Las actividades del nivel de host a menudo requieren cooperación entre hosts. Por ejemplo, hosts individuales intercambian información de enlace para actualizar sus mapas de topologías internas. En la carga de inicialización remota, un host, en la reiniciación, debería identificar un servidor de "carga" confiable para abastecerse de la información (por ejemplo, una copia del sistema operativo) requerida para una correcta inicialización.

Usuario-host. Un usuario consigue acceso a un sistema distribuido identificándose en un host en el sistema. En un ambiente de acceso abierto donde los hosts son diseminados a través de áreas no restringidas, un host puede ser arbitrariamente comprometido, necesitando autentificación mutua entre el usuario y el host.

Proceso-proceso. Existen dos sub-clases principales:

- Comunicación de proceso-igual. Procesos iguales deben ser satisfechos con cada identidad de los otros antes que pueda empezar la comunicación privada.
- Comunicación cliente-servidor. Una decisión de acceso respecto a una solicitud de cliente puede ser hecha solamente cuando la identidad del cliente es afirmada. Un cliente es complaciente de ceder información valiosa a un servidor solamente después de verificar la identidad del servidor.

2.1.4 PARADIGMAS DE LOS PROTOCOLOS DE AUTENTIFICACIÓN

La autentificación en los sistemas distribuidos debe ser generalmente llevada a cabo con protocolos. Un protocolo es una secuencia definida, precisamente, de pasos de cálculos y comunicación. Un paso de comunicación transfiere mensajes desde un principal (transmisor) hacia otro (receptor). Mientras que el paso de calculo actualiza un estado interno del principal.

Dos estados distintos pueden ser identificados sobre la terminación del protocolo, uno significando autentificación satisfactoria y el otro de falla.

Aún cuando el objetivo de cualquier autentificación es para verificar la identidad demandada de un principal, estados específicos del éxito y de la falla son altamente dependientes del protocolo. Por ejemplo, el éxito de una autentificación durante la fase de establecimiento de conexión de un protocolo de comunicación es generalmente indicado por la distribución de una llave de sesión actual, entre dos procesos iguales autentificados mutuamente. Por otra parte, en una autentificación de conexión de usuario, el éxito generalmente resulta en la creación de un proceso de conexión ("login") en nombre del usuario.
En este apartado presentaremos protocolos en el formato siguiente:

Un paso de comunicación por el que P envía un mensaje M para Q es representado como P→ Q: M. Mientras que un paso de calculo de P es escrito como P: . . . , donde ". . ." es una especificación del paso de computo. Por ejemplo, el protocolo de conexión (login) típico entre el host H y el usuario U es dado abajo. (f es una función de un solo sentido: Esto es, dado y, es computacionalmente imposible encontrar una x tal que f(x) = y.).

U → H: U
H → U: " Por favor dar su clave ("password")
U → H: p
H : calcula y = f(p)
: recupera los registros del usuario (U, $f(password_U)$) de la base de datos del usuario.
: si y = $f(password_U)$ entonces acepta; de otra forma rechaza

A continuación examinamos las ideas clave a subrayar del diseño de protocolos de autentificación presentando varios paradigmas de protocolos.

En vista de que los paradigmas de los protocolos de autentificación directamente utilizan criptosistemas, sus principios de diseño básicos también siguen finalmente el tipo de criptosistema utilizado.

Hay que notar que el paradigma del protocolo ilustra solamente los principios de diseño básicos. Un protocolo realista es necesariamente un refinamiento de este paradigma básico y suposiciones de ambientes de débiles direcciones, poscondiciones más fuertes, o ambos. También, un protocolo realista podría utilizar tanto criptosistemas simétricos y asimétricos.

Protocolos basados en criptosistemas simétricos. En un criptosistema simétrico, conociendo la llave compartida, permite al principal encriptar[9] y desencriptar[10] mensajes arbitrariamente. Sin tal conocimiento, un principal no puede obtener la versión encriptada de un mensaje o desencriptar un mensaje encriptado. En consecuencia, los protocolos de autentificación pueden ser diseñados de acuerdo con el principio (SYM): *Si un principal puede correctamente encriptar un mensaje utilizando una llave que el verificador cree que es conocida solamente para un principal con la identidad demandada (fuera de lugar del verificador), este acto constituye suficiente prueba de identidad.*

[9] Acto de la encripción
[10] Acto de la desencripción

En este caso (SYM) incluye la prueba –por-principio de conocimiento para la autentificación, eso es, un conocimiento del principal es indirectamente demostrado a través de encripción (ver "Aproximaciones a la autentificación".) Utilizando (SYM), inmediatamente obtenemos el protocolo básico siguiente (k es una llave simétrica compartida entre P y Q.)

P : crea m = "Yo soy P"
: calcula $m' = \{m\}_k$
P → Q: m, m´
Q : Verifica $\{m\}_k$?= m´
: si es igual entonces acepta; de otra forma rechaza.

Claramente, el principio de diseño (SYM) es bueno solamente si el criptosistema fundamental es fuerte (Uno no puede encontrar la versión encriptada de un mensaje sin conocimiento de la llave) y la llave es secreta (esta es compartida solamente entre el principal autentico y el verificador) Notemos que este protocolo presenta solamente autentificación de un solo sentido. Autentificación mutua puede ser conseguida invirtiendo los roles de P y Q.

Una debilidad del protocolo es su vulnerabilidad a reenvíos. Específicamente, un adversario puede enmascararse como P grabando los mensajes m´ y después reenviando este a Q. Como se menciono, los ataques de reenvío pueden ser controlados utilizando nonces o time stamps. Es decir, modificando el protocolo agregando un paso de reto-respuesta utilizando nonces (donde n es un nonce.)

P : "Yo soy P"
Q → P: n
P : calcula $n' = \{n\}_k$
P → Q: n´
Q : Verifica $\{n\}_k$?= n´; si es igual entonces acepta; de otra forma rechaza.

El reenvío es frustrado por lo resiente de n. Por lo tanto, aún si un escuchante no autorizado ha monitoreado todas las conversaciones de autentificación previas entre P y Q, este no podría todavía producir el correcto n´. (Esto también señala la necesidad para los criptosistemas para soportar el sabido ataque de "plaintext [11]". En otras palabras, el criptosistema debe ser irrompible dando el conocido par "plaintext-ciphertext".) El paso reto-respuesta puede ser repetido cualquier numero de veces hasta que el nivel deseado de confidencia es alcanzado por Q.

Este protocolo es impractico como una solución general en escalas-largas, por que cada principal debería guardar en memoria la llave secreta para cada uno de los otros principales que quiera autentificarse. Esto presenta mayor inicialización (la predistribución de llaves secretas) y problemas de almacenaje. Más aún, el compromiso de un principal puede potencialmente comprometer al sistema completo. Estos problemas pueden ser reducidos significativamente postulando un servidor de autentificación centralizado "A" que comparta una llave secreta k_x con

[11] En algunos textos se puede encontrar esta traducción como "texto en claro" o "texto plano" pero a lo largo de este documento lo manejaremos sin traducción. De igual forma lo aplicaremos para el "ciphertext" al cual se le encuentra como traducción " texto cifrado".

cada principal "X" en el sistema como se ve en [8]. El protocolo de autentificación básico entonces queda:

P : "Yo soy P"
Q → P: n
P : calcula $n' = \{n\}_{kP}$
P → Q: n'
Q : calcula $n'' = \{P,n'\}_{kQ}$
Q → A: n''
A : recupera (P, n') desde n'' desencriptando con k_Q
: calcula $m = \{\{n'\}_{kP}\}_{kQ}$
A → Q: m
Q : Verifica $\{n\}_{kQ}$?= m
: si es igual entonces acepta; de otra forma rechaza.

Así el paso de verificación de Q es precedido por el paso de traducción de llave de A. La corrección del protocolo ahora también descansa en la confiabilidad de A. Que A podrá propiamente desencriptar utilizando la llave de P y nuevamente encriptar utilizando la llave de Q. Los problemas de inicialización y almacenaje son ampliamente aliviados por que ahora cada principal necesita mantener una sola llave. El riesgo de comprometerse es en su mayor parte trasladado a A. Cuya seguridad puede ser garantizada por varias medidas, tal como encriptando las llaves guardadas, empleando una llave maestra y poniendo a "A" físicamente en un cuarto seguro.

Protocolos basados en criptosistemas asimétricos. En un criptosistema asimétrico, cada principal P publica su llave publica k_P y mantiene secreta su llave privada k_P^{-1}. De esta forma solamente P puede generar $\{m\}k_P^{-1}$ para cualquier mensaje m firmando este empleando k_P^{-1}. El mensaje firmado $\{m\}k_P^{-1}$ puede ser verificado por cualquier principal con conocimiento de k_P (asumiendo un criptosistema asimétrico conmutativo.). El principio de diseño básico es (ASYM): *Si un principal puede correctamente firmar un mensaje empleando la llave privada de la identidad demandada, este acto constituye la prueba suficiente de la identidad.*

Este principio (ASYM) sigue el principio de prueba-por-conocimiento para autentificación, en el que un conocimiento del principal es indirectamente demostrado a través de su capacidad de firmado. Utilizando (ASYM), podemos obtener un protocolo básico como se muestra (donde nuevamente n es un nonce):

P → Q : "Yo soy P"
Q → P: n
P : calcula $n' = \{n\}k_P^{-1}$
P → Q: n'
Q : Verifica n ?= $\{n'\}k_P$
: si es igual entonces acepta; de otra forma rechaza.

Este protocolo depende de la garantía que $\{n\}k_P^{-1}$ no pueda ser producida sin el conocimiento de k_P^{-1} y la correcta k_P como la publicada por P y guardada por Q.

Como en los protocolos que emplean llaves simétricas, los problemas de inicialización y almacenamiento pueden ser aliviados postulando una autoridad de certificación centralizada "A"

que mantenga una base de datos de todas las llaves publicas publicadas. El protocolo puede entonces ser modificado como sigue.

P → Q : "Yo soy P"
Q → P: n
P : calcula $n' = \{n\}k_P^{-1}$
P → Q: n'
Q → A: "Necesito la llave publica de P"
A : recupera $c = \{P, k_P\}k_A^{-1}$ desde la base de datos de llaves
A → Q: P, c
Q : Recupera (P,k_P) desde c desencriptando con k_A
. Verifica $n?= \{n'\}k_P$
: si es igual entonces acepta; de otra forma rechaza.

Así el certificado "c" de llave publica representa una manifestación de certificación por A que la llave publica de P es k_P. Otra información tal como un tiempo de vida especificado y la clasificación del principal P (para control de acceso mandatorio) puede también ser incluida en el certificado (tal información es omitida aquí en este momento.) Cada principal en el sistema necesita guardar solamente una copia de la llave k_A de A.

En este protocolo, A es un ejemplo de una autoridad de certificación en línea, la cual soporta preguntas interactivas y está activamente involucrada en intercambios de autentificación. Una autoridad certificadora puede también operar fuera de línea de modo que un certificado de llave publica es emitido para cada principal solamente una vez. El certificado es guardado por el principal y reenviado durante un intercambio de autentificación, de tal forma eliminando la necesidad de preguntar a A interactivamente. Alternativamente, el certificado puede ser guardado en una base de datos en línea que sea públicamente accesible. La falsificación es imposible, ya que un certificado es firmado por la autoridad certificadora.

Aproximaciones a la autentificación

Todos los procedimientos de autentificación involucran comprobación de información conocida respecto a una identidad demandada contra información adquirida del demandante durante el procedimiento de verificación de identidad. Por lo que la comprobación puede estar basada en las siguientes tres aproximaciones:

Prueba por conocimiento. El demandante conoce información respecto de la identidad demandada que puede solamente ser conocida o producida por un principal con esa identidad. Por ejemplo, el conocimiento del password es necesario para la mayoría de los procedimientos de conexión. Una prueba por conocimiento puede ser guiada por una demostración directa, como escribiendo en un password, o por una demostración indirecta, tal como respuestas calculadas correctamente para verificar retos. La demostración directa no es preferible desde el punto de vista de la seguridad, ya que un verificador comprometido puede grabar el conocimiento enviado y después hacer "Impersonation[12] " al demandante presentando los conocimientos guardados. La demostración indirecta puede ser diseñada para incluir alta confidencialidad en el verificador sin dejar ninguna pista de cómo las respuestas de los demandantes son calculadas. Por ejemplo, Feige, Fiat y Shamir, [9] propusieron un protocolo de cero-conocimiento para prueba de

[12] Acceder a un sistema utilizando la identidad de otra persona

identidad. Este protocolo permite al demandante "D" probar al verificador "V" que "D" conoce como calcular respuestas a los retos sin revelar las respuestas. Estos protocolos son probablemente seguros (bajo suposiciones de complejidad) Sin embargo, refinamientos adicionales son necesarios antes de que ellos puedan ser aplicados en sistemas prácticos.

Prueba por posesión. El demandante produce un "item [13]" que puede solamente ser poseído por un principal con la identidad demandada, por ejemplo, un ID ofrecido. El item tiene que ser inolvidable y guardado seguramente.

Prueba por propiedad. El verificador directamente mide ciertas propiedades del demandante con técnicas biométricas como impresión de dedo, o impresión de retina. La propiedad medida tiene que estar distinguiendo, que es, única entre todos los principales posibles.

La prueba por conocimiento y posesión (y combinaciones de estos) pueden ser aplicadas a todos los tipos de autentificación necesarios en un sistema distribuido seguro, mientras que la prueba por propiedad es generalmente limitada a la autentificación de usuarios humanos por un host equipado con instrumentos de medida especializados.

Noción de confianza. La corrección de ambos tipos de paradigmas de protocolos requieren mas que la existencia de canales de comunicación seguros entre principales y los apropiados servidores de autentificación (o autoridades certificadoras.). De hecho, tal corrección es críticamente dependiente de la capacidad de los servidores (autoridades) para seguir los protocolos fielmente. Cada principal basa sus criterios en sus propias observaciones (mensajes enviados y recibidos) y su confianza en los criterios del servidor.

Desde un punto de vista particular, menos confianza es requerida de una autoridad de certificación que de un servidor de autentificación, por que toda la información guardada por la autoridad es publica (excepto por su propia llave privada) Además, una autoridad de certificación no tiene sentido de enmascararse como un principal por que una llave privada de un principal no es compartida.

Nuestro entendimiento formal de confianza en un sistema distribuido es aún en el mejor caso inadecuado. En particular, un entendimiento formal de autentificación requeriría ambos, una especificación formal de confianza y un riguroso método razonable en donde la confianza es un elemento básico.

2.1.5 DEFECTOS DE PROTOCOLOS DE AUTENTIFICACIÓN

A pesar de la aparente simplicidad de los principios básicos de diseño para protocolos de autentificación, diseñar protocolos realistas es notoriamente difícil. Varios protocolos publicados han exhibido sutiles problemas de seguridad. [2,8,10].

Hay varias razones para que estas dificultades existan. La mayoría de los criptosistemas realistas satisfacen las identidades algebraicas vistas en las ecuaciones (1) y (2). Estas propiedades extra podrían generar efectos indeseables cuando son combinados con la lógica del protocolo.

---

[13] partida de información

En segundo lugar, aún asumiendo que el criptosistema fundamental es perfecto, una interacción inesperada entre los pasos del protocolo puede protagonizar sutiles defectos lógicos.

En tercer lugar, las suposiciones respecto del ambiente y las capacidades de un adversario no siempre son explícitamente especificadas, haciendo esto extremadamente difícil para determinar cuando un protocolo es aplicable y que estados finales son alcanzados.

Ilustraremos la dificultad mostrando un protocolo de autentificación propuesto por Needham y Schroeder [8], el cual contiene unas sutiles debilidades. [2] Las llaves simétricas k_P y k_Q son compartidas entre P y A, y Q y A, respectivamente, donde A es un servidor de autentificación. Siendo k una llave de sesión.

(1) $P \rightarrow A$: P, Q, n_P
(2) $A \rightarrow P$: $\{n_P, Q, k, \{k, P\}k_Q\}k_P$
(3) $P \rightarrow Q$: $\{k,P\}k_Q$
(4) $Q \rightarrow P$: $\{n_Q\}_k$
(5) $P \rightarrow Q$: $\{n_Q + 1\}_k$

El mensaje $\{k, P\}k_Q$ en el paso 3 puede ser solamente desencriptado y en consecuencia entendido por Q. El paso 4 refleja el conocimiento de k por parte de Q, mientras que el paso 5 infunda confianza a Q del conocimiento de k por parte de P. Por tanto el "handshake" de autentificación esta basado completamente en el conocimiento de k. La sutil debilidad en el protocolo se origina desde el hecho que el mensaje $\{k, P\}k_Q$ enviado en el paso 3 no contiene información para Q para verificar su actualidad. (Hay que notar que solamente P y A conocen que k sea actual.). De hecho, este es el primer mensaje enviado a Q respecto a la intención de P de establecer una conexión segura. Un adversario el cual tiene comprometida una llave de sesión vieja k' puede impersonar a P por el reenvío del mensaje guardado $\{k', P\}k_Q$ en el paso 3 y subsecuentemente ejecutando los pasos 4 y 5 utilizando k'.

Para evitar los defectos de los protocolos, métodos formales podrían ser empleados en el diseño y verificación de los protocolos de autentificación. Un método de diseño formal debería involucrar los principios básicos de diseño. Más aún, el razonamiento informal debería ser formalizado dentro de un método de verificación. Tal razonamiento informal incluiría proposiciones como "si tu crees que solamente tu y Bob conocen k, entonces tu deberías creer que cualquier mensaje que tú recibas encriptado con k fue originalmente enviado por Bob", el cual debe ser formalmente especificado.

Los primeros intentos de verificación formal de protocolos de seguridad, principalmente siguieron una aproximación o propuesta algebraica [11]. Los mensajes intercambiados en un protocolo son vistos como términos en un álgebra. Varias identidades involucrando los operadores de encripción y desencripción (por ejemplo, las ec. (1) y (2)) son tomados para ser reglas de reescritura de término.

Un protocolo es seguro si de este es imposible derivar ciertos términos (por ejemplo, él termino que contiene la llave) desde los términos obtenidos por un adversario. La aproximación algebraica es limitada, puesto que se ha utilizado principalmente para ocuparse de un aspecto de la seguridad, llamado secreto. Recientemente, aproximaciones lógicas han sido propuestas para estudiar protocolos de autentificación [10]. La mayoría de estas lógicas adoptan un modelo base,

con creencia y conocimiento como nociones principales. Las propuestas lógicas aparecieron para ser mas generales que las algebraicas, pero en un principio, carecían de los fundamentos rigurosos de sistemas más establecidos como lógicas de primer orden y temporales. Para más detalles de la lógica empleada en esta tesis revisar el anexo A

2.1.6 ESTRUCTURA DE LA AUTENTIFICACIÓN

Sintetizando algunos conceptos básicos dentro de una estructura de autentificación que pueden ser incorporados dentro del diseño de sistemas distribuidos seguros. Se identifican cinco aspectos de diseño de sistemas distribuidos seguros. Y de las necesidades de autentificación asociadas. Estos aspectos son:

- Iniciación del host. Todas las ejecuciones de procesos toman lugar dentro de los hosts. Algunos hosts (como estaciones de trabajo) también actúan como puntos de entrada al sistema permitiendo a los usuarios conectarse. La autentificación puede ser utilizada para implementar "bootstrapping" seguro.

- "logins" de usuario. La identidad de usuarios es establecida en el "login", y todas las actividades del usuario subsecuentes son atribuidas a esta identidad. Una autentificación mutua usuario-host puede alcanzar las garantías requeridas.

- Comunicaciones iguales. Los sistemas distribuidos pueden distribuir una tarea sobre múltiples hosts para alcanzar un alto desempeño o mayor utilización balanceada que en los sistemas centralizados. La corrección de una tarea distribuida depende de que sean procesos iguales participando en la tarea que pueda identificar correctamente la identidad del otro. La autentificación puede identificar amigos o enemigos.

- Interacciones cliente-servidor. El modelo cliente-servidor proporciona un paradigma atractivo para la construcción de sistemas distribuidos. Los servidores están gustosos de proporcionar servicio solamente a los clientes autorizados, mientras que los clientes están interesados en pactar solamente con servidores legítimos. La autentificación puede ser usada para verificar una relación potencial consumidor-proveedor.

- Comunicación inter-dominios. La mayoría de los sistemas distribuidos no son centralmente administrados o de un solo dueño.

En ambientes maliciosos con amenazas como las tratadas en la sección 2.1, algunas suposiciones básicas respecto al sistema deberían ser satisfechas para alcanzar un nivel razonable de seguridad. (Para revisar algunas posibles suposiciones, ver Abadi et al, [12] y Linn [7]). La figura 2.1 también describe estas suposiciones

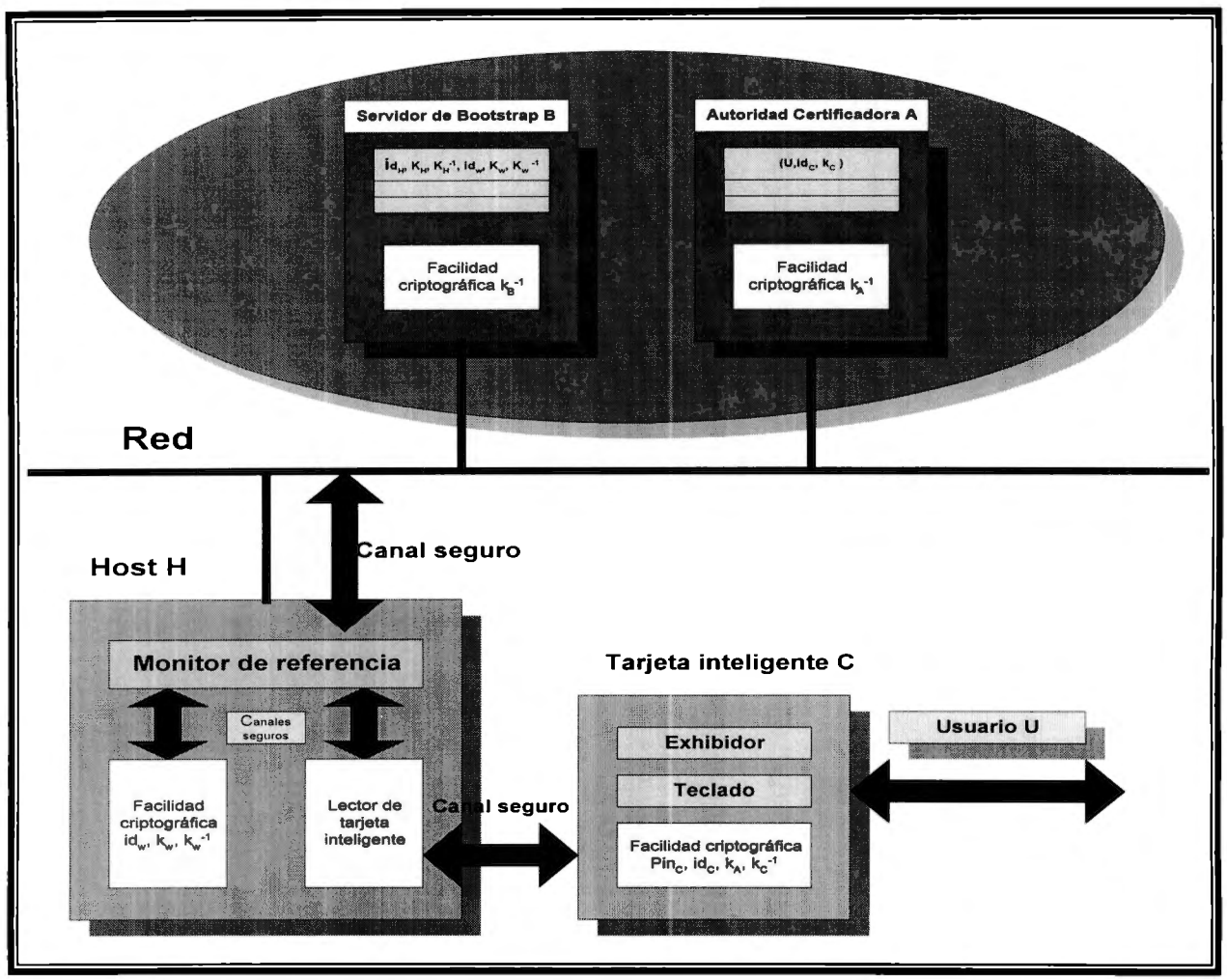


Figura 2.1. Configuración del sistema distribuido

A continuación se puede ver el protocolo que describe la grafica anterior.

Protocolo de " bootstrap" seguro

(1) W → todos: "boot" , id_W, $\{n_W, id_W\}k_W$
(2) B : Recupera lo guardado (Id_H, k_H, k_H^{-1},id_w, k_w, k_w^{-1})
para W desde la base de datos
: recupera n_W desde $\{n_W\ id_W\}k_W$, desencriptando con k_W^{-1}
: genera una llave aleatoria k
: calcula m = $\{n_W, k_A, k_B, k\}k_W$
(3) B → W : m
(4) W : recupera (n_W, k_A, k_B, k) desde m
desencriptando con k_W^{-1}
(5) W → B : $\{n_W, \text{"listo"}\}_k$
(6) B → W : $\{n_W, n_B, id_H, \{k_H^{-1}\}k_W, OS\}_k$
(7) W → B : $\{\{n_B\}k_H^{-1}\}_k$
(8) B → W : $\{id_H, id_W, k_H, T\}k_B^{-1}$
(9) H : valida el certificado $\{id_H, id_W, k_H, T\}k_B^{-1}$
encriptando con k_B

Protocolo de autentificación de usuario-host

(1) C → H : id_C, n_C
(2) H → A : id_C, $\{id_H, id_W, k_H, T\} k_B^{-1}$)
(3) A : checa el "time stamp" del certificado
Si el "time stamp expiro, aborta.
(4) A → H : $\{ id_H, id_W, k_H, T\} k_A^{-1}$, $\{U, id_C, k_C\} k_A^{-1}$
(5) H : genera una nueva delegación del par de llaves (k_d , k_d^{-1})
(6) H → C : $\{ id_H, id_W, k_H, T\} k_A^{-1}$, $\{U, k_d, n_C\} k_H^{-1}$
(7) C → U : id_H, id_W
(8) U : verifica si id_H / id_W es el host deseado
: si no aborta
(9) U → C : Pin
(10) C : Verifica Pin ¿=? Pin_C
: si no es igual aborta
(11) C → H : $\{U, id_H, k_d, T_c\}k_C^{-1}$
(12) H : verifica la delegación correcta desencriptando el certificado de login
$\{U, id_H, k_d, T_c\}k_C^{-1}$ con k_C

Protocolo de autentificación igual-igual.

Este tipo de autentificación mutua y negociación de parámetros criptográficos es presentado en la fase de conexión-establecimiento de un protocolo orientado a conexión seguro.

(1) P → Q: P, Q
(2) A → P: $\{Q, k_Q\}k_A^{-1}$
(3) P → Q: $\{n_P, P\}k_Q$
(4) Q → A: Q, P, $\{n_P\}k_A$
(5) A → Q: $\{P, k_P\}k_A^{-1}$, $\{\{n_P, k, Q\}k_A^{-1}\}k_Q$
(6) Q → P: $\{\{n_P, k, Q\}k_A^{-1}, n_Q\}k_P$
(7) P → Q: $\{n_Q\}k$

Autentificación inter-dominios. Esto es, asumiendo una autoridad de confianza de certificación centralizada para todos los principales. Sin embargo, un sistema distribuido realista es a menudo compuesto de subsistemas independientemente administrados por diferentes autoridades. Utilizando él termino "dominio" para referirnos a tal subsistema. Cada dominio D mantiene su propia autoridad certificadora A_D que tiene jurisdicción sobre todos los principales dentro del dominio. Autentificación intra-dominio se refiere a un intercambio entre dos principales perteneciente al mismo dominio, mientras que autentificación inter-dominio se refiere a un intercambio que involucra dos principales pertenecientes a diferentes dominios.

2.1.7 Algunos Casos de estudio

En esta sección analizamos dos protocolos de autentificación: Kerberos y SPX. Ambos puntualizan las necesidades de autentificación de cliente-servidor. Sus servicios son generalmente disponibles para un programa de aplicación a través de una interfase programable. Mientras Kerberos utiliza un criptosistema simétrico, SPX utiliza un criptosistema asimétrico.

Kerberos

Protocolo de inicialización de credencial.

U → H : U
H → Kerberos: U, TGS
Kerberos : recupera k_U y k_{TGS} desde una base de datos
: genera una nueva llave de sesión k
: crea un certificado de concesión-de ticket
$tick_{TGS} = \{U, TGS, k, T, L\}k_{TGS}$
Kerberos → H: $\{TGS, k, T, L, tick_{TGS}\}k_U$
H → U : " Password ?"
U → H : passwd
H : calcula $f = f(passwd)$
:recupera k, $tick_{TGS}$ desencriptando $\{TGS, k, T, L, tick_{TGS}\}k_U$ con f
: si la desencripción falla, aborta el "login"
: de otra forma retiene $tick_{TGS}$ y k
: borra passwd de la memoria

Protocolo de autentificación cliente-servidor de Kerberos

(1) C→ TGS : S, $tick_{TGS}$, $\{C,T_1\}_k$
(2) TGS : recupera k desde $tick_{TGS}$ desencriptando con k_{TGS}
: recupera T_1 desde $\{C, T_1\}_k$ desencriptando con k
: checa la oportunidad de T_1 con respecto al reloj local
: crea un certificado de servidor $tick_S = \{C, S, k', T', L'\}k_S$
(3) TGS → C : $\{S, k', T', L', tick_S\}_k$
(4) C : recupera k', $tick_S$ desencriptando con k
(5) C → S : $tick_S$, $\{C,T_2\}_{k'}$
(6) S : recupera k'desde $tick_S$ desencriptando co k_S
: recupera T_2 de $\{C, T_2\}_{k'}$ desencriptando co k'
: checa la oportunidad de T_2 con respecto al reloj local
(7) S → C : $\{T_2 + 1\}_{k'}$

SPX

Protocolo de iniciación de credencial de STX

(1) U → H : U, "passwd"
(2) H → LEAF : U, $\{T,n,h_1(paswd)\}k_{LEAF}$
(3) LEAF → CDC : U
(4) CDC → LEAF : $\{\{k_U^{-1}\}_{h2(passwordU)}, h_1(password_U)\}_k$, $\{k\}k_{LEAF}$
(5) LEAF : Recupera k desencriptando con k_{LEAF}^{-1}
:recupera $\{k_U^{-1}\}_{h2(passwordU)}$ y $h_1(password_U)$ desencriptando con k
: verifica $h_1(passwd) =? h_1(password_U)$
: si no es igual, aborta
(6) LEAF → H : $\{\{k_U^{-1}\}_{h2(passwordU)}\}n$

(7) H : recupera k_U^{-1} desencriptando primero con n y luego con h_2(paswd)
: genera (RSA) el par de llaves delegadas (k_d, k_d^{-1})
:crea el "ticket" $tick_U = \{L,U,k_d\}k_U^{-1}$
(8) H → CDC : U
(9) CDC → H : $\{A, k_A\}k_U^{-1}$

Donde: LEAF = "Login Enrollment Agent Facility"
CDC = "Certificate Distribution Center"

Protocolo de intercambio de autentificación entre cliente y servidor.

(1) C → CDC : S
(2) CDC → C : $\{ S, k_s\}_{k^{-1}AS}$
(3) C → S : T, $\{k\}_{ks}$, $ticket_C$, $\{ k_d^{-1}\}_k$
(4) S → CDC : C
(5) CDC → S : $\{C, k_C\}_{k^{-1}AC}$
(6) S : valida el $ticket_C$ por la encripción con k_C
(7) S→ C : $\{T + 1\}_k$

Comentarios

Con el crecimiento en la escala de sistemas distribuidos, la seguridad se ha convertido en una preocupación importante y un factor limitador en su diseño. La seguridad se ha abogado fuertemente como uno de los apremios principales del diseño en instituciones importantes, tales como el Massachussets Institute of Technology y Carnigie Mellon University entre otras. La mayoría de los sistemas distribuidos existentes, sin embargo, no tienen una estructura de seguridad bien definida y utilizan la autentificación solamente para sus aplicaciones más criticas. Aunque las nuevas y recientes versiones ya empiezan a contemplar e implementar dichas preocupaciones. Más aún, los protocolos hasta aquí presentados son prácticamente factibles. Y su adopción y uso debería ser una cuestión de necesidad.

2.2 DISEÑO, VERIFICACIÓN E IMPLEMENTACIÓN DE UN PROTOCOLO DE AUTENTIFICACIÓN

En esta sección, se presenta el análisis de un ciclo desarrollado completamente. (Diseño, especificación, verificación e implementación) de un protocolo de autentificación realista, el cual es parte de una arquitectura de seguridad propuesta por Y.C.Woo y Simon S. Lam. [13]

Para la implementación de este protocolo, se adoptaron los estándares propuestos GSS-API. Se describe el mapeo desde este protocolo hacia el GSS-API, el cual sirve como una referencia para otras implementaciones de protocolos.

2.2.1 DISEÑO DEL PROTOCOLO

Es un protocolo de autentificación "de igual-a-igual", similar en funcionalidad y estructura a otros (Kerberos, SPX.) Específicamente, se autentifica mutuamente a las dos partes qué se estén comunicando con la ayuda de una 3ª parte de confianza. En este estudio se consideran solamente la autentificación intra dominio (todos los usuarios están bajo una simple autoridad y confían en un servidor común)[14].

El diseño de este protocolo sigue un proceso de refinamiento de pasos. Se empieza con un protocolo muy simple que es intuitivamente correcto y luego se analiza sucesivamente y se refina el mismo con la finalidad de mejorar ciertos aspectos. Preservando las propiedades de seguridad. La mayoría de los pasos de refinamiento están basados en principios de diseño conocidos [14].

La idea principal en este diseño es la separación de la distribución de llave y las funciones de autentificación mutua. Es decir, se parte el diseño del protocolo en dos componentes pequeños:

1. El diseño de un protocolo que distribuye una nueva llave de sesión a dos usuarios.
2. El diseño de un protocolo que autentifica mutuamente a 2 usuarios empleando la nueva llave de sesión distribuida. El nuevo protocolo es obtenido de la combinación de estos dos protocolos.

También se hacen otras consideraciones de diseño:

- Primero, el protocolo no debe estar basado solamente en el empleo de criptosistemas simétricos. Criptosistemas asimétricos deben ser empleados para funciones de control (por ejemplo, distribución de llaves), mientras que los criptosistemas simétricos son preferidos para el intercambio ("handshake") de llave y transferencia de datos. Esto en parte, por que los criptosistemas asimétricos generalmente permiten manejos de llaves más fáciles y requieren menos confianza, y en parte, por que el resto de esta arquitectura de seguridad (por ejemplo, el protocolo de "login") hace uso de criptosistemas asimétricos.

- Segundo, el protocolo debe evitar el uso de timestamps; En su lugar deberán ser empleados exclusivamente nonces. (Esto elimina la necesidad de relojes sincronizados).

- Tercero, el protocolo debe ser simétrico con respecto a los usuarios participantes, en otras palabras, los requerimientos del procesamiento de ambas partes son similares, por lo tanto, sus roles pueden ser intercambiables fácilmente.

[14] El caso de inter-dominios, se trata en el capitulo siguiente, dentro de las propuestas de los protocolos trabajados para esta tesis.

Notación

Para la descripción de esta parte utilizamos la siguiente notación (muy similar a la antes ya utilizada):

- Los usuarios son denotados por letras mayúsculas, por ejemplo, P y Q.
- La llave compartida entre P y Q es denotada por k_{PQ}
- Las llaves publica y privada de P son denotadas respectivamente por k_P y k_P^{-1}.
- La concatenación de mensajes m y m´ es denotada por m, m´.
- La encripción de un mensaje m por una llave k es denotado por $\{m\}_k$
- El protocolo es presentado como una secuencia de pasos de protocolo, cada paso de protocolos escrito en la forma: "P → Q: m" representa la comunicación del mensaje m desde P hacia Q. o en la forma "P: acción" lo cual representa una acción interna de P, como ya sé había definido previamente.

Distribución de llave

Se comienza con el siguiente protocolo de distribución simple, a ser referido como Π_1:

(1) S : genera "s" secreto
(2) S → P: $\{\{P,T, s\}k_S^{-1}\}k_P$

S es un servidor (un principal fijo) cuya única función es generar secretos y distribuirlos a varios usuarios en el sistema. T es un time stamp.

La corrección de Π_1 puede ser informalmente argumentada como sigue:
Asumiendo relojes sincronizados, T provee una garantía de "oportunidad". La encripción interna por la llave privada de S, certifica que "s" verdaderamente viene desde S (esto es, autenticidad de origen), asumiendo que la llave publica de S es conocida por todos los principales en el sistema. La encripción exterior por llave publica de P asegura la secrecia de "s". Finalmente la inclusión del nombre P, liga P con s, de este modo se hace explicito al destinatario recibir s.

El problema de la distribución de llave es simplemente una forma especializada de distribución secreta, en la cual un secreto (esto es, una llave de sesión nueva) es distribuida para un par de principales. De este modo, Π_1 puede ser refinado para el siguiente protocolo de distribución de llave, Π_2:

(1) S : generar llave k
(2a) S → P : $\{\{P, Q, T, k\}k_S^{-1}\}k_P$
(2b) S → Q : $\{\{P, Q, T, k\}k_S^{-1}\}k_Q$

La segunda y tercera líneas reflejan que el orden de sus ejecuciones es insignificante. Hay que notar que cada llave k es generada para un par especifico de principales.

Π_2 es insatisfactoria en dos sentidos:

1). La distribución es iniciada por S en lugar de P ó Q. por lo que, P y Q talvez tengan que esperar indefinidamente antes de que la comunicación pueda comenzar.

2). El empleo de un timestamp requiere sincronización de relojes. Esto se arregla en Π_3. Específicamente, un paso de reto-respuesta (challenge/response) con un nonce empleado en lugar de timestamps. Hay que notar que cada principal genera su propio nonce.

(1a) P : generación de nonce, n_P
(1b) Q : generación de nonce, n_Q
(2a) P → S : P, n_P, Q
(2b) Q → S : Q, n_Q, P
(3) S : generación de llave k
(4a) S → P : $\{\{P, Q\, n_P, k\}k_S^{-1}\}k_P$
(4b) S → Q : $\{\{P, Q\, n_Q, k\}k_S^{-1}\}k_Q$

Aunque Π_3 permite a cualquiera de los principales iniciar, esto no elimina el problema completamente. Específicamente, las peticiones de P y Q son asíncronas, y S responde solamente después de que esta ha recibido ambas peticiones, por lo que es posible para cualquiera de las partes esperar indefinidamente.

Se puede remediar fácilmente esto, teniendo la solicitud desde un principal reenviándose a través del otro principal, como se muestra abajo en Π_K:

(K_1) P : generación de nonce, n_P
(K_2) P → Q : P, n_P
(K_3) Q : generación de nonce, n_Q
(K_4) Q → S : P, n_P, Q, n_Q
(K_5) S : generación de llave k
(K_6) S → Q : $\{\{P, n_P, Q, n_Q, k\}k_S^{-1}\}k_Q$
(K_7) Q → P : $\{\{P, n_P, Q, n_Q, k\}k_S^{-1}\}k_P$

Estrictamente hablando, la línea (K_7) no es un paso de reenvío. Q tiene que Desencriptar el mensaje recibido en el flujo (K_6) y reencriptar el resultado con la llave publica de P antes de que este pueda ser enviado en la línea (K_7). Esto es sin embargo "equivalente" a un simple reenvío si la llave publica de P es consistentemente conocida por ambos S y Q.

Autentificación mutua

El protocolo de distribución de llave en la sub-sección anterior, establece una nueva llave de sesión secreta entre dos principales. Esta llave puede ser empleada para autentificar mutuamente a los dos principales (key hanshake).
A continuación se muestra un protocolo de autentificación mutua de dos partes directo equitativamente. Nos referiremos a este como Π_M. Hay que notar que Π_M asume que una llave k ya esta compartida entre las partes.

(M1) P : generar el nonce n_P
(M2) P → Q : P, n_P
(M3) Q : generar el nonce n_Q
(M4) Q → P : $\{n_P, n_Q\}_k$
(M5) P → Q : $\{n_Q\}_k$

Es posible que parezca que Π_M es susceptible a ataques de "interleaving" o interpolación [15]. Sin embargo, no es el caso, ya que se asume que k es de nuevo u otra vez distribuida solo cada momento antes de un intercambio de autentificación. Por lo que, diferente del típico escenario de una autentificación mutua de dos-partes usual, la autentificación nunca es presentada nuevamente empleando la misma llave k.

Un numero considerable de otros protocolos de autentificación mutua de dos-partes han sido propuestos en la literatura [15]. Muchos de estos son convenientes o apropiados para los requerimientos aquí demandados. Se prefiere el que sé ha comentado en el párrafo de arriba, por su simplicidad. En cualquier caso, con la propuesta de diseño compuesta, esto es relativamente fácil para poder sustituir otro protocolo en lugar de Π_M.

Protocolo combinado

Se obtiene el protocolo final por la composición de los 2 sub-protocolos Π_K y Π_M. El protocolo resultante, denotado por Π_{KM} se muestra a continuación.

(KM1) P : generar el nonce n_P
(KM2) P → Q: P, n_P
(KM3) Q : generar el nonce n_Q
(KM4) Q → S: P, n_P, Q, n_Q
(KM5) S : General la llave k
(KM6) S → Q: $\{\{P, n_P, Q, n_Q, k\}k_S^{-1}\}k_Q$
(KM7) Q → P: $\{\{P, n_P, Q, n_Q, k\}k_S^{-1}\}k_P$, $\{n_P, n_Q\}_k$
(KM8) P → Q: $\{n_Q\}_k$

La composición es como sigue:

Los nonces, n_P y n_Q generados en Π_K pueden ser reutilizados en Π_M, de este modo, los pasos (M1) a (M3) pueden ser combinados con los pasos (K1) a (K3) en Π_K y por lo tanto eliminado. El mensaje enviado en el paso (M4) de Π_M puede ser llevado ("piggybacked") dentro del mensaje en el paso (K7) de Π_K. En el resultado, nos referiremos a P como el iniciador y Q como el receptor. (el que responde)

Además de la satisfacción de los requerimientos del diseño anterior, el protocolo Π_{KM} es interesante apreciarlo desde otro punto de vista; este puede ser visto como una extensión "segura" del ordinario "three way handshake" empleado en el establecimiento de la conexión de TCP. Específicamente, los pasos (KM2), (KM7) y (KM8) corresponden a los 3 pasos del "three way handshake" en (KM2) el iniciador comunica su numero de secuencia (nonce n_P) hacia el receptor. En (KM7) el que contesta reconoce el numero de secuencia del iniciador como bueno al reenviarle su propio (nonce n_Q). Y finalmente, el iniciador reconoce él numero de secuencia del receptor en (KM8.). La encripción requerida para (KM2), (KM7) y (KM8) junto con los mensajes extra para S representan el costo de agregar seguridad al "three way handshake"

2.2.2 ESPECIFICACIÓN Y VERIFICACIÓN DEL PROTOCOLO

Existen tres tareas en la verificación del protocolo:

(1) Especificar el protocolo con una notación que tenga una semántica definida precisamente.
(2) Formalizar los objetivos de alto nivel del protocolo, esto es, como propiedades correctas en la forma de afirmaciones con semántica bien-definida.
(3) Demostrar que las afirmaciones de propiedades correctas son satisfechas por las especificaciones del protocolo basado sobre una buena noción definida de la satisfacción.

En este caso se emplea la metodología de verificación propuesta por Woo y Lam [16,17] (ver anexo A) la cuál fue diseñada en un nivel de abstracción relativamente cerrado para la implementación del protocolo. En otras palabras el vacío o brecha semántica entre los protocolos formales verificados y los protocolos implementados existentes es pequeño y puede ser intuitivamente justificada (como opuesto a la mayoría de las propuestas lógicas, tales como la lógica de BAN cuyos pasos de idealización podrían introducir una brecha semántica grande.).

Revisión de la metodología

Típicamente las metas de alto nivel de un protocolo de autentificación son:

- Autentificación. Para cada principal participante, sobre una terminación exitosa de sus ejecuciones del protocolo, debe asegurarse que esta "hablando" con el principal que se tiene en mente.
- Distribución de llave. La nueva llave de sesión distribuida debe a lo más, ser conocida por los principales que están destinados a esta.

En la metodología de Woo y Lam las metas son formalizadas empleando dos tipos de propiedades de corrección llamadas, correspondencia y secrecia:

Correspondencia. Informalmente, especifica que diferentes principales en un protocolo de autentificación, ejecutan el protocolo en una modalidad de pasos-cerrados.
Secrecia. Especifica que cierta información (por ejemplo, llaves privadas, nuevas llaves de sesión) no deben ser accesibles para un intruso. La secrecia dirige la meta de distribución de llave.

Especificación del protocolo

Las condiciones iniciales para la especificación del protocolo de Π_{KM} , se da de la siguiente manera:

(CI1) $\forall x, y$: x tiene y $\wedge$ x tiene k_y
(CI2) $\forall x, y$: x tiene $k_y^{-1} \Leftrightarrow x = y$
(CI3) **Z** tiene **X** $\Leftrightarrow$ [**X** $\in$ SYS $\vee \exists x$: **X** = $k_x \vee$ **X** = k_Z^{-1}]

Protocolo del iniciador (i):

(I1) Comienzo-ini (r)
(I2) Nuevo-nonce (n)
(I3) Envía (r, [i,n])
(I4) Recibe (r, [$\{\{i,n,r,\mathbf{N},\mathbf{K}\}_k{}^{-1}{}_s\}_{ki}$, $\{n, \mathbf{N}\}_{\mathbf{K}}$])
(I5) Envía (r, [$\{\mathbf{N}\}_{\mathbf{K}}$])
(I6) Acepta (**K**)
(I7) Fin-ini (r)

Protocolo del receptor (r):

(R1) Comienzo-Resp (i)
(R2) Recibe (i, [i, **N**])
(R3) Nuevo-nonce (n)
(R4) Envía (S, [i, **N**, r,n])
(R5) Recibe (S, [$\{\{i, \mathbf{N}, r, n, \mathbf{K}\}_k{}^{-1}{}_s\}_{kr}$])
(R6) Envía (i, [$\{\{i, \mathbf{N}, r, n, \mathbf{K}\}_k{}^{-1}{}_s\}_{ki}$, $\{n, \mathbf{N}\}_{\mathbf{K}}$])
(R7) Recibe (I, [$\{n\}_{\mathbf{K}}$])
(R8) Acepta (**K**)
(R9) Fin-Resp (i)

Protocolo de servidor (S):

(S1) Recibe (r, [i, n, r, n′])
(S2) Nuevo-secreto ({i,r}, k)
(S3) Envía (r, [$\{\{i, n, r, n', k\}_k{}^{-1}{}_s\}_{kr}$])

La correspondiente explicación es:

SYS denota al grupo de todos los usuarios (principales) en el sistema). En particular, Z, denota un principal distinguido que representa al intruso y S, un principal que representa el servidor fijo, pertenecientes al SYS.
Las variables en minúsculas (por ejemplo, i, r, n) especifican la gama de términos primitivos.
Además, i,r,p se asumen para extenderse sobre nombres en SYS – {S,Z}, mientras que x, y sobre SYS. Las variables en negritas (por ejemplo, **N**) especifican los términos no nulos arbitrarios sobre la gama.

CI1, especifica que todos los principales conocen los nombres de cada principal (incluyendo el propio) y su llave publica. CI2, especifica que la llave privada de un principal es solo conocida por él. CI3, especifica que Z conoce precisamente los términos permitidos bajo CI1 y CI2.

Especificación de corrección

La meta de autentificación mutua es formalizada utilizando las dos afirmaciones de correspondencia mostradas abajo.
Donde (3), especifica que siempre que un iniciador (i) termina la ejecución de su protocolo local, deberá ser el caso en que el receptor (r) tenido por objetivo esta teniendo parte en el intercambio. (4), especifica una propiedad similar, pero con respecto al receptor.

(i, Fin-ini (r)) $\hookrightarrow$ (r, Comienza-Resp (i)) .....(3)
(r, Fin-Resp (i)) $\hookrightarrow$ (i, Comienza-ini (r)) .....(4)

Sumado a la Condición de Secrecia General (GSC), se especifica la siguiente afirmación de secrecia (5), la cuál dice que la llave privada de cada principal con excepción de Z no se debe aprender por ningún otro principal como resultado de ejecutar el protocolo. Se exceptúa Z porque no es limitado por ningún protocolo y está libre divulgar su propia llave privada

$\forall x, p$: x tiene $k_p^{-1} \Leftrightarrow x = p$ .....(5)

La especificación de corrección final es: ({(3), (4)}, {(5)})

Resumen de la prueba

Primero se prueba que Π_{KM} satisface (5). Informalmente, se puede ver que (5) es satisfecha por que las llaves privadas de los principales no-servidores (excepto Z) son únicamente utilizadas internamente para Desencriptar mensajes entrantes; ellos nunca se utilizaron en mensajes de salida. Para S, su llave privada es solamente utilizada en la encripción de mensajes y nunca es transmitida como un componente en el mensaje.

Lema 1.

Π_{KM} satisface $\forall$ i, r, n, n′, k : $\neg$ (Z tiene $\{i, n, r, n', k\}_k^{-1}{}_s$).

Este lema dice que Z nunca puede aprender un termino de la forma $\{i, n, r, n', k\}_k^{-1}{}_s$ Por que, todos estos términos son siempre bajo la encripción de cualquier k_i ó k_r y por lo tanto no puede ser recuperado por Z.

Lema 2.

(x, r, Comm ([{{i, n′, r, n, k}$_k^{-1}{}_s$}$_{kr}$])) $\hookrightarrow$ (S, y, Comm([{{i, n′, r, n, k}$_k^{-1}{}_s$}$_{kr}$]))

El lema 2 dice que cualquier mensaje recibido por el receptor en el paso R5 debe ser originalmente enviado por S en el paso S3.

Lema 3.

(x, i, Comm ([{{i, n, r, n′, k}$_k^{-1}{}_s$}$_{ki}$, {n, n′,}$_k$])) $\hookrightarrow$
(r, y, Comm ([{{i, n, r, n′, k}$_k^{-1}{}_s$}$_{ki}$, {n, n′,}$_k$]))

El lema 3 es similar al anterior, dice que cualquier mensaje recibido en el paso I4 debe ser originalmente enviado por el receptor en el paso R6.

Lema 4. (x, r, Comm ([{n}$_k$])) $\hookrightarrow$ (i, y, Comm([{n}$_k$]))

El lema 4 dice que cualquier mensaje recibido en el paso R7 debe ser originalmente enviado por el iniciador en el paso I5.
Utilizando los lemas anteriores se puede probar que Π_{KM} satisface la condición de secrecia general (GSC), (3) y (4). Así se concluye que el protocolo es correcto.

Proposición.

Π_{KM} es correcto con respecto a ({(3), (4)}, {(5)}).

2.2.3 IMPLEMENTACIÓN DEL PROTOCOLO

El protocolo Π_{KM} adopta el estándar GSS-API [18]. Generic Security Service Application Program Interface (GSS-API) es una implementación de interfase independiente a través de la cual los servicios de seguridad son proporcionados para visitantes. Considerando el protocolo Π_{KM} el control de flujo básico del iniciador y el receptor es descrito en la siguiente figura 2.2

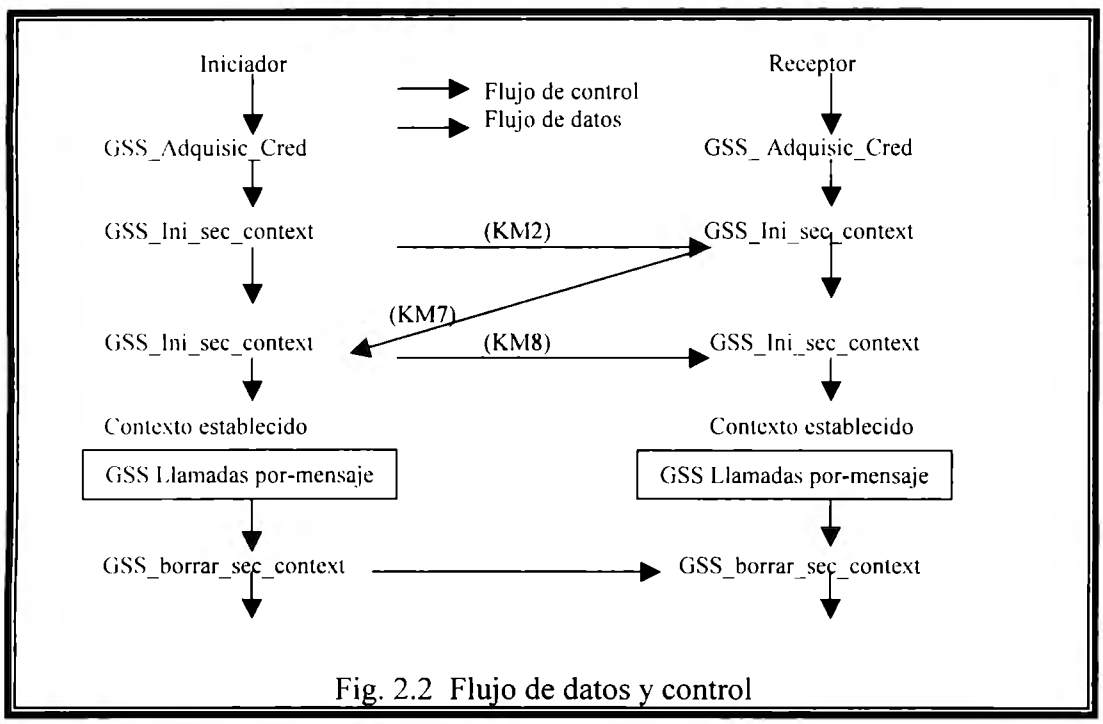


Fig. 2.2 Flujo de datos y control

Comentarios

Como se pudo observar esta es una manera de desarrollar de forma sistemática un protocolo de autentificación realista, específicamente tomando en cuenta el diseño, especificación, verificación e implementación. Aunque el proceso es informal, es útil en el diseño de protocolos de autentificación.

2.3 USO METÓDICO DE TRANSFORMACIONES CRIPTOGRÁFICAS EN PROTOCOLOS DE AUTENTIFICACIÓN

Como hemos visto, el diseño de protocolos criptográficos para autentificación y manejo de llaves se sabe que es un problema difícil. Aún cuando muchas investigaciones se han dedicado a las técnicas de análisis, sigue siendo un problema la carencia de diseño desde la base. Ahora identificaremos otro método común de diseño de protocolos, el cual contribuye en un número de alternativas o caminos para los problemas de los protocolos. Esta es la practica de encriptamiento y relevantes campos utilizando una transformación criptográfica reversible. Un nuevo principio de diseño y una notación complementaria ha sido analizado en el trabajo presentado por [19], lo cual ayuda a los diseñadores de protocolos a identificar que forma de encripción es realmente requerida. Aquí solo presentamos algunos ejemplos más, los cuáles son utilizados para ilustrar los métodos de desencripción aplicados ampliamente para la autentificación y para constatar como el principio de diseño y notación debe ser empleado en la practica. En esta investigación se revelan varios problemas presentes en los protocolos de autentificación tratados, además se demuestra como rescribir protocolos sin un uso innecesario de desencripción. Y se llega a una especificación de protocolo mas completa, para mayor detalle ver [19]. De acuerdo con lo comentado sé deriva la pregunta de la siguiente sección.

¿ Desencriptar o no desencriptar?

Un ejemplo simple. Considerar la autentificación de "login" en una computadora compartida (o incluso en un cajero automático de banco). Ahora, asúmase que el usuario Alicia quiere hacer login. Abajo tenemos un protocolo imaginario con el cual el host autentifica la autenticidad o genuinidad del usuario Alicia.

(a) Alicia envía a la computadora su password como es solicitado.
(b) El host encuentra la forma encriptada del password de Alicia guardado en la computadora y decripta este.
(c) El host compara el resultado con el valor dado por Alicia.

El protocolo en el paso (b) demanda al host ejecutar la desencripción por que el password del usuario debe estar guardado en forma encriptada. Guardando passwords en forma de plaintext y utilizando los mecanismos de protección de archivo de los sistemas operativos solo para prevenir accesos no es un método seguro ya que, por ejemplo, una cinta de respaldo puede fácilmente ser leída en otra computadora. Aun con la medida de protección especificada este protocolo no es seguro todavía ya que el sistema necesita una llave criptográfica. Entonces el problema de proteger esta llave es el mismo problema de cómo proteger "passwords" guardados en forma de "plaintext". Needham sugiere una forma diferente para proteger passwords, ilustrado por el protocolo siguiente.

1) Alicia envía a la computadora su password como es solicitado.
2) El host encripta el valor dado por Alicia.
3) El host compara el resultado con el valor guardado en la computadora.

Este protocolo no requiere del host para realizar una desencripción. La encripción ejecutada en el paso 2 puede simplemente utilizar un password como llave para encriptar una constante. El

segundo protocolo es claramente mejor que el primero. No hay secreto a ser protegido, en contraste, el secreto convenido con el primer protocolo no puede de hecho ser protegido. Verdaderamente como será visto en el resto del documento, una entidad de autentificación puede ser mejor alcanzada sin depender de pasar ningún secreto. (hay que notar que la distribución de llave secreta es generalmente un objetivo).

2.3.1 DESENCRIPCIÓN: MÉTODO APLICADO AMPLIAMENTE PARA AUTENTIFICACIÓN

En esta sección se listan otros de los métodos de autentificación publicados y varios protocolos bien conocidos para demostrar el uso de la desencripción para la autentificación. Fue mostrado por Gong que el típico mecanismo de challenge/response para autentificación basado en el uso de llaves simétricas puede ser resumido dentro de cuatro clases como se muestra:

2.3.2 Resúmenes de protocolos

Uso 0

$A \rightarrow B: \{T_A\}_k$

Uso 1

1. $A \rightarrow B: N_A$
2. $B \rightarrow A: \{N_A\}_k$

Uso 2

1. $A \rightarrow B: \{N_A\}_k$
2. $B \rightarrow A: N_A$

Uso 3

1. $A \rightarrow B: \{N_A\}_k$
2. $B \rightarrow A: \{f(N_A)\}_k$

Para estos casos también T_x es un timestamp y N_x es un número aleatorio o nonce generado por el principal X.

A continuación se listan varios de estos protocolos bien conocidos de manera concreta, los cuales hacen uso de la desencripción. Las amenazas de identificadores frescos en los protocolos caen dentro de los cuatro usos mencionados anteriormente.

El protocolo de Otway-Rees [20]

1. $A \rightarrow B$: M, A, B, $\{N_A, M, A, B\}k_{AS}$
2. $B \rightarrow S$: M, A, B, $\{N_A, M, A, B\}k_{AS}$, $\{N_B, M, A, B\}k_{BS}$
3. $S \rightarrow B$: M, $\{N_A, K_{AB}\}k_{AS}$, $\{N_B, K_{AB}\}k_{BS}$
4. $B \rightarrow A$: M, $\{N_A, K_{AB}\}k_{AS}$

Este protocolo envuelve dos autentificaciones de un solo sentido para los dos clientes; Alicia y Bob, respectivamente, con el servidor S de autentificación de confianza. Aquí los nonces N_A y N_B son tratados como si ellos fueran secretos. Los receptores tienen que presentar desencripción para recuperar estos. Se podría considerar este un tratamiento de nonce para ser de uso 3.

Protocolo de Kerberos [21, 22]

1. $A \rightarrow S$: A,B
2. $S \rightarrow A$: $\{K_{AB}, B, T_S, L, \{K_{AB}, A, T_S, L\}k_{BS}\}k_{AS}$
3. $A \rightarrow B$: $\{K_{AB}, A, T_S, L\}k_{BS}$, $\{A, T_A\}k_{AB}$
4. $B \rightarrow A$: $\{T_A+1\}k_{AB}$

Esta es una especificación simplificada del protocolo el cual es bosquejado desde la versión (V5) [12]. L es un tiempo de vida que marca el tiempo de expiración de la llave de distribución K_{AB}. El tratamiento del T_S en el protocolo es de uso 0 y T_A es de uso 3. (un timestamp puede ser pensado como un nonce).

Protocolo de Yahalom [23]

1. $A \rightarrow B$: A, N_A
2. $B \rightarrow S$: B, N_B, $\{A, N_A\}k_{BS}$
3. $S \rightarrow A$: N_B, $\{B, K_{AB}, N_A\}k_{AS}$, $\{A, K_{AB}, N_B\}k_{BS}$
4. $A \rightarrow B$: $\{A, K_{AB}, N_B\}k_{BS}$, $\{N_B\}k_{AB}$

El tratamiento de los nonces en este protocolo no cae exactamente dentro de ningún uso de los resúmenes de protocolo. Sin embargo, los dos nonces son enviados en plaintext por sus respectivos iniciadores, quienes pueden considerar que los nonces son enviados fuera del servidor. Desde este punto de vista se podrían considerar el tratamiento de estos nonces para ser de uso 1 donde la interpretación de desencripción aplica.

Protocolo en el documento ISO/IEC CD 11770.

El siguiente protocolo es una versión simplificada del "mecanismo de establecimiento de llave 6" en ISO / IEC JTC1/SC27 del documento N832 [24] (Debe hacerse notar que este documento no tuvo el estatus de estándar y había sido cambiado en la versión de estandarización final.).

1. $A \rightarrow S$: N_A, B
2. $S \rightarrow A$: $\{N_A, K_{AB}, B\}\ k_{AS}$, $MAC_{KAB}(\{N_A, K_{AB}, B\}k_{AS})$, $\{T_S, K_{AB}, A\}k_{BS}$, $MAC_{KAB}(\{T_S, K_{AB}, A\}k_{BS})$
3. $A \rightarrow B$: $\{T_S, K_{AB}, A\}k_{BS}$, $MAC_{KAB}(\{T_S, K_{AB}, A\}k_{BS})$
4. $B \rightarrow A$: $\{N_B, A\ \}k_{AB}$

Este protocolo esta basado en el protocolo de Needham-Schroeder [25], pero hace uso de time stamps para remover un ataque propio de Denning y Sacco [26]. El tratamiento del timestamp T_S es de uso 0 y del nonce N_A es uso 1. después de que sé complete la corrida del protocolo, una sesión de comunicación entre Alicia y Bob subsecuentemente tomara lugar la cual le asegura a Bob que Alicia ha desencriptado el mensaje de la línea 4 utilizando la nueva llave de sesión. Por lo que el tratamiento del nonce N_B puede ser considerado esencialmente como uso 2.

2.4 AUTENTIFICACIÓN Y AUTORIZACIÓN EN AMBIENTES MÓVILES

En servicios ofrecidos sobre redes de información, como actividades bancarias electrónicas, validar la identidad de un usuario y las autorizaciones que este tiene son fundamentales.

Existen varios caminos para realizar estas operaciones, algunas de las cuales proporcionan un alto grado de certeza y son fáciles de utilizar comparadas con otras. Las soluciones basadas en Infraestructura de Llave Publica (PKI) son generalmente consideradas las más seguras y confiables.

En un ambiente móvil, donde los mismos servicios pueden ser utilizados a través de diferentes canales, como WEB o WAP, la emisión de autentificación y autorización es a menudo más compleja. Por ejemplo, en una PKI el manejo de la llave privada de manera que pueda ser utilizada, sin ser comprometida, en diversas plataformas de dispositivos, es un desafío.

La estandarización de tecnologías a ser utilizadas para resolver el problema esta avanzando a un paso constante y rápido. Las implementaciones actuales se están retrasando un paso detrás, especialmente cuando vienen a desarrollar las soluciones totales para autentificación y autorización.

Autentificando la identidad y la autorización de los usuarios con un alto grado de certeza en ambientes no fiables y abiertos ha sido uno de los problemas funcionales más significativos con los cuales se ha luchado, cuando los servicios se convierten accesibles sobre el Internet. Muchos esquemas diferentes para lograr el resultado deseado han sido ideados. Algunos de las cuales han sido mejor que otros en términos de lo amigables con los usuarios y del nivel de certeza y de la seguridad.

Ahora que está amaneciendo la era de Internet móvil hay varias revisiones de los procedimientos que se utilizan para la autentificación y la autorización actuales. El dilema al que nos enfrentamos es que ahí debería existir un mecanismo que esté de la manera más conveniente para los usuarios y que daría una certeza perfecta del resultado de la autentificación.

2.4.1 ¿CUAL ES EL ALCANCE?

Ahora discutiremos los métodos actuales para autentificar y autorizar a usuarios en el ambiente móvil. No solamente nos concentraremos en las aplicaciones inalámbricas, sino que abordaremos una definición propia del ambiente móvil que se utiliza, como se comenta en la sección 2.4.2.3. La discusión se centra en soluciones basadas PKI, aunque también otros medios de

autentificación se presentan. La necesidad real de la autentificación y de la autorización también se discute de manera abreviada, pues la materia no parece siempre estar totalmente clara. La discusión de criptografía de llave-pública se hace en un nivel muy superficial en este momento ya que esta ya ha sido tratada.

En algunas áreas no es sensible pegarse a describir cuál metodología es actualmente posible, pues la tecnología se está desarrollando muy rápidamente. En este apartado se discute en parte de la tecnología que no está aún extensamente disponible comercialmente. El Módulo de Identidad Inalámbrico (WIM) es un ejemplo de tal equipo.

El objetivo de esta sección es clarificar los conceptos de la autentificación y de la autorización, así como otros conceptos importantes que se derivan de estas, además de discutir cómo los conceptos y los procedimientos relacionados con ellas se pueden poner en ejecución o realizar en el ambiente móvil. En él capitulo siguiente también se harán propuestas para nuevos, procedimientos de autentificación aplicable de manera general que confían en las tecnologías discutidas en el mismo.

2.4.2 DEFINICIONES

Para las siglas utilizadas en este apartado ver apéndice.

2.4.2.1 Autentificación

"La autentificación de la identidad es el proceso por el que un cierto atributo elegido de una entidad del mundo real (el carácter que lo distingue o la personalidad de un individuo) se demuestra pertenecer a esa entidad." [27] Es decir, la autentificación es el proceso de determinar si alguien o algo es quien se demanda o se declara ser.

Hay cinco métodos de autentificar la identidad de un principal [27]:

1. Algo que el demandante sabe
2. Algo el demandante posee
3. Algo el demandante es
4. El demandante está en un lugar determinado (en un tiempo determinado)
5. La autentificación es establecida por terceras partes de confianza (Trusted third party)

2.4.2.2 Autorización

La autorización esta relacionada estrechamente con la autentificación. La autorización es de hecho el proceso de dar a alguien el permiso de hacer o de tener algo. Esto implica que para poder dar a alguien acceso a algo, este alguien necesita primero ser autentificado. En otras palabras la autorización es precedida lógicamente por la autentificación. La autentificación real de la identidad de una entidad o de un individuo, sin embargo, no se requiere siempre, para hacer la autorización, si la autorización se puede validar por algunos otros medios (véase la sección 2.1.4.2.6.).

En este sentido la autentificación por sí misma sé podría decir que tiene poco uso; y sería justamente el medio de conseguir la información necesitada para la autorización.

2.4.2.3 Ambiente móvil

En el contexto de esta tesis el concepto "ambiente móvil" se utiliza para describir una configuración donde un usuario del sistema, que tiene acceso por Internet o por otras redes abiertas, no se limita a un cierto lugar, dispositivo o canal de acceso para utilizar servicios proporcionados por el sistema. Un ejemplo de una configuración como esta, es un sistema de actividades bancarias que es accesible a través de los canales de acceso WAP y el Internet.

2.1.4.2.4 Infraestructura de Llave Pública y Criptografía de llave-pública

(PKI) es "un sistema para publicar los valores de llave-pública utilizados en criptografía de llave-pública" [28].
La criptografía de llave-pública es una técnica introducida inicialmente por Diffie y Hellman en 1976 [29]. La criptografía de llave-pública confía en los problemas matemáticamente complejos que se ocupan de la teoría de números primos, que son imposibles de solucionar en un tiempo razonable sin la entrada de información correcta (la llave.) La criptografía de llave-pública se utiliza esencialmente, para las secuencias que encriptan y desencriptan datos. En la criptografía de llave-pública hay fundamentalmente dos llaves (un par de llaves) implicadas: una llave que es mantenida secreta por su propietario (la llave privada) y una llave que se hace pública (la llave pública) entre las cuales no hay conexión que pueda descubrir la llave privada. Los datos que son encriptados utilizando la llave pública pueden ser desencriptados solamente utilizando la llave privada y viceversa. En esta sección no se discutirá específicamente la criptografía de llave-pública y los algoritmos relacionados con ella a un nivel más alto de detalle que esto. Se hace la suposición de que los algoritmos criptográficos que se están utilizando son fuertes y las llaves que están en uso son suficientemente grandes para que el proceso sea confiable.

Aunque, una PKI no es totalmente un concepto nuevo (véase [29]) a partir de 1976 de una u otra forma ha estado extensamente en uso en casi todo el mundo, y todavía las discusiones continúan alrededor del tema. Generalmente por el hecho de que las autoridades de gobierno, especialmente en los Estados Unidos desean tener acceso a las llaves privadas de todos los individuos para "los propósitos de la seguridad nacional." Resulta un tema de inquietud e interés mundial [30]. La discusión adicional al respecto está fuera del alcance de esta tesis.

Actualmente existe un número de acercamientos levemente diversos tomados para implementar PKIs pero hay por lo menos dos cosas que son comunes a todos ellos:

- **Certificación**, la cual es el proceso de dar un valor de llave-pública a un individuo o atributo.
- **Validación,** la cual es el proceso de verificar la validez del anterior.

En él articulo "Una inspección de la PKI" [28], Branchaud define la certificación como "un medio por el que valores de llave-pública, e información perteneciente a ese valor, sean

publicados" y el certificado para ser " La forma en que un PKI comunica valores de llave pública o información sobre llaves públicas, o ambos".

Un tema de discusión serio que se solucionará en una PKI y la implantación del sistema criptográfico de llave pública, es la administración de las llaves privadas de los individuos. Las llaves deben mantenerse en secreto, pero el propietario de la llave debe siempre tener acceso a ella.

2.4.2.5 Autoridades de Certificación y Certificados Digitales

La Autoridad Certificadora (CA) es una entidad que publica y verifica certificados digitales. El certificado digital es una cadena de datos, que contiene una llave pública y sus atributos, tales como; la información sobre su propietario y la firma de una CA. El hecho de que el certificado es firmado por una CA, nos da ciertas garantías, y nos permite fácilmente ganar la certeza de que la información contenida en el certificado es válida. Esto es posible puesto que las CA´s son entidades confiables y la firma de la CA hace también valido al certificado.

En una PKI también existe normalmente una red o una jerarquía de CAs [28, 31]. Esto hace posible, que una CA lejana sea validada por alguna otra CA que sea validada por la raíz CA (que es confiable implícitamente) o una CA en el camino a la raíz, que sea CA confiable también.

Debe observarse que cuando una CA publica un certificado firmándolo digitalmente, tiene que estar seguro que la llave pública y la información que se está asociando a ella en el certificado es correcta, esto es, la información sobre el subscriptor del certificado es correcta y coherente. La CA confía normalmente en una autoridad de registro (RA) para hacer esto.

2.4.2.6 SPKI – Infraestructura de Llave Pública Simple

El SPKI se ha definido en el RFC 2693 [32] del IETF. La idea de los certificados de SPKI es conectar una autoridad a una llave. Esto difiere del PKI tradicional, donde la información de la identificación se conecta a una llave en un certificado. En otras palabras los certificados de SPKI se centran más en los derechos obligatorios y autoridades para las llaves que la información de identificación de los individuos para las llaves, que es el caso interno con implementaciones PKI convencional como el X.509 [31,32, 33].

Para los propósitos de ambientes abiertos esto es conveniente, puesto que en esta configuración un grado de anonimato es proporcionado - un usuario puede presentar un certificado que lo publique una autoridad para tener acceso a algunos servicios o recursos, sin necesariamente revelar su identidad. La confianza se basa en una red de confianza y autorización delegada.

Aunque es un método ideal para proporcionar mecanismos de autentificación y de autorización en los ambientes abiertos, no hay ninguna implementación o bien, los productos que lo soporten en los mercados actualmente [33].

2.4.2.7 X.509 PKI

X.509 es el estándar de PKI utilizado más extensamente. Define todos los aspectos de la infraestructura, como las estructuras de certificados y los lazos jerárquicos entre las CAs.

Los certificados X.509 tienen el contenido siguiente [33]:

Versión que describe la versión del certificado codificado
Número de serie que es un número de identificación único
Identificador de algoritmo de firma de CAs
Nombre del emisor que identifica la CA quien firmó y emitió el certificado
Validez que indica el periodo de validez del certificado
Sujeto identificando el poseedor de la llave privada
Información de llave pública del sujeto
Identificador único del emisor
Identificador único del sujeto
Campo(s) de extensión

Es útil entender la estructura de los certificados X.509, pues se utilizan también en la capa de seguridad de WAP WTLS (Seguridad Inalámbrica de la Capa de Transporte). Que se discute más adelante.

Hay varias versiones del estándar X.509. Las últimas versiones se especifican con "vX" por ejemplo: X.509v3.

2.4.3. LA NECESIDAD DE LA AUTENTIFICACIÓN Y DE LA AUTORIZACIÓN

La autentificación de la identidad de una persona o de una entidad tiene que ser hecha, siempre que se requiera, para saber explícitamente quién es la persona o la entidad realmente. La autorización tiene que ser hecha, siempre que una decisión tiene que ser tomada, si o no, alguien tiene permiso de hacer o tener algo. Por ejemplo, cuando una persona visita un banco, para retirar el dinero de su cuenta el empleado del banco tiene primero que autentificar a la persona, (podría ser; su identificación firmada legalmente.) Luego se tiene que revisar el balance de la cuenta de la persona para saber cuánto dinero le autorizan retirar. En este ejemplo se requieren tanto la autentificación, como la autorización. Los billetes de banco se pueden utilizar como símbolos de pagos (autorizan las transacciones del pago.). Al pagar con los billetes de banco la identidad de la persona que los usa no tiene que ser ampliamente autentificada. Los billetes de banco son firmados por el gobierno para certificar su autoridad (es decir la denominación).

Esto es un arreglo conveniente para todas las partes;

1) El consumidor tiene derecho a la privacidad - nadie sabrá necesariamente en que gasta su dinero

2) La identidad de un consumidor no tiene que ser autentificada cada vez que él paga algo, lo cual ayuda a hacer transacciones más fáciles y rápidas.
3) Las autoridades, es decir el gobierno que publica el billete de banco no tiene que tratar con los consumidores directamente

En las redes de información los mismos principios se aplican. La autentificación de un usuario no es necesaria, si solamente la autorización se puede validar de alguna u otra manera. Incluso la autorización no es necesaria para los servicios o los recursos que son accesibles para cada uno. En teoría, si hay por lo menos una persona o entidad que no debería ser autorizada para tener acceso a un cierto servicio o recurso, la autorización de cada uno debe ser validada, antes de permitir el acceso al servicio o al recurso.

La autentificación de la identidad de una persona no es necesaria casi tan a menudo como sé podría pensar o utilizar, por ejemplo en eCommerce, hoy en día. La mayoría de los "eStores" en el Internet, te requiere autentificarte al hacer una compra. ¿Cuántos almacenes ordinarios te requieren presentar tu identificación cuando checas tus compras de libros o de la tienda de comestibles? no muchos! Actualmente lo hacen normalmente solamente si tienen que tener cierto aseguramiento de tu crédito. Esta misma práctica debe también aplicarse al "e-mundo" - todos deberían dar derecho a la privacidad, la autentificación debe ser requerida solamente cuando no hay otra manera de conseguir el aseguramiento de la credibilidad. Hay muchas iniciativas para proporcionar un "e-Money" seguro que se pueda utilizar en el Internet (para ambos; móvil y alámbrico) - serán las herramientas para permitir la autorización de pagos sin compromiso de la privacidad, como lo comentan en [34, 33]

2.4.4. MEDIOS DE LA AUTENTIFICACIÓN Y DE LA AUTORIZACIÓN

En este capitulo los métodos reales o los procesos que se utilizan para autentificar las identidades de usuario son discutidos. Todos los métodos hacen uso de una o más de las cinco características descritas en la sección 2.4.2.1

La autorización del usuario para tener acceso al servicio o recurso se puede realizar en un sistema, después que el usuario haya sido autentificado y su identidad sea resuelta a través del uso de listas de control de acceso (ACL), para determinar qué determinado usuario es autorizado a hacerlo.

La autorización está así en lo máximo, tan exacta y correcta como el proceso de la autentificación. Un mecanismo como el SPKI se podría utilizar, para evitar la autentificación del usuario, pero para proporcionar todavía una autorización confiable.

2.4.4.1 Las palabras de paso ("passwords")

Las palabras de paso asociadas a los nombres del usuario (algo que la persona sabe) son una manera simple de autentificación. Pueden tener uno-a-uno o uno-a-muchos lazos a las personas reales. Es decir cada usuario de un sistema puede tener una diversa palabra de paso, o un grupo puede tener la misma palabra de paso. Poner un sistema en ejecución donde las palabras de paso

se utilizan para autentificar a los usuarios es absolutamente directa (Aquí solo tiene que ser un depósito, que debe ser seguro, donde se salvan las palabras de paso.)

La desventaja de usar palabras de paso es que son notoriamente vulnerables a los ataques. Por lo menos los tipos siguientes de ataques son posibles:

- El acceso externo (external disclosure),
- Conjeturar (guessing),
- Escuchar las comunicaciones (communications eavesdropping),
- Ataques de reenvío (replay) y
- El compromiso del host [27].

Hay varios esquemas de autentificación que hacen uso de palabras de paso conjuntamente con algún otro factor. Una extensión simple de palabras de paso son las palabras de paso de una sola vez. Proporcionan a un usuario una lista de palabras de paso a través de un canal seguro. Él usuario utiliza solamente cada palabra de paso individual una vez para autentificarse con el sistema. El uso de este esquema elimina la posibilidad de comunicaciones que se escuchan y ataques de reenvío.

2.4.4.2 La palabra de paso con "token"

Las palabras de paso se puede utilizar conjuntamente con un cierto objeto físico (algo que la persona posee). Por ejemplo, una tarjeta de cajero automático (ATM) con un código "NIP", la tarjeta por sí sola es inútil sin el "NIP", al igual que el "NIP" sin la tarjeta.

Este concepto se ha extendido con el uso de las tarjetas de circuito integrado (ICC) o tarjeta inteligente. Estas son los dispositivos de la prueba de la interferencia que pueden obrar recíprocamente con un dispositivo directamente, al igual que el caso del "SIM-card" en los teléfonos GSM. El SIM no se puede utilizar sin el conocimiento del código del NIP, que el "SIM-card" verifica por sí mismo. El NIP relacionado con el SIM nunca se revela del SIM. Un método de desafío y respuesta se utiliza en la autenticidad del usuario.

"Palabras de paso de una sola vez sincronizadas" [27] es otra técnica similar. El usuario tiene un dispositivo de prueba de interferencia que producen los códigos que se pueden utilizar como palabras de paso, a menudo conjuntamente con un NIP y un nombre de usuario para producir un trío de autentificación. El código es producido en el dispositivo por un algoritmo conocido también por el autentificador, que de esta manera sabe qué código se ha producido esta vez y puede realizar la autentificación. Este es un esquema más elaborado que solamente usar las listas de las palabras de paso de una sola vez, puesto que el uso del dispositivo elimina la necesidad de distribuir las listas

2.4.4.3 Biométricos

Las técnicas de autentificación biométricas, incluyen el reconocimiento de la huella digital, exploración retiniana, la exploración de la geometría de la mano y el reconocimiento de voz,

principalmente [27]. Todas estas técnicas se basan en las características físicas de una persona (algo que él posee / que es.).

Puesto que, las características físicas de diversos individuos se diferencian tan solo algunas veces levemente de uno a otro, pero pueden diferir con el tiempo en un solo individuo, es en ocasiones imposible conseguir un nivel suficientemente alto de certeza en la autentificación. Además la precisión de las tecnologías disponibles es limitada y algunas de las tecnologías son muy costosas. En algunas aplicaciones estas técnicas son perfectamente convenientes, pero en ambientes abiertos siguen siendo a menudo demasiado difíciles de poner en ejecución dentro de la práctica

2.4.4.4 Firmas Digitales

Cuando una PKI se establece en un lugar, las firmas digitales se pueden utilizar para autentificar usuarios. De manera general la secuencia siguiente de acciones tiene que ser realizada para autentificar a un usuario por su firma digital:

1. Las peticiones de usuario tienen acceso al servicio o al sistema

2. El sistema genera ciertos datos para que el usuario pueda encriptar con su llave privada. Entonces los datos se envían al usuario

3. El usuario concatena los datos recibidos del sistema y de un grupo fecha/hora (time stamp) y encripta la secuencia entera. (Esta es una buena práctica que por ejemplo, un timestamp se concatene a los datos, para no poder descifrar totalmente los datos que se cifrarán por alguna parte intrusa. Esto debe evitar la posibilidad de un ataque de "Chosen plain-text" (escoger el plain-text) de acuerdo con lo descrito en [35].) Entonces los datos encriptados (el texto cifrado) son enviados de nuevo al sistema. Junto con los datos encriptados una conexión al certificado (o al certificado por sí mismo) del usuario se envía

4. El sistema desencripta la información recibida con la llave pública del usuario, y encuentra el certificado.

5. El sistema verifica que la información desencriptada esté compuesta de los datos originalmente generados y de un timestamp válido. Si esto es ACEPTABLE, el sistema ha autentificado con éxito al usuario

2.4.4.5 Propiedades de un buen mecanismo de autorización y autentificación

En esta sección se enumeran las características que un buen mecanismo de autentificación y autorización de la identidad deben poseer. Algunas de las características citadas, están en contradicción de la otra parte, pero sobre todo debe ser posible alcanzar un nivel aceptable de conformidad con cada uno de los criterios.

Corrección. Los resultados de cada caso individual de autentificación o de autorización realizada deben estar correctos. Si es posible autentificar al usuario, el resultado debería ser

siempre que cualquiera sea encontrado, que el usuario es quién él afirma o lo encuentran siendo un fraude. De acuerdo con esta autentificación perfectamente correcta es más a fondo posible autorizar al usuario a tener acceso a esos servicios y recursos que se definieron para ser accesibles para él o al grupo o a los grupos a los que él pertenece. En la práctica es imposible conseguir una certeza absoluta en la autentificación. Solamente un nivel razonable de la certeza puede ser ganado.

Posibilidad para la privacidad y el anonimato. La autentificación de la identidad debe ser hecha solamente cuando es absolutamente necesaria. La autorización siempre es posible sin que sea revelada la identidad de los usuarios, esto debe ser hecho de tal manera.

Velocidad. E l proceso de la autentificación debe ser rápido. El usuario no tendría que esperar por el resultado por más de un segundo o dos.

Resistencia al ataque. El mecanismo perfecto de la autentificación debe ser resistente contra cualesquiera de los tipos sabidos o desconocidos de ataques.

No costosos. El mecanismo no debe requerir inversiones extensas ni de los usuarios ni de los autentificadores.

Amigable al usuario. El mecanismo debe producir "overhead" pequeño al usuario como sea posible. Debe también ser fácil de utilizar y entender. En la situación óptima el usuario no tiene que realizar ninguna acción para autentificarse. El usuario no debe ser forzado a llevar cualquier equipo adicional, tarjetas magnéticas o inteligentes, lista de palabras de paso u otros objetos físicos para utilizar el sistema.

Universalidad. Debe ser posible que el usuario utilice los mismos medios o métodos de autentificación en todos los servicios y por todas partes.

2.4.5. POSIBILIDADES TÉCNICAS OFRECIDAS POR LOS DISPOSITIVOS MÓVILES ACTUALES.

Existen ya microteléfonos móviles GSM permitidos en Internet disponibles en el mercado. Él término "Internet permitido" significa que el microteléfono tiene un WAP interno. El primero de estos dispositivos que llegó a estar comercialmente disponible fue el Nokia 7110. Después de este, Ericsson introdujo el R320 y muchos otros siguieron a estos. En algunos de los modelos y de las versiones actuales los "browsers" de WAP implementan el WTLS.

2.4.5.1 Seguridad Inalámbrica de la Capa de Transporte.

El WTLS es el protocolo de la capa de seguridad para el WAP. Ha sido especificado por la iniciativa del foro de WAP. El propósito del WTLS es definido por el foro de WAP como sigue:

"La meta fundamental de la capa de WTLS es proporcionar privacidad, integridad de los datos y autentificación entre dos aplicaciones en comunicación" [36].

El protocolo de intercambio de WTLS establece una conexión segura entre el cliente y el servidor. También da la posibilidad para que el servidor autentifique al cliente, enviando su certificado. Lo mismo aplica para el cliente. Puede autentificarse enviándole su certificado o una conexión a este, si la autentificación del cliente es solicitada por el servidor [36].

La tabla **2.1** presenta el protocolo de intercambio del WTLS. Las partes marcadas con las muestras del (*) son opcionales.

Client		Server
ClientHello	--->	
		ServerHello
		Certificate *
		ServerKeyExchange *
		CertificateRequest *
	<---	ServerHelloDone
Certificate *		
ClientKeyExchange *		
CertificateVerify *		
[ChangeCipherSpec]		
Finished	--->	
		[ChangeCipherSpec]
	<---	Finished
Application Data	<--->	Application Data

Tabla **2.1**: El protocolo de intercambio de WTLS

La especificación de WTLS no esta realmente fijada a cualquier requisito para la autentificación, es decir, de todos modos, el cliente debe ser autentificado por su certificado. De hecho en los teléfonos móviles actuales, sus browsers de WAP y las tarjetas SIM que son entregados por el estándar, no tienen certificados apropiados. Para estandardizar el almacenaje de los certificados y de las llaves privadas en las terminales de WAP, el foro de WAP ha publicado otra especificación, "La especificación del módulo de la identidad inalámbrica" [37].

El nivel real de la seguridad proporcionada por el WTLS aún se está discutiendo constantemente. La encripción y la autentificación de los usuarios están solamente entre el Gateway-WAP y el equipo móvil (no todo el camino del servicio.). Esto significa que hay un boquete en la seguridad dentro del Gateway de WAP - los datos existen en texto claro y su integridad podría posiblemente ser comprometida así. La autentificación y el intercambio no están tampoco hechas por un estándar de extremo a extremo.

2.4.5.2 Módulo de Identidad Inalámbrica.

El módulo de la identidad inalámbrica es definido por el foro de WAP para ser un dispositivo resistente a la interferencia, que se utiliza para "Realizar funciones de seguridad del nivel de aplicación y WTLS, y especialmente, para guardar y procesar información necesaria para la identificación de usuario y la autentificación" [37]. El WIM es considerado para ser puesto en ejecución en una tarjeta inteligente, en el SIM o una tarjeta separada.

El WIM se utiliza principalmente para los propósitos siguientes [37]:

- Para guardar la llave privada certificada permanentemente (la llave privada nunca sale del WIM)
- Para guardar el certificado del cliente o una conexión a él
- Para realizar operaciones criptográficas durante el intercambio de WTLS, especialmente aquellas necesarias para la autentificación del cliente
- Para firmar datos con la llave privada para los propósitos de la aplicación
- Para "unwrap" por ejemplo, llaves firmadas con la llave pública relacionada con la llave privada en los WIM
- Para guardar certificados de la CA, esta puede ser cambiada
- Para generar los números al azar para los propósitos de la criptografía

Mientras que puede ser visto, a WIM como el que desempeña realmente un papel crucial, en el WTLS. Sin WIM, un mecanismo propietario para manejar las llaves, los certificados y las operaciones de la firma tienen que ser puestos en ejecución. Pues el WIM se diseña como un dispositivo de prueba a la interferencia y proporciona a todas las funciones necesitadas para hacer uso del PKI, no debería realmente ser ningún punto en ir sobre la implementación de una manera diversa.

Actualmente no hay ningún teléfono móvil comercialmente disponible que tengan una capacidad de WIM. Aunque Ericsson ya lo tiene, y fue anunciado al publico el desbloquear tal modelo para inicios del 2001 todavía no es un hecho realmente.

2.4.6. PRACTICAS Y ESTÁNDARES EN EL AMBIENTE MÓVIL.

El uso de Internet y servicios accesibles a través de este, por dispositivos tales como teléfonos móviles están siendo realidad en los últimos años. Los modelos de la autentificación que han sido utilizados se han heredado en gran parte de las contrapartes alámbricas. Pero existen, de todos modos, algunos métodos para hacer autentificación en el Internet móvil que son distintos a los utilizados en redes alámbricas, e inclusive muchas veces no es posible utilizarlos para otra parte.

En el anexo C se encuentra una sección que discute el "status" así como algunos de los estándares más significativos y los progresos que están llegando a ser disponibles a corto plazo.

Comentarios

Autorizar por ejemplo pagos es una de las preocupaciones centrales hechas frente, al hacer comercio en redes abiertas. Por otra parte, cuando los servicios y los recursos se pueden alcanzar sobre diversas clases de redes de información, móvil y fijo, y diversa clase de dispositivos y de interfaces, un mecanismo que permite la autorización de la autentificación debe ser hecho independiente de los atributos del canal o del modo de acceso.

PKI proporciona la base para tales funcionalidades. Los certificados X.509 se especifican ya, para ser utilizados y en parte ya en uso en los teléfonos actuales del GSM equipados de browsers de WAP. El manejo de llaves es, sin embargo, un problema integral hecho frente en estas

configuraciones. La misma llave privada debe ser útil en diversos ambientes, al trabajar en la estación de trabajo en la oficina y al navegar en el Web con un teléfono WAP o del mismo tipo.

El WIM puede ser la solución para este dilema. Su especificación es, sin embargo, aún muy joven y parece haber una cierta agitación en la comunidad de la estandardización de WAP. La iniciativa de MeT formada recientemente por Nokia, Ericsson y Motorola y unida a ellos posteriormente Siemens han comenzado a abordar la misma problemática que el foro de WAP está intentando solucionar con las especificaciones de WTLS y de WIM. En un futuro a corto plazo se mostrará qué dirección adquirirá la estandardización de esta materia. Con la tecnología actual, es ya posible instalar los sistemas con la autentificación y la autorización que se basan en PKI y proporcionar una utilidad de nivel relativamente alto. Solamente el futuro mostrará cuál será el modelo eventual que se adoptará.

Esto nos da ayuda a tener un panorama más amplio aunque para nuestro trabajo de tesis queda como una de las líneas a seguir.

2.5 AUTENTIFICACIÓN DE USUARIOS MÓVILES EN SISTEMAS DE COMUNICACIÓN PERSONAL

Desde los inicios de los PCS existían muchos problemas técnicos sin resolver. Uno de ellos es la autentificación de usuarios móviles, como se ha venido comentando. Debido a lo ya conocido del desarrollo de los PCS basado en radio disponible y redes alámbricas, siendo así muy difícil realizar seguridad y privacidad en tal ambiente de red heterogéneo [40, 41].

Resulta importante plantearnos una estructura lógica de los PCS, para poder analizar y plantear nuestros esquemas de trabajo en esta fase de estudio y análisis, así como para nuestras propuestas manejadas en capitulos posteriores. Así pues, la estructura lógica de los PCS es dividida dentro de tres capas:

- Capa inteligente
- Capa de transmisión
- Capa de acceso

La autentificación de los usuarios móviles es realizada en la capa inteligente. La estructura de la base de datos de PCS difiere de los sistemas existentes, tales como GSM e IS-41, entre otros, donde es centralizada. Por lo tanto los esquemas de autentificación utilizados en sistemas existentes no pueden satisfacer los requerimientos de PCS. Por tal razón se buscan esquemas que estén de acuerdo con las necesidades del ambiente en cuestión. En esta sección, se analiza un esquema de autentificación de usuarios de acuerdo con la estructura de base de datos en la capa inteligente de PCS. Comparando con los esquemas de autentificación existentes, se tienen las siguientes propiedades:

1. La autentificación de usuarios móviles en redes visitadas no es controlada por su red local. La carga del trafico de la actualización y de la interrogación de la localización se reduce ampliamente.

2. El protocolo de autentificación adopta el criptosistema híbrido llave publica / llave secreta.
3. Cuando la red de transmisión de datos adopta Red de Señalización de Canal Común, el esquema aún tiene muy alta seguridad.

Además comparado con los sistemas de comunicación móvil celular existentes, los PCS adoptan los conceptos de la red inteligente (IN) como anteriormente se menciono [42, 43]. La capa inteligente consiste de muchas bases de datos que realizan principalmente dirección de la red y el control de tráfico. Estas bases de datos deben tener las siguientes funcionalidades [44]:

1. Seguimiento y registro de la localización
2. Salida de la llamada
3. Autentificación y verificación
4. Encripción y desencripción
5. Administración del perfil del recurso y del servicio
6. Servicios especiales (marcación rápida, llamada en espera, etc.)
7. Facturación

Por lo que, la seguridad relacionada con funciones de la red es realizada en la capa inteligente de PCS. En los sistemas existentes, tales como GSM, IS-41, entre otros, es utilizada la estructura de base de datos centralizada. Por ejemplo, GSM define 4 bases de datos [45], las cuales son:

- Registro de Localidad Local (HLR)
- Registro de Localidad Visitada (VLR)
- Registro de Identidad de Equipo (EIR) y
- Centro de Autentificación (AC)

La estructura de la base de datos es mostrada en la figura 2.3.

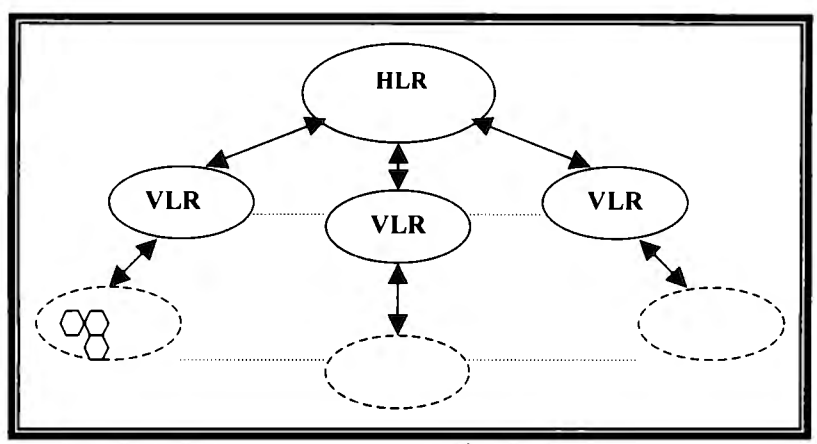


Fig.2.3 La estructura centralizada de GSM

La figura 2.4 muestra las tres estructuras de las bases de datos distribuidas. De acuerdo con esta estructura de bases de datos distribuida, es propuesta una estrategia de privacidad y autentificación (PyA) para PCS. Además se analiza el desempeño de seguridad de los protocolos asociados a esta estructura.

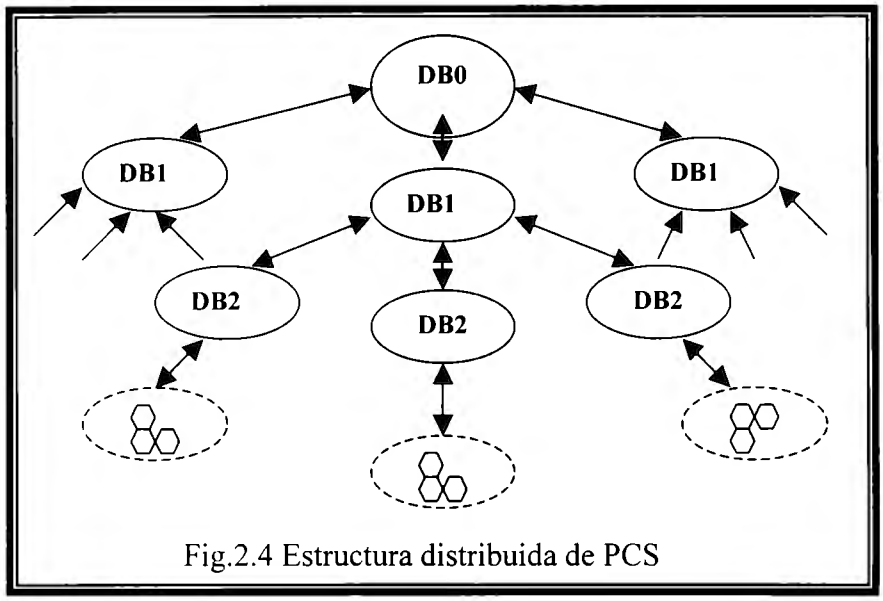


Fig.2.4 Estructura distribuida de PCS

2.5.1 ESTRUCTURA DE SISTEMA

En [46] el autor propone un esquema de PyA híbrido, en el cual tanto sistemas de llave publica y de llave secreta son utilizados. Una autoridad certificadora (CA) se encarga de distribuir certificados para todos los usuarios y controladores de acceso. Aunque, este modelo es muy eficiente por su simplicidad, no es conveniente para expansión de servicios y adición de nuevos usuarios. Además es casi imposible para solamente una CA distribuir todos los certificados en dicha red global. En [47] el autor propone una estructura de autentificación jerárquica conveniente para Internet, basado en esta estructura, una estructura de cuatro capas de CA distribuida es propuesta para PCS. La figura 2.5 muestra la estructura del sistema.

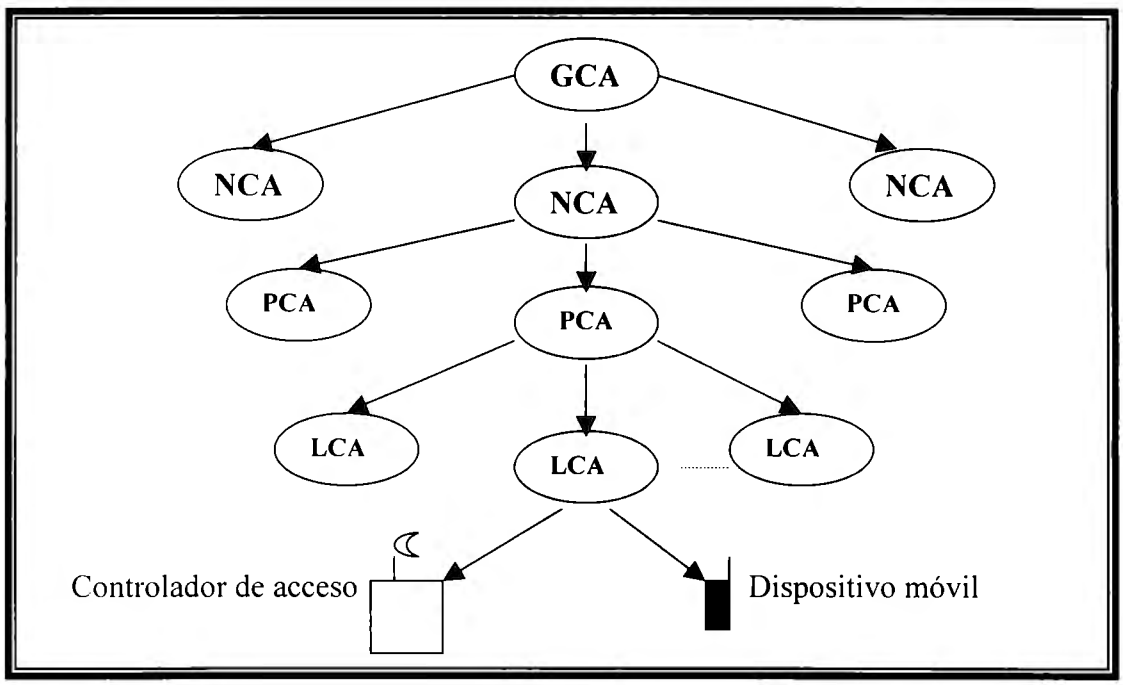


Fig.2.5 Estructura del sistema distribuido de 4 capas

En esta estructura distribuida, las CA, son clasificadas dentro de 4 niveles. Ellas son responsables de distribuir llaves publicas, identidades de usuarios y certificados para cada una de las CA´s en la capa inferior.

1) Las CA´s en la capa superior son llamadas Autoridad Certificadora Global (GCA). Estas son responsables de la distribución de cadenas de datos hacia las CA´s en el 2° nivel. Un ejemplo de las cadenas es como se ve a continuación:

$$\{ Pk^{(0)}, ID_i^{(1)}, C_i^{(1)}, i = 1,2....,I\}.$$

2) Las CA´s en la segunda capa son llamadas Autoridad Certificadora Nacional (NCA). Estas distribuyen cadenas de datos hacia las CA´s en el 3er nivel. La cadena de datos es:

$$\{ Pk^{(0)} || PK_i^{(1)}, ID_i^{(1)} || ID_j^{(2)}, C_i^{(1)} || C_j^{(2)}, j = 1,2....,J\}$$

3) Las CA´s en la tercera capa son llamadas Autoridad Certificadora Provincial (PCA). Estas distribuyen cadenas de datos hacia las CA´s en el 4° nivel. La cadena de datos es:

$$\{ Pk^{(0)} || PK_i^{(1)} || PK_j^{(2)}, ID_i^{(1)} || ID_j^{(2)} || ID_k^{(3)}, C_i^{(1)} || C_j^{(2)}) || C_k^{(3)}, k = 1,2,....,K\}$$

4) Las CA´s en la capa mas baja son llamadas Autoridad Certificadora Local (LCA). Estas distribuyen cadenas de datos a cada controlador de acceso y usuarios de dispositivos móviles "handset" . La cadena de datos para los usuarios es:

5) $\{ Pk^{(0)} || PK_i^{(1)} || PK_j^{(2)} || PK_k^{(3)}, ID_i^{(1)} || ID_j^{(2)} || ID_k^{(3)} || ID_u, C_i^{(1)} || C_j^{(2)}) || C_k^{(3)} || C_u, u = 1,2,....,U\}$

La cadena de datos para los controladores de acceso es:

$$\{ Pk^{(0)} || PK_i^{(1)} || PK_j^{(2)} || PK_k^{(3)}, ID_i^{(1)} || ID_j^{(2)} || ID_k^{(3)} || ID_r, C_i^{(1)} || C_j^{(2)}) || C_k^{(3)} || C_r, \quad r = 1,2,....,R\}$$

Cabe hacer notar que la distribución de todas las cadenas de datos es llevada a cabo fuera de línea[15].

Un ejemplo del certificado en este caso es dado como:

$$\{C_A^{(\ell+1)} = TE || E_{SK^{(\ell)}_{CA}}(h(ID_A^{(\ell+1)} || PK_A^{(\ell+1)} || TE)), \ell=0,1,2,3)$$

donde, el subíndice A representa al dueño del certificado en la capa (ℓ+1), y TE es el tiempo de expiración del certificado $SK^{(\ell)}_{CA}$ es la llave secreta de la CA en el nivel ℓ-th , E es el algoritmo de encripción de llave pública y h es la función hash.

2.5.2 PROTOCOLO DE AUTENTIFICACIÓN PARA REGISTRO DE USUARIOS (APUR)

En esta autentificación, el MS debe establecer un dato de autentificación secreto con el controlador de acceso de la red que esta visitando actualmente.

---

[15] Proceso de distribución no involucrado en el tiempo real del protocolo

1) Primero, el usuario envía un mensaje de pregunta hacia el controlador de acceso. Cuando el controlador de acceso recibe el mensaje, este difundirá el tiempo t de referencia, y la cadena de dato en "plaintext". La cadena es:

$$\{ Pk^{(0)} || PK_i^{(J)} || .... || PK_r, C_i^{(1)} || C_j^{(2)} | .... || C_r, ID_i^{(1)} || ID_j^{(2)} || ... || ID_r, \}$$

2) Tan pronto como el usuario recibe la cadena de datos, este verifica la cadena de certificados primero. Si la verificación pasa, el usuario determinara que información debe ser enviada al controlador de acceso basado en los siguientes 2 casos:
   a. Si la autoridad certificadora de confianza del usuario y el controlador de acceso es un LCA el usuario enviara una cadena de datos corta como se indica:

$$\{ Pk_k^{(3)} || PK_u, C_u, E_{PKr} (ID_u || x || t_u) \}$$

En esta ecuación, él número aleatorio x es un dato de autentificación muy secreto, escogido por el usuario. Obviamente, la información de identidad del usuario y el número x son encriptados con la llave publica del controlador de acceso. Ya que el usuario vaga en su red local la mayor parte del tiempo este caso es generalmente uno.

   b. Si la CA de confianza mutua del MS y el controlador de acceso es localizada en la capa l-th ($0 \leq \ell \leq 3$). El usuario enviara una cadena de datos como sigue:

$$\{ Pk_{no}^{(\ell)} || PK_{n1}^{(\ell+1)} || .... || PK_u, Cn_1^{(\ell+1)} || Cn2^{(\ell+2)} || .... || C_u, E_{PKr} (ID_{n1}^{(\ell-1)} || ID_{n2}^{(\ell+2)} || ID_u || x || t_u) \}$$

3) Cuando el controlador de acceso recibe de la cadena de datos, desencripta la cadena de identidad de usuario utilizando su llave secreta, y obtiene la identidad en "plaintext", x y él time stamp t_u . Después de esto el controlador de acceso verificara lo correcto de t_u y desarrolla la cadena de certificación.

$$(Cn_1^{(\ell+1)} || Cn2^{(\ell+2)} || .... || C_u)$$

Utilizando la llave publica de $PK_{no}^{(\ell)}$ en turno desde ($Cn_1^{(\ell+1)}$ a C_u .) Si la verificación pasa la llave publica del usuario será considerada a ser legal. Entonces el controlador de acceso guardara el dato de autentificación secreto x, y envía un número aleatorio "challenge" "m" hacia el usuario.

4) Cuando el usuario recibe el "challenge m" este inmediatamente firmara el dato utilizando su llave secreta, esto es; $S = Sign_{Sku}$ (m) después se enviará la firma hacia el controlador de acceso.

5) El controlador de acceso verifica la firma utilizando la Pku obtenida en el paso 3. si la verificación pasa, el usuario será considerado legal, y el controlador de acceso enviará el mensaje de reconocimiento al usuario.
   De ahora en adelante el proceso de autentificación de usuario acaba. Un dato de autentificación secreto x ha sido fijado entre el usuario y el controlador de acceso.

2.5.3 PROTOCOLO DE AUTENTIFICACIÓN PARA DISPOSICIÓN DE LLAMADA (APCS)

Ya que el dato de autentificación secreto ha sido fijado durante el periodo de registro, la autentificación mutua entre el usuario y el controlador de acceso durante la dependencia de llamada en esta información secreta.

El protocolo de autentificación para disposición de llamada es:

Dispositivo móvil "handset" **Controlador de acceso**

E_{PKr} (UID), Tu, Un, MAC_{Kur} (t_u, N_u, UID) →

← N_r, T_{ru}, MAC_{Ks} (N_r, Un)

Explicación

1). Primero, el usuario calcula el código de autentificación del mensaje (MAC) utilizando el dato x como la llave, esto es: $K_{ur} = MAC_x$(UID, RID)

UID y RID representan la identidad del usuario móvil y la del controlador de acceso. Esto es:

$UID = ID_{n1}^{(r+1)} || ID_{n2}^{(r+2)} || .... || ID_u$
$RID = ID_{n1}^{(r+1)} || ID_{n2}^{(r+2)} || ... || ID_r$

$MAC_Y(Y)$ puede ser calculada por el empleo de función hash tal como MD5, esto es:

$MAX_Y(Y) = h(Y||x)$
Entonces, el usuario genera un numero N_u y guarda el actual t_u y calcula MAC_{Kur} (t_u, N_u, UID) utilizando K_{ur} como llave. Luego el usuario encripta UID utilizando la llave publica del controlador de acceso y envía el flujo de datos hacia el controlador.

2) Cuando el controlador recibe el flujo de datos, será primero desencriptado E_{PKr} (UID) y consigue el UID en "plaintext" , luego calcula K'_{ur}. Si los dos códigos de autentificación del mensaje $MAC_{K'ur}$ (•) y MAC_{Kur} (•) son iguales, lo cual significa $K'_{ur} = K_{ur}$, el requerimiento de acceso será aceptado; si no, será rechazado. Una vez que él requerimiento fue aceptado, el controlador genera un nonce N_r y calcula MAC_{Kur} (N_r, N_u, t_u). Finalmente este genera una llave de sesión K_s para esta disposición de llamada, calcula
$T_{ru} = MAC_{Kur}$ (N_r, N_u, t_u) + K_s y MAC_{Ks} (N_r, N_u) y envía el flujo de datos al usuario móvil.

3) Cuando el usuario recibe el flujo de datos, este sustraerá K'_s desde T_{ru} y calcula
$MAC_{K's}$ (N_r, N_u) Si $MAC_{K's}$ (•) = MAC_{Ks} (•), el usuario considerara al controlador legal. Esto es, el usuario guardará la llave de sesión K_s, de otra manera, el usuario considerara que el controlador es ilegal y el "handset" alarmara al usuario. De aquí en adelante, la autentificación

mutua de las dos partes comunicantes y la distribución de llave ha terminado y el enlace secreto ha sido fijado.

2.5.4 ANÁLISIS DE SEGURIDAD DE LOS PROTOCOLOS DE AUTENTIFICACIÓN

El análisis de seguridad cualitativo de los protocolos será dado a continuación. El análisis cuantitativo esta fuera del alcance de este documento. Se hace un examen de seguridad de los protocolos basados en los siguientes ataques:

1) Ataque de spoof. Un atacante pos si mismo puede generar un par de llaves publicas PK'_u / SK'_u y tratar de forjar una firma en su llave pública. El controlador de acceso detectará la falsificación por la verificación de la cadena de certificación.
2) Ataque de replay. Por que el protocolo adopta un nonce y un "timestamp" en ambos APUR y APCS, este ataque de replay será prevenido con éxito.
3) Ataque de impersonación. En APUR, la impersonación de usuarios legales será encontrada por el controlador en la verificación de la firma del "challenge m", cuando la impersonación del puerto de red sea encontrado por los usuarios en la verificación de la cadena de certificación recibida. En APCS, debido a que el atacante no puede obtener el dato x de autentificación secreta desde APUR, la impersonación de ambos usuarios legales y puertos de red es imposible.
4) Ataque de seguimiento de localidad. En ambos protocolos la identidad del usuario móvil es encriptada por la llave publica del controlador. No hay ningún escape de información del movimiento al atacante.
5) Ataque de "tampering" (tratar de forzar). Un atacante puede modificar el flujo de datos en el canal de comunicación de tal forma que el usuario no pueda extraer la llave de sesión K_s del flujo de datos. Pero, esta amenaza será detectada por el usuario al verificar el código de autentificación de mensaje MAC_{Ks} (N_r, N_u)

2.6 PROTOCOLO DE AUTENTIFICACIÓN DINÁMICA PARA SISTEMAS DE COMUNICACIÓN PERSONAL

Ahora analicemos la propuesta de [49], que dice ser un protocolo seguro dinámico adecuado para las PCS, el cual se basa en llave publica y algoritmo híbrido de llave secreta, timestamp y una variable pequeña de acumulación incorporada en el carácter dinámico.

El protocolo aborda el problema de amenazas de un ataque de copia, que no puede ser resistido en él articulo [46] y remedios para los problemas existentes en este trabajo [50, 51] en el cual los máximos tiempos de llamada hallan sido limitados. Más aún, la implementación del protocolo comienza a ser más conveniente. Se analizan las características de este protocolo y se discute que el protocolo puede efectivamente resistir al tipo de ataque de replica, ataque de copia, del cual muchos esquemas de seguridad no pueden vencer.

En los PCS futuros, 3 factores han realzado la necesidad de contar con algoritmos de llave publica para aplicaciones en PCS.

1° Los PCS pueden ofrecer ancho de banda amplio sobre celular convencional, el cual permite una más rápida transferencia libre de error en ambientes de llave publica grandes cruzando el canal de PCS.

2° Con el desarrollo acelerado en las técnicas microelectrónicas las habilidades de calculo de procesadores de bajo costo se incrementan ampliamente.

3° La introducción de nuevos algoritmos de llave publica proporcionará la misma seguridad que sé tenia en los primeros, pero la cantidad de cálculos pueden disminuir.

Dan Brown a propuesto un protocolo de autentificación y privacidad para PCS conveniente, basado en llave-pública y un algoritmo híbrido de llave secreta [46], sin embargo, el protocolo de seguridad no puede resistir ataques de copia donde el atacante copie información valida y transmitir a otro.

Los artículos [50] y [51] también avanzaron los protocolos de seguridad basados en llave pública y algoritmo híbrido de llave secreta para PCS. Estos protocolos tienen buena seguridad. Sin embargo, en ellos existen serias limitaciones; resultado de la llave de autentificación C_j finita, que es guardada en la red por adelantado. Todas las llaves de autentificación las cuales fueron calculadas con formula de recurrencia en la red local, son transmitidas hacia la red visitada durante el registro de usuario. Él número de llaves que son guardadas en la red visitada limita al subscriptor a un numero máximo de llamadas validas. A pesar de que se pueden proveer suficientes llaves de autentificación durante el registro, la limitación intrínseca es un detalle de este esquema.

En este caso, el protocolo seguro dinámico puede ser establecido en PCS. Este incluye 2 secciones:

- Protocolo de registro de subscriptor y
- Protocolo de disposición de llamada

2.6.1. TRABAJO PRELIMINAR DEL PROTOCOLO

Antes de ver el protocolo de seguridad, revisemos las siguientes descripciones que tienen ventajas en la comprensión del protocolo y son valor de referencia para aplicación practicable en PCS.

1. Todos los subscriptores y redes de servicio son equipados con los instrumentos que pueden adquirir señales de reloj exacto.
2. Todos los subscriptores y redes de servicio adoptan el mismo algoritmo de llave publica (E)
3. La llave secreta de cualquier subscriptor y red no pueden hacerse idénticas para evitar debilidad de la privacidad de la información después de una encripción doble.
4. El protocolo ignora autentificación mutua entre subscriptor y la unidad móvil, esto significa que el subscriptor y la unidad móvil son considerados como partes enteras.

5. Suponiendo que Δt es el retardo de tiempo grande desde el subscriptor hacia la red, pensaríamos a Δt como el valor de ventana permisible de tiempo para verificar el "timestamp".
6. Para entender el protocolo, la siguiente es una lista de la notación utilizada en el protocolo.

r_{ih}, r_{iv}: El valor del modulo de tiempos de llamada de subscriptores en red local y visitada.
r_{hn}, r_{vn}: El valor del modulo relacionado ha tiempos de llamada de subscriptores a ser guardados en la red local y visitada.
E: Algoritmo de llave publica en el protocolo.
h(a,b): Función hash de un solo sentido con 2 parámetros.
e_{iu}, d_{iu}: Llaves de desencripción y encripción de subscriptores, donde e_{iu} se mantiene secreta, d_{iu} se hace pública.
e_{hn}, d_{hn}: Llaves de desencripción y encripción de la red local, donde e_{hn} se mantiene secreta, d_{hn} se hace pública.
e_{vn}, d_{vn}: Llaves de desencripción y encripción de la red visitada, donde e_{vn} se mantiene secreta, d_{vn} se hace pública.
ID_i, HID, VID: Identidades de subscriptor, red local y red visitada.
TID_i: Identidad temporal del subscriptor dada por la red visitada.
t_{iuk}, t_{hnk}, t_{vnk}: Timestamp de subscriptor, red local y red visitada cuando transmiten información, k=1,2,3.....
t_{iu}, t_{hn}, t_{vn}: Reloj del subscriptor, red local y red visitada cuando reciben información.
Δt, Δt_1: La ventana permisible de tiempo máximo entre usuario y red, red y red.
p: El valor del modulo, podría ser un valor pequeño.

2.6.2 PROTOCOLO SEGURO EN PCS

Los protocolos de seguridad son divididos en dos partes:

Protocolo de registro de subscriptor y protocolo de disposición de llamada. A continuación se muestran estos protocolos.

2.6.2.1 Protocolo de registro de subscriptor

El siguiente procedimiento es invocado cuando vaga el subscriptor dentro de una área de servicio nueva, y aplica para el registro.

1. El subscriptor calcula:

$\alpha = Ed_{hn}(ID_i, Ee_{iu}, (r_{ih}))$
Luego, envía HID, t_{iu1}, α a la red visitada

2. La red visitada verifica t_{iu1}, para ver si t_{iu1} pertenece a $(t_{vn}-\Delta t_1, t_{vn})$. Si es así, esta genera un numero aleatorio a y calcula:

$\beta = Ed_{hn}(VID, Ee_{vn}, (a))$, posteriormente, esta envía VID, t_{vn1}, α, β, hacia la red local.

3. La red local verifica t_{vn1}, si el timestamp falla, es rechazada la solicitud, de lo contrario, es desencriptada α con su llave secreta e_{hn} para conseguir ID_i y $Ee_{iu}(r_{iu})$. Luego, busca la llave publica d_{in} de ID_i y desencripta Ee_{iu} (r_{ih}) para verificar sí r_{ih} es igual a r_{hn}; Si no es así, es rechazada para esta solicitud; de otra manera, la red local reconoce al llamante como un subscriptor legitimo. Luego esta desencripta β para conseguir VID y Ee_{vn} (a). Además, consigue d_{vn} y desencripta Ee_{vn} (a) para conseguir a. Finalmente, la red local envía Ed_{vn} (a), Ed_{iv} (r_{hn}), t_{hn1}, d_{iu} de regreso a la red visitada.
4. La red visitada examina la validez del timestamp, si el timestamp es bueno, esta desencripta Ed_{vn} (a) con e_{vn} para ver si a existe. Si es así, esta asigna al subscriptor una identidad temporal TID, y guarda d_{in}. Finalmente, envía t_{vn2}, d_{vn}, VID, Ed_{iu} ($Ee_{vn}(TID_i)$), $Ed_{iv}(r_{hn})$ para suscribir i, y poner su variable local r_{vn} para 0.
5. El subscriptor examina si t_{vn1}, pertenece a $(t_{iu}-\Delta t, t_{iu})$; si cumple, desencripta Ed_{iv} (r_{hn}) para conseguir r_{hn} , si r_{hn} es igual a r_{ih}, se desencripta Ed_{iu} (Ee_{vn} (TID_i)) con e_{iu} y d_{vn} para conseguir TID_i, Luego, se pone la variable local r_{iv} a 0. Finalmente, se salva $\{VID, TID_i, d_{vn}, r_{vni}\}$.

Hasta ahora, el subscriptor que vaga esta registrado satisfactoriamente en la red visitada legítimamente. El subscriptor y la red visitada utilizaran la información acumulada en esta fase para autentificarse mutuamente sin interactuar con la red local. Cualquier solicitud de llamada hecha por el subscriptor invocara el siguiente protocolo.

2.6.2.2 Protocolo de disposición de llamada

El protocolo lleva a cabo los siguientes pasos:

1. El subscriptor genera un número aleatorio Rand y calcula
$\alpha = Ed_{vn}$ (TID_i, Ee_{iu}, (r_{iv}, m, Rand)
Luego, , calcula $r_{iv} = (r_{iv}+ 1)$ mod p y envía t_{iu1} a la red de servicio. Si el subscriptor no recibe la respuesta de la red durante el tiempo dado, $r_{iv} = (r_{iv} -1)$ mod p debe ser calculado.

2. La red de servicio verifica t_{iv1}, si t_{iv1} pertenece a $(t_{vn}-\Delta t, t_{vn})$, desencripta α con la llave secreta e_{vn} para conseguir TID_i, luego busca d_{in} y desencripta Ee_{iu} (r_{iv}, m, Rand) para conseguir r_{iv}, m y Rand. Si r_{iv} es igual a r_{vn}, esta calcula $r_{vn} = (r_{vn}+ 1)$ mod p y escoge un numero aleatorio Rand1. Para una solicitud valida, pone la llave de sesión K_c = h(Rand, Rand1). Finalmente envía t_{vn1}, β hacia el subscriptor, aquí $\beta = Ed_{iu}$ (Rand, Ee_{vn}, (m_1, Rand1, r_{vn})).
3. El subscriptor verifica t_{vn1}, si t_{vn1} pertenece a $(t_{iv}-\Delta t, t_{iv})$, desencripta β con la llave secreta e_{iu} para ver si Rand está presente. Si es así, calcula $r_{iv} = (r_{iv} + 1)$ mod p luego esta utiliza d_{vn} para desencriptar Ee_{vn} (m_1, Rand1, r_{vn}) para conseguir m_1, Rand1 y r_{iv}. Más aún, si r_{iv} es igual a r_{vn}, la respuesta se pensaría ser valida. Finalmente, el subscriptor calcula K_c = h(Rand, Rand1) como llave de sesión. De ahora en adelante, el subscriptor ha establecido conexión de llamada satisfactoria.

Los protocolos de seguridad discutidos solo están basados en las condiciones de que el servicio sea provisto en un estado de movimiento. De hecho, si el subscriptor se halla por casualidad en su red local, el roll de red visitada es realmente tomado por la red local, y la interacción entre la red visitada y la local no existe durante el protocolo de registro. Es decir es transparente para el subscriptor.

2.6.3 ANÁLISIS DEL PROTOCOLO

Características mejores del protocolo

Comparando con muchos otros protocolos, la mejor característica del protocolo es que la seguridad es además mejorada a través de la introducción de timestamp y variables tipo-r, donde las variables tipo-r se refieren a r_{iu} de subscriptores y r_{iv}, r_{hn} de red local y r_{vn} de red visitada.

Este protocolo tiene las siguientes características:

1. Las variables tipo-r pueden ser muy pequeñas, por ejemplo, es suficiente para asignar acumuladores de bits pequeños.
2. Las variables tipo-r son consideradas como la principal certificación de autentificación mutua entre la red local y el subscriptor durante la etapa de registro de subscriptor. Más aún, mejora la seguridad de autentificación entre el subscriptor y las redes de servicio durante la etapa de disposición de llamada. Además, las ventajas de seguridad aportadas por la introducción de variables tipo-r tienen un análisis detallado en la siguiente sección.
3. Por la adopción de variables las cuales guardan los tiempos de llamada pueden remplazar el trabajo de timestamps y variables tipo-r, pero el acumulador limitado de las variables permite al problema similar para el problema de tiempos de llamada limitados en el articulo [50]. En este protocolo, a través de variables tipo-r pueden ser muy pequeñas, esto no limita los tiempos de llamada.

2.6.3.1 Seguridad del protocolo

El protocolo tiene características seguras buenas, el cual puede prevenir de ataques ilegales tradicionales tales como; "eavesdropping", enmascaramiento y reemplazo de información, entro otros. Además, este todavía tiene las siguientes característica de seguridad:

1. Independencia de la llave de sesión
   La llave de sesión K_c, la cual es utilizada para encriptar información transmitida entre subscriptor y red, solamente confía en las funciones hash de que parámetro permite a los dos números aleatorios ser generados por el subscriptor y la red durante cada llamada. El método no solamente provee confidencia de la transmisión de la información, también asegura la independencia de la llave de sesión. Además, incluso si un atacante ha roto una información válida por suerte, es inútil romper la otra información.

2. Protección de la localidad del subscriptor
   En el protocolo, por la protección de la llave publica en él ID del subscriptor, cualquier intruso (externo) no podrá conseguir un ID valido del subscriptor. Por lo tanto, el protocolo puede proteger el mensaje de localidad del subscriptor en efecto.

3. No-repudiación del servicio
   Después de que el subscriptor o la red provean el servicio a cada uno, ellos pueden rechazar reconocer los servicios. El protocolo resuelve este problema. Por que la llave secreta de un subscriptor es solamente conocida por este subscriptor y acreditable por

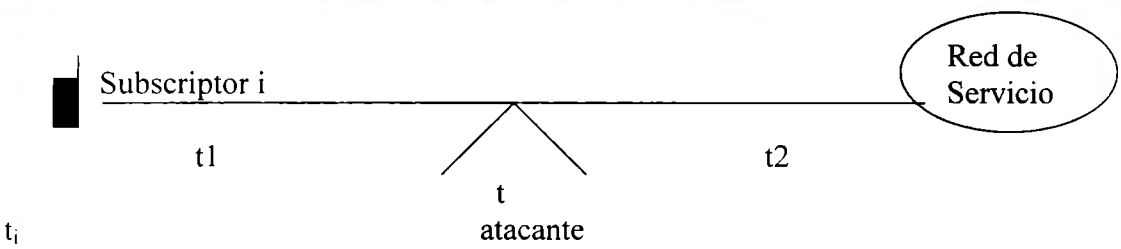
AC, ningún otro subscriptor no puede fraguar la información del subscriptor en cuestión. Por la misma razón, cualquier otro atacante no puede enmascarar a la red.

4. Resistencia de ataques de copia tipo-retransmisión

El ataque de copia es el más efectivo entre todos los métodos de ataque. Si solamente un atacante ilegal posee el equipo de transmisión y recepción correspondiente, él podrá realizar el ataque valido en cualquier tiempo. Incluso si el atacante no conoce ninguna información relacionada con la red o subscriptor, él podrá destruir el sistema. En los protocolos, el principal efecto del timestamp y las variables tipo-r son para tener éxito en la defensa contra ataques de copia.

La efectividad del timestamp puede prevenir de ataques de copia, por que un mensaje de copia valida en tiempo anterior comenzara un mensaje invalido en otro tiempo. Por lo que, es generalmente considerado que adoptando un timestamp puede evitar la amenaza resultado del ataque de copia. Al menos, en este documento se enfatizan el ataque de copia tipo-reenvío, el cual hace timestamps para hacerlo inútil al ataque.

Ya que la verificación de timestamp debe reservar una cierta ventana de tiempo de permiso, esta da al atacante una oportunidad para romper el sistema. Suponiendo un subscriptor ilegal halla capturado y copiado un mensaje valido, este inmediatamente envía el mensaje a otro. Luego, si el tiempo alcanzado del mensaje es satisfactorio con el requerimiento de ventana de permiso de tiempo, puede pasar la verificación de timestamp. El mensaje de reenvío se convierte en un mensaje valido. Por lo tanto, el protocolo que solamente toma timestamps sencillos no puede resistir este ataque de copia. El método de ataque es mostrado a continuación (figura 2.6).



t_i: El tiempo del mensaje enviado por el subscriptor
t_n: El tiempo del mensaje recibido por la red
t1 + t2 + t : desde el tiempo t_i hasta el tiempo cuando el mensaje copiado del atacante es enviado a la red
si $t1 + t2 + t \in (t_n - \Delta t, t_n)$?

Figura 2.6. Ataque de copia tipo- retransmisión

La introducción de variables tipo-r en el protocolo previene efectivamente de amenazas de ataque de tipo-reenvío. Cuando la red recibe un mensaje valido que es transmitido por el usuario, se examina el timestamp del subscriptor para verificar si dicho timestamp es apropiado. Si es así, entonces trae r_n que es guardado en la red para verificar la r_u del

subscriptor. Si r_u es igual a r_n el mensaje es tomado como solicitud valida, en el mismo momento el valor de r_n es cambiado adicionando 1. Suponiendo que la red recibe una copia de mensaje que es transmitido por un subscriptor ilegal, y el mensaje pasa satisfactoriamente el chequeo de timestamp. Subsecuentemente, la red verifica la r_u del mensaje copia. Por que la r_n que se salvara en la red no es la misma que r_u, el mensaje copia es pensado como un mensaje invalido y será rechazado.

Comentarios

Los protocolos vistos en esta sección se resumen de la siguiente manera:

- Se construye un protocolo seguro dinámico que es útil para sistemas de comunicación personal.
- La característica esta en el uso de timestamp y variable de acumulación.
- El protocolo no solamente supera el problema de amenazas de ataque de copia del articulo [46] sino, también ofrece remedio a los tiempos de llamada máximos a ser limitados.
- Por ultimo, se pone énfasis en la discusión que el protocolo puede resistir efectivamente a ataques de copia del tipo-reenvío de los cuales muchos esquemas seguros no pueden conquistar.

2.7 PROTOCOLO DE AUTENTIFICACIÓN PARA SISTEMAS DE COMUNICACIÓN PERSONAL

Los protocolos de autentificación mutua de GSM e IS-41 son computacionalmente eficientes e impiden el enmascaramiento y el fisgoneo o ataque de escucha. Sin embargo estos protocolos no soportan el no repudio del servicio. En este apartado se estudia la propuesta de un protocolo de autentificación para sistemas de comunicación personal enfatizando la no repudiación y prevención de ataques de"play back" [52]. Se extiende el núcleo de las funciones de autentificación de GSM para incluir una función de un solo camino que establece veracidad entre las unidades móviles y las localidades de registro visitadas (VLR).

2.7.1 REVISIÓN DE GSM

La autentificación y los protocolos de registro en GSM e IS-41 tienen similitudes en su fundamento. En GSM [53], La unidad portátil (PU) se comunica por radio con la estación base (BS). La BS es conectada con los centros de "switcheo" móvil (MSC), las cuales son conectadas con las redes alámbricas existentes. Por cada proveedor de servicio a móviles, existen dos bases de datos:

- Una es el Registro Ubicación Local (Home Location Register HLR),
- y la otra es el Registro de la Ubicación Visitada (Visited Location Register VLR).

HLR guarda información de sus propios subscriptores.
VLR guarda la información de los subscriptores visitantes.

Existe un Centro de Autentificación (AC) en la arquitectura el cual guarda las llaves secretas de los subscriptores y genera los parámetros de seguridad sobre las peticiones del HLR. Cada subscriptor tiene una única identidad; Identidad de Subscriptor Móvil Internacional (IMSI) y una llave secreta (k_i). Durante la autentificación, 2 funciones de un solo sentido (A_3, A_8), y un algoritmo de llave de encripción/desencripción (A_5) es empleado. Cuando el subscriptor va por primera vez a una nueva localidad, este envía su IMSI a la VLR, la cual en su momento envía la pregunta a la correspondiente HLR. Entonces HLR busca una secuencia de 3-tuples (RAND, SRES, Kc) desde la AC y envía estos hacia el VLR. RAND es un numero de pregunta aleatoria, SRES es una respuesta firmada de la pregunta y Kc la llave calculada. Para verificar la identidad del subscriptor, el VLR envía el RAND hacia el subscriptor, quien emplea K_i para calcular Kc y SRES. La respuesta SRES es enviada de regreso hacia el VLR. Si la SRES almacenada checa con la recibida SRES, VLR envía una encriptada, TMSI al subscriptor. Un segundo intercambio es empleado para convencer al subscriptor de que VLR es legitimo. La voz es cifrada empleando Kc. La figura 2.7, esquematiza el protocolo de autentificación de GSM. Los mensajes son especificados por [ ] y la dirección de la transmisión por →.

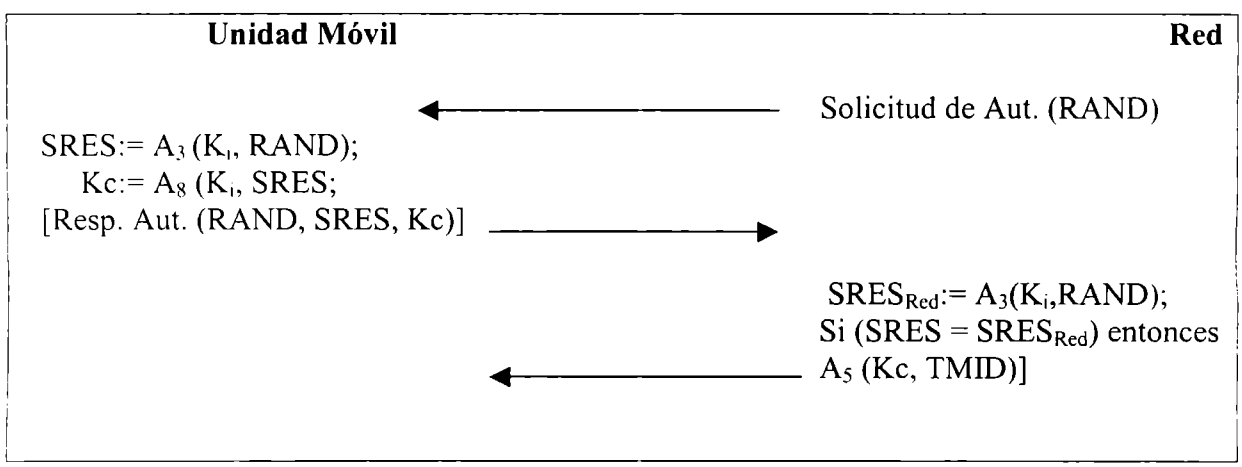


Figura 2.7. Autentificación de GSM

En términos generales, GSM es un protocolo simple y eficiente para PCS. Sin embargo, este no soporta no-repudiación del servicio. La VLR siempre depende de HLR para la autentificación.

2.7.2 Enriquecimiento de la autentificación por no-repudiación del servicio.

La principal motivación de este trabajo es la de incluir no-repudiación del servicio por el empleo de una función de un solo sentido. El protocolo incluye 2 fases:

Registro y establecimiento de sesión. No se quiere que este protocolo sea incompatible con los fundamentos de autentificación de GSM o IS-41 ya que este procedimiento es empleado en ambos; registro y establecimiento de la sesión. Por lo que se adoptan, notaciones generales para el protocolo que no sugieren un particular método de encripción o implementación funcional. Se asume que la unidad móvil comparte una llave secreta con HLR y la comunicación entre el VLR y HLR es segura.

MU	Unidad Móvil
HLR	Registro de Ubicación Local
VLR	Registro de ubicación visitado
ID	Identidad en tramos largos de la MU
TID	Identidad de MU temporalmente
$E_k( )$	Función de encripción tradicional con llave k
$D_k( )$	Función de desencripción tradicional con llave k
f( )	función de un solo sentido
+	Computación lógica, or exclusivo
Kc	llave de sesión
r_v	Numero aleatorio seleccionado por VLR
r_m	Numero aleatorio seleccionado por la unida móvil
r_h	Numero aleatorio seleccionado por HLR.

Registro

Los siguientes pasos son tomados para el registro de llamada.

- VLR Transmite un nonce (numero aleatorio) r_v para frescura.
- La MU escoge un número aleatorio r_m y encripta r_v, r_m con k (la llave es compartida con HLR). MU envía (ID, r_v, E_k (r_v), $E_k(r_m)$) para VLR.
- VLR pasa los parámetros (ID, r_v, E_k (r_v), $E_k(r_m)$) a HLR.
- Recibidos los parámetros desde VLR, HLR desencripta E_k (r_v) y checa que el resultado sea igual a r_v. Si se cumple, HLR esta convencido que MU es un subscriptor legal. HLR escoge un numero aleatorio r_h , una llave de sesión K_c, y encripta estas con k . HLR también calcula f^n ($r_h \oplus r_m$) y envía (ID, E_k (r_h), f^n ($r_h \oplus r_m$), K_c E_k (K_c), n) hacia VLR. Aquí

  f_n ( ) = f (f^{n-1} ( )), f^1( ) = f( ).

- VLR guarda (f^n ($r_h \oplus r_m$), K_c , n) bajo el nombre del ID, elige una identidad temporal TID para el usuario registrado, emplea f^n (K_c) para encriptarlo y envía (E_k (r_h), $E_k(K_c)$, n, $E_{f^n(Kc)}$) hacia MU.
- MU desencripta E_k (r_h), E_k (K_c) para conseguir r_h, K_c y luego calcula f^n (K_c) para desencriptar TID. El registro es en este momento completo.

Registro

VLR → MU: r_v
VLR → HLR: r_v
MU : Escoge r_m
VLR → MU: (ID, r_v, E_k (r_v), $E_k(r_m)$)
VLR → HLR: (ID, r_v, E_k (r_v), $E_k(r_m)$)
HLR : Desencripta E_k r_v) y checa que r_v sea valida si se cumple HLR valida MU, HLR escoge r_h, K_c y las encripta con k. HLR calcula: f^n ($r_h \oplus r_m$)
HLR → VLR: (ID, E_k (r_h), f^n ($r_h \oplus r_m$), K_c E_k (K_c), n)

VLR : Guarda (f^n ($r_h \oplus r_m$), K_c , n) bajo el nombre de ID, escoge TID emplea f^n (K_c) para encriptarlo

VLR → MU: (E_k (r_h), $E_k(K_c)$, n, $E_{f^n(Kc)}$)

MU : Desencripta E_k (r_h), E_k (K_c) Para conseguir r_h, K_c. Entonces Calcula f^n (K_c) para Desencriptar TID

El registro es completo ahora.

Establecimiento de la sesión

- Para iniciar una sesión MU envía (TID), f^{n-1} ($r_h \oplus r_m$) hacia VLR.
- VLR emplea la función de un solo sentido f( ) para checar si f(f^{n-1} (($r_h \oplus r_m$)) = f^n ($r_h \oplus r_m$) .
- Si se cumple, entonces VLR calcula f^{n-1} (K_c) y utiliza esta para encriptar f^{n-1} ($r_h \oplus r_m$) y envía esta de regreso a MU, decrementa n, y actualiza f^n ($r_h \oplus r_m$) con f^{n-1} ($r_h \oplus r_m$).
- Una vez recibido $E_f{}^n{}_{(Kc)}$ (f^{n-1} ($r_h \oplus r_m$)) MU trata de desencriptar f^{n-1} ($r_h \oplus r_m$).
- Si esta es la misma que fue enviada anteriormente, entonces MU cree que esta comunicándose con un VLR legal, actualiza n por (n-1) , y establece la llave de sesión como f_{n-1} (K_c).

MU → VLR: (TID), f^{n-1} ($r_h \oplus r_m$)

VLR : Emplea f( ) para checar si f(f^{n-1} (($r_h \oplus r_m$)) = f^n ($r_h \oplus r_m$). Si se cumple, entonces VLR calcula f^{n-1} (K_c) y utiliza esta para encriptar f^{n-1} ($r_h \oplus r_m$) y envía esta de regreso a MU, decrementa n, y actualiza f^n ($r_h \oplus r_m$) con f^{n-1} ($r_h \oplus r_m$).

VLR → MU: $E_f{}^{n-1}$(KC) (f^{n-1} ($r_h \oplus r_m$))

MU : trata de Desencriptar f^{n-1} ($r_h \oplus r_m$). Si esta es la misma que fue enviada anteriormente, entonces MU valida que esta comunicándose con un VLR legal, actualiza n por (n-1) , y establece la llave de sesión como f_{n-1} (K_c).

2.7.3 ANÁLISIS

La comunicación segura es garantizada por el establecimiento de la llave de sesión. El servicio de no-repudiación es lo que más concierne en este caso.

Argumentos.

1. Protegiendo ID del llamante. En este protocolo, después de que él MU ha sido registrado en la VLR, se le asigna una TID para proteger la identidad real del "caller".
2. Autentificación mutua. Este protocolo soporta este tipo de autentificación durante la sesión de inicialización cuando la sesión i-esima es inicializada MU envía f^{i-1} ($r_h \oplus r_m$) a VLR para indicar su identidad VLR envía $E_f{}^{i-1}{}_{(Kc)}$ (f^{i-1} ($r_h \oplus r_m$)) para demostrar la posesión de f^{i-1} (K_c). En este sentido VLR prueba su identidad al MU.
3. No-repudiación. Se emplea función de un solo sentido en la sesión i-esima, MU provee f^{n-1} ($r_h \oplus r_m$) por la función "de un solo sentido", pero esta puede no conseguir f^{n-1} ($r_h \oplus r_m$) desde f^{n-i+1}($r_h \oplus r_m$). Por lo tanto f^{n-i} ($r_h \oplus r_m$) puede ser empleada como prueba de i-esima conexión, Siempre que un "challenge" ocurra el VLR puede ser requerido para mostrar f^{n-i} ($r_h \oplus r_m$).

4. Evitando el ataque de "play back". En este protocolo r_v debe ser actualizado regularmente, por que ($r_h \oplus r_m$) es determinado por ambos MU y HLR, será diferente en cada momento y previamente guardado, no puede haber "play back".

Comentarios

En esta sección, se revisan los modelos de seguridad empleados en Sistemas de Comunicación Personal y autentificación GSM. Los requerimientos de seguridad para sistemas de comunicación personal fueron sumados y un protocolo de autentificación mejorado es propuesto como una generalización de la autentificación de GSM. Este protocolo adopta la tradicional encripción de una sola llave para simplificar los procesos de autentificación. Una función de un solo sentido es utilizada para implementar servicio de no-repudiación. En el anexo C se puede profundizar mas en las interfases aéreas y módulos de seguridad

2.8 PROTOCOLO DE AUTENTIFICACIÓN SIN TERCERA PARTE DE CONFIANZA

Un protocolo de autentificación segura, el cual soporta tanto la privacidad de mensajes como la autenticidad de las partes que se comunican es el que se analiza ahora [54]. La tercera parte de confianza (centro de información de llaves) no es necesario una vez que el sistema de red segura es levantado. Autentificación mutua y distribución de llave pueden ser logrados con 2 mensajes solamente entre las dos partes involucradas.

El primer esquema basado en "ID" (ID-based), propuesto por Shamir [55], soporta solamente firma digital en vez de encripción de mensajes. Tsujii propuso otro criptosistema "ID-based", el cual esta basado en el problema de logaritmos discretos [56], mismo que sufre del problema de conspiración y necesita un alto "overhead" de cálculos exponenciales. Okamoto y Tanaka extendieron la idea de Shamir y combinaron firma digital y distribución de llave en un simple esquema basado en ID (ID-based) [57]. El cual soporta encripción de mensajes y problemas de conspiración. Sin embargo en el esquema las identificaciones de usuario pueden ser forjadas, información secreta de usuario puede ser divulgada y el alto overhead de los cálculos exponenciales es necesario.
En dicho esquema (Shiuh-Pying Shieh et al), proponen un nuevo protocolo de autentificación en el cual el centro de información de llaves es necesario solamente cuando el sistema de red segura esta levantado o cuando nuevos usuarios solicitan registro. No solamente este ultimo protocolo necesita de pocos cálculos exponenciales sino también resuelve el problema de seguridad que apareció en el esquema de Okamoto y Tanaka.

2.8.1 PROTOCOLO DE AUTENTIFICACIÓN SEGURO

Tanto el esquema de ID-based y la técnica de criptografía simétrica son utilizadas en este protocolo.

El esquema de "ID-based" es empleado para autentificación y disposición del sistema, mientras que la criptografía simétrica es empleada para encripción de mensajes subsecuentes para obtener mejor desempeño de comunicación. Hay dos fases en este protocolo de autentificación:

La fase inicial, es completada en el centro de información de llave para levantar el sistema, y
La fase de autentificación, es ejecutada entre las dos partes comunicantes para alcanzar autentificación mutua e intercambiar la llave de sesión común.

fase inicial

El centro de información no es responsable ni de la autentificación mutua ni de la generación de claves comunes. El rol del centro es simplemente generar información publica y secreta para nuevos usuarios registrados. Cuando el sistema de red segura es creado, el Centro de Información de Llaves (KIC) ejecutara los siguientes pasos.

1) Escoge 2 números primos grandes p y q y permite calcular $n = p \cdot q$

2) Obtiene la información "d" secreta del centro de los siguientes cálculos, donde "d" es solamente conocida por el centro.

$$3 \cdot d \pmod{(p-1) \cdot (q-1)} = 1 \quad \ldots\ldots(6)$$

3) Encuentra un entero "g" el cual es un elemento primitivo en ambos GF(p) y GF(q), donde "g" es información publica del centro.

4) Deja a ID_i denotar la identidad del usuario i quien solicita registro para esta red segura. El ID_i puede ser compuesto de nombre, dirección, ...y así.

5) Escoge una función de un solo sentido f para calcular la identidad extendida (EID_i) de i, como sigue:

$$EID_i \equiv f(ID_i) \pmod{2^N}$$
$$\equiv (EID_{i1}, EID_{i2} \ldots\ldots\ldots EID_{iN},) \quad \ldots\ldots(7)$$

Donde N denota la longitud del bit de EID

6) Después de calculado EID_i, se calcula la información secreta del usuario S_i como:

$$S_i \equiv EID_i^d \pmod{n} \quad \ldots\ldots(8)$$

Y de las relaciones arriba mencionadas, la siguiente ecuación sería obtenida

$$EID_i \equiv S_i^3 \pmod{n} \quad \ldots\ldots(9)$$

7) Envía (n,g, f(x), S_i) de regreso al usuario sobre un canal seguro, por ejemplo un certificado y correo sellado. Sobre lo recibido de la información, el usuario i debe mantener S_i secreto y guardar la información publica (n, g, f(x)).

Una vez que el sistema de red seguro es creado, el KIC no es necesario ya, excepto cuando nuevos usuarios se unen.

La información "d" del centro debe ser guardada secretamente para usos subsecuentes. Sin embargo, los enteros p y q no serán ampliamente utilizados y deberían ser desechados secretamente. Cuando un nuevo usuario solicita unirse, él envía al centro su ID. Sobre lo recibido de los ID del usuario, el centro repite los pasos 5 al 7.

Fase de autentificación

El protocolo de autentificación aquí propuesto solamente necesita 2 mensajes para completar la autentificación mutua. Sobre lo recibido del 1er mensaje desde el usuario i, el usuario j verifica el contenido del mensaje. Sí la verificación tiene éxito, el cree que el mensaje es enviado por el usuario i, por lo tanto, el usuario j autentifica al usuario i. Similarmente el usuario i autentifica al usuario j con el 2° mensaje. Los pasos de ejecución para autentificación mutua e intercambio de llave para una sesión son listados como sigue:

1) Si el usuario i desea comunicarse con el usuario j, el genera un número aleatorio r_i y calcula los siguientes 2 enteros:

$$X_i \equiv g^{3r_i} (\text{mod } n) \quad \ldots\ldots(10)$$
$$Y_i \equiv S_i \cdot \text{tiempo}_i \cdot g^{2r_i} (\text{mod } n) \quad \ldots\ldots(11)$$

Donde tiempo_i es el tiempo en que el calculo los 2 enteros.

2) El usuario i envía estos dos enteros X_i y Y_i juntos con ID_i y tiempo_i al usuario j.

3) Sobre lo recibido del mensaje, el usuario j compara tiempo_i con el tiempo local presente. Sí la diferencia entre tiempo_i y el tiempo local presente es más corta que el periodo valido, el mensaje recibido es considerado valido. De acuerdo a ambientes de red diferentes, la longitud del periodo valido puede ser ajustada.
Luego el usuario j calcula $EID_i = f(ID_i)$ y checa si la siguiente ecuación se tiene:

$$EID_i \cdot \text{time}_i^3 = Y_i^3 / X_i^2 \quad \ldots\ldots(12)$$

4) Sí la ecuación se tiene, el usuario j, cree que el mensaje es enviado por el usuario i y mantiene X_i para la generación de la posterior llave común. Luego genera un numero aleatorio r_j y calcula los siguientes 2 enteros:

$$X_j \equiv g^{3r_j} (\text{mod } n) \quad \ldots\ldots(13)$$
$$Y_j \equiv S_j \cdot \text{tiempo}_j \cdot g^{2r_j} (\text{mod } n) \quad \ldots\ldots(14)$$

5) Él usuario j envía estos dos enteros X_j y Y_j de un extremo a otros juntos con ID_j y tiempo_j al usuario i.

6) Sobre lo recibido del mensaje, el usuario i checa si tiempo_i es idéntico al que él envió. (tiempo_i aquí dentro puede ser considerado como un nonce del usuario i, el cual es solamente usado por una vez). Si es así, calcula $EID_j = f(ID_j)$ y checa si la siguiente ecuación se tiene:

$$EID_j \cdot time_j^3 = Y_j^3 / X_j^2 \quad \text{.....(15)}$$

7) Si esto es verdadero el usuario i calcula la llave de sesión K_{ij} como sigue:

$$K_{ij} = X_j^{r_i} = g^{3 r_i r_j} \quad \text{.....(16)}$$

8) En el mismo sentido, el usuario j calcula la llave de sesión K_{ji} como sigue:

$$K_{ji} = X_i^{r_j} = g^{3 r_j r_i} \quad \text{.....(17)}$$

Los usuarios i y j utilizan $K_{ij} = K_{ji}$ como la llave común de esta sesión para encriptar los mensajes de comunicación.

2.8.2 COMPUTO DE OVERHEAD

En el esquema de Tanaka y Okamoto, cada parte necesita 5 cálculos exponenciales para completar la autentificación mutua e intercambiar la llave común para cada sesión (uno para X_i, uno para Y_i, dos para chequeo de ecuación y uno para el calculo de la llave común).

En este protocolo se reduce el numero de cálculos exponenciales para cada sesión de comunicación de 5 a 2 en la fase de autentificación, primero calcula g^{r_i}, luego calcula X_i y Y_i como sigue:

$$X_i \equiv g^{r_i} \cdot g^{r_i} \cdot g^{r_i} \pmod n \quad \text{.....(18)}$$
$$Y_i \equiv S_i \cdot tiempo_i \cdot g^{r_i} \cdot g^{r_i} \pmod n \quad \text{.....(19)}$$

No hay cálculos exponenciales pero, multiplicaciones son necesarias en estas dos ecuaciones. La verificación de la identidad del emisor [ver ec. 12] puede solo ser cumplida sin cálculos exponenciales en el mismo sentido. Por lo que, este protocolo necesita solamente dos cálculos exponenciales (uno para g^{r_i}, y otro para la llave común $(X_i)^{r_i}$).

2.8.3 ANÁLISIS DE SEGURIDAD

Este protocolo proporciona encripción de mensajes y la autenticidad de las partes comunicantes para garantizar la privacía y seguridad de la comunicación de red. Este no tiene el problema de conspiración existente en el esquema de Tsujii por que su seguridad recae en la dificultad del problema de calcular los logaritmos discretos. Si un falsificador quiere enmascarar al usuario i para comunicarse con otros, este tendría que encontrar dos enteros x y y. Que satisfagan la siguiente ecuación:

$$y^3 = EID_i \cdot tiempo_i^3 \cdot x^2 \quad \text{.....(20)}$$

El uso de exponentes públicos moderado en esta ecuación disminuye la dificultad para romper (y,x). Aunque el falsificador pueda conseguir un par de enteros (y^3, x^2) que hace tener la ecuación, el par (y,x) es irrealizable por que calculando (y,x) desde (y^3, x^2) es un problema de los logaritmos discretos.

Este protocolo puede también proteger usuarios de ataques de "Hastad". Hastad propuso un ataque al utilizar RSA con bajos exponentes en una red de llave publica [58]. Para ilustrar este ataque, supongamos que un mensaje "m" es transmitido para 3 partes en la cual los exponentes públicos son $e_1 = e_2 = e_3 = 3$, y en la cual los $modulo_i$ son n1, n2, n3. Los mensajes encriptados son:

$m^3 \bmod n1$, $m^3 \bmod n2$, $m^3 \bmod n3$.

Utilizando el teorema chino de los restos, uno puede encontrar; $m^3 \bmod n1n2n3$. De cualquier modo, $m^3 < n1n2n3$ por que $m < n1,n2,n3$. Por lo que m_3 no es afectado siendo reducido modulo n1n2n3, y el mensaje puede ser recuperado tomando la raíz cúbica de m_3. Este ataque no sucede en el protocolo aquí planteado, ya que los mismos módulos n son usados para todas las partes.

No obstante que utilizan un "time stamp" para checar la llegada del mensaje, un ataque de "replay" no puede suceder en el protocolo, aún si la suposición de relojes sincronizados no exista. Considerando el siguiente escenario, un intruso escucho una sesión de comunicación, por ejemplo, la comunicación entre usuario i y usuario j. el intruso puede repetir un mensaje de autentificación viejo, capturado en una sesión vieja. Sobre lo recibido de los mensajes viejos, el usuario j checa la legalidad del $tiempo_i$. Si el tiempo de reloj del sistema esta sincronizado, el conoce que el mensaje es invalido examinando $tiempo_i$ y por lo tanto descarta este mensaje de autentificación. Si el reloj del sistema no es sincronizado, él podría considerar el mensaje. Entonces el escoge un nuevo numero aleatrorio r_j' y replica los siguientes mensajes al intruso:

$$X_j' \equiv g^{3 r_j'} \pmod n \quad \text{.....(21)}$$
$$Y_j' \equiv S_j \cdot tiempo_i \cdot g^{2 r_j'} \pmod n \quad \text{.....(22)}$$

Sin embargo, la llave común de esta sesión de comunicación es $K' = g^{3 r_i r_j'}$, en lugar de la llave común vieja $K = g^{3 r_i r_j}$. El intruso no puede calcular la nueva llave común K' sin conocer él numero aleatorio r_i. Ya que los mensajes viejos son todos encriptados por la llave común vieja K, el no podrá satisfactoriamente repetir los mensajes viejos que escucho. El usuario j puede tratar de desencriptar estos por la nueva llave común K', pero la desencripción falla. Consecuentemente, el cierra la conexión y el ataque falla.
Este protocolo también no cuenta con las dos debilidades que aparecen en el esquema de Tanaka y Okamoto.

1) Este protocolo utiliza dos pequeños números primos 3 y 2 en lugar de los dos enteros e y c del esquema de Okamoto. Debido a que no existe la gran posibilidad que e pueda ser un factor de c, la información secreta del usuario no será descubierta en este protocolo.

2) El ataque de un mensaje de autentificación falsificado fallara en este protocolo debido a la función de un solo sentido f(x). Si un usuario malicioso quiere enviar un mensaje falsificado a un usuario j. Tendrá que escoger aleatoriamente un par de números (X', Y'). Aunque un falso EID' puede ser calculado desde (20) el atacante no podrá derivar el correcto ID' desde el EID' debido a la función f(x) de un solo sentido. Sí él escoge aleatoriamente una información de identidad

ID′′, y envía esta junto con (X′, Y′), el tiempo en que él escribió el mensaje, y el mensaje falsificado, sobre lo recibido del paquete, el usuario j conseguirá EID′′ desde f(ID′′) en lugar de EID′. Consecuentemente la verificación de (12) fallara, y el usuario j regresara la solicitud falsificada, por lo que, este protocolo podrá proteger la comunicación del usuario de un ataque de solicitudes falsas.

2.9 AUTENTIFICACIÓN Y PROTOCOLO "DE ACUERDO DE LLAVE" PARA PROCESAMIENTOS EN SISTEMAS DE COMUNICACIONES MÓVILES.

Se presenta el análisis de otro protocolo, el cual combina técnicas de llave-secreta y llave-publica [59]. Este protocolo híbrido tiene las ventajas sobre otros protocolos convencionales de no estar manteniendo una base de datos de llaves secretas en las redes de PCS, y no estar teniendo que presentar complejidad y consumiendo tiempo en cálculos en el momento de la autentificación de los portátiles.

2.9.1 PROTOCOLO DE CHALLENGE-RESPONSE

En este protocolo las partes involucradas en el proceso de autentificación se asume que sean:

MS - Estación Móvil
BS - Estación Base
AC - Centro de Autentificación

1. La MS y AC comparten una llave secreta "K".
2. Para autentificarse así mismos con la red MS envía su identidad a la AC vía BS.
3. La AC busca la identidad de MS en su base de datos para obtener la correspondiente K de MS.
4. Luego la AC genera un número aleatorio (challenge). Este número y K son "hasheados" utilizando funciones hash [60, 61] para obtener la "response" esperada y la llave de sesión K_s.
5. La AC envía el número generado, la "response" esperada, y la llave de sesión K_s hacia la BS.
6. La BS reenvía el "challenge" a MS
7. La MS calcula la "response" y la K_s utilizando K y las funciones hash.
8. La MS envía la "response" calculada a BS.
9. BS compara las "responses", recibidas de AC y MS si son iguales, MS es autentificado. Después de la autentificación BS y MS utilizan K_s como llave de sesión para proteger la voz del usuario y comunicación de datos entre ellos.

Este protocolo es mostrado en la figura 2.8.

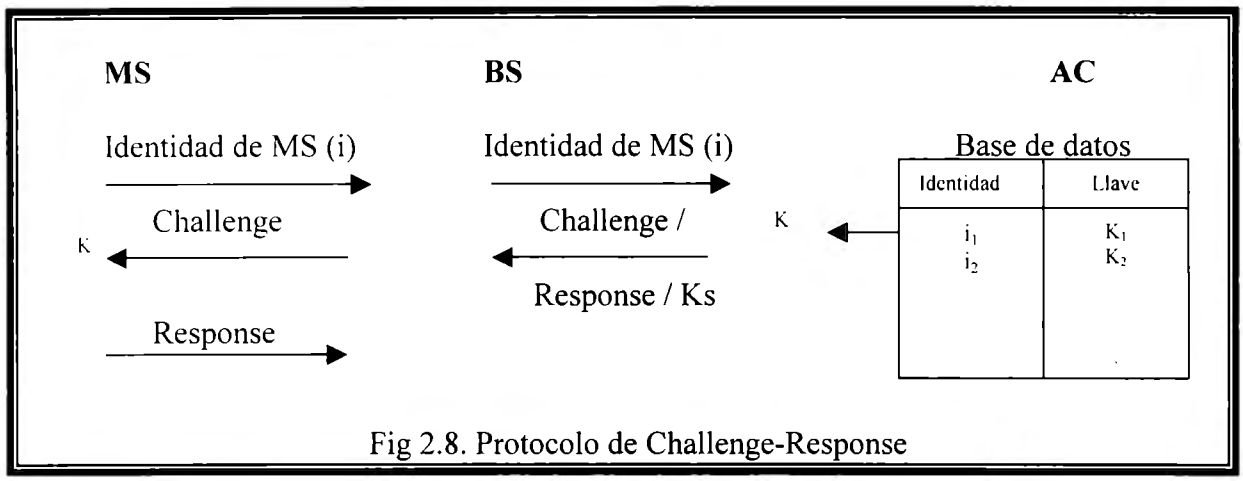


Fig 2.8. Protocolo de Challenge-Response

2.9.2 PROTOCOLOS DE LLAVE-PÚBLICA

En los protocolos de llave pública [62, 63] una autoridad de certificación (CA) es asumida para dar confianza por los MS y las redes de PCS. Los usuarios acceden a la CA con sus credenciales y alguna información de identidad acerca de sus MS's. La CA verifica la exactitud de la información proveída y firma una versión codificada de esta información [46] empleando su llave privada (figura 2.9 (a)). Esta firma es llevada al usuario como certificado. Cualquier red PCS puede checar la validez del certificado utilizando la llave publica de CA. (fig. 2.9 (b)). La CA emite certificados para las redes PCS en una manera similar. Esto permite la autentificación subsiguiente de las redes PCS por los MS's.

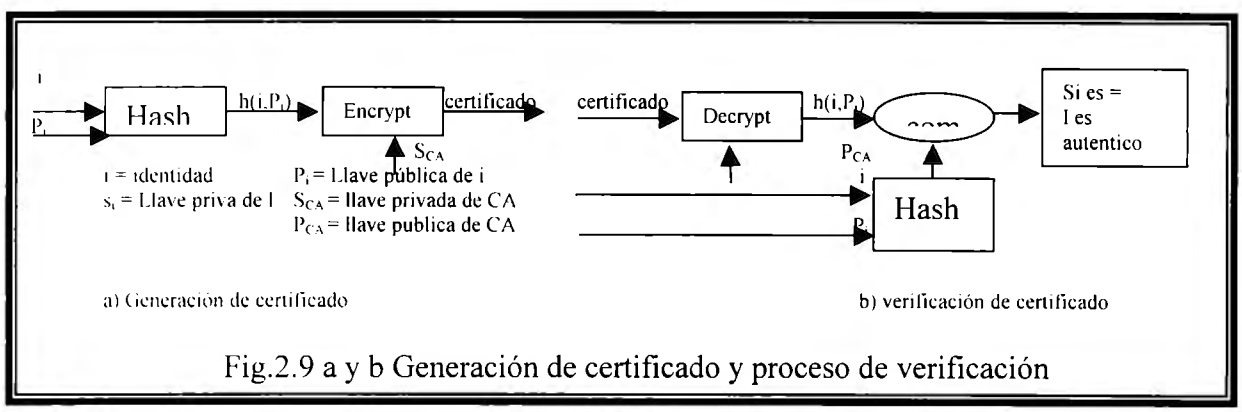


Fig.2.9 a y b Generación de certificado y proceso de verificación

2.9.3 PROTOCOLO HÍBRIDO

En este protocolo el requerimiento de una base de datos en la red para guardar llaves secretas de autentificación es eliminado por el uso de criptosistemas de llave pública, para transferir las llaves de autentificación entre MS y AC. En la MS el tiempo y complejidad que consumen los cálculos de llave pública se hacen fuera de la llamada, esto evita esos cálculos complejos en tiempo real, minimizando los requerimientos del poder de procesamiento de los portátiles.

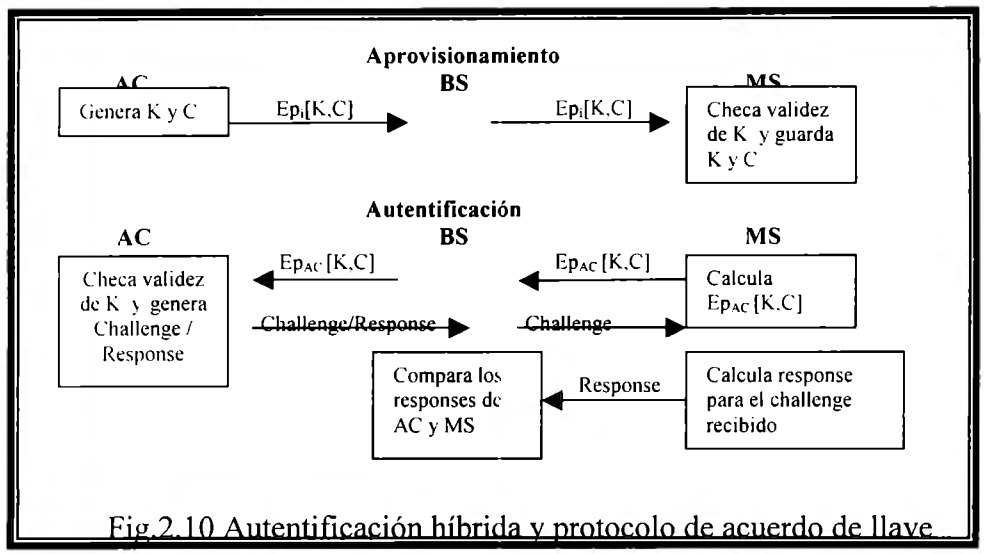


Fig.2.10 Autentificación híbrida y protocolo de acuerdo de llave

La AC provee a MS de las llaves por el protocolo de llave pública. Esto es P_i y S_i junto con la llave pública P_{AC}. Esto es proveído en el paso de aprovisionamiento, estos es, antes o en el momento de la suscripción (en el momento en que el cliente compra un microteléfono) La llave secreta (K) para el protocolo de challenge-response es generada por la AC y distribuida seguramente por aire hacia el MS junto con un certificado utilizando el método de llave pública. (Fig. 2.10). Para evitar ataques de "replay" en el aprovisionamiento de K, AC incluye un timestamp en el mensaje llevando la llave secreta K y el certificado para darle un tiempo de vida limitado para la transmisión. Cuando el MS requiere servicio desde la red, este envía la llave secreta y el certificado de regreso a la AC empleando el método de llave publica. La AC checa la validez del certificado y comienza un protocolo de challenge-response basado en la llave secreta K enviada por MS para autentificar al MS.

La generación de la llave secreta K y el proceso de distribución utilizando el sistema de llave pública es mostrado en la figura 2.11 (a).

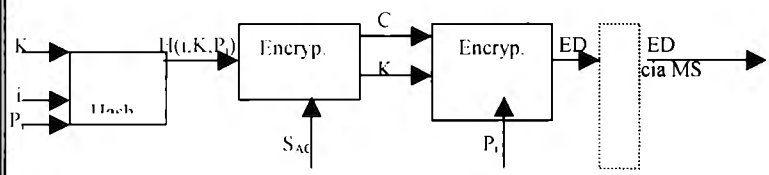
Después que AC envió K hacia MS, esta K se borra de la memoria, el MS checa la validez de K y el certificado como se muestra en la figura 2.11 (b) y guarda K y el certificado en su memoria.

Cuando el MS requiere servicio desde la red es enviado K y el certificado de regreso hacia la AC vía la BS (ver fig 2.11 (c)). Encriptando bajo la llave publica de AC (P_{AC}) con la petición de servicio. Este calculo de llave pública puede ser hecho delante del momento de autentificación y el resultado puede ser guardado en el MS para usarse en el momento de autentificación y el resultado puede ser guardado en el MS para usarse en el momento de autentificación. Esto elimina el requerimiento para que los portátiles tengan un elevado poder de procesamiento para presentar los cálculos de complejidad de llave pública en tiempo real.

Cuando AC recibe K y el certificado desde MS esta checa la validez de los dos items (fig. 2.11 (d)) y usa K para generar un challenge y un correspondiente response. Entonces la AC empieza la fase de challenge-response del protocolo por el envío del challenge-response hacia BS como en el protocolo challenge-response original para autentificar a MS.

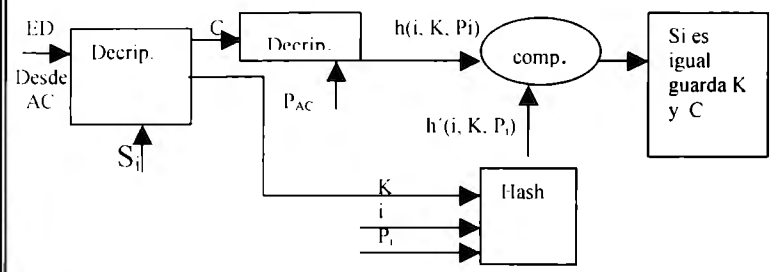
Ya que AC no mantiene un record de las llaves secretas de los móviles y del hecho que K es enviada de regreso para AC empleando el esquema de llave publica, el requerimiento de una base de datos segura para guardar las llaves secretas en la red es eliminado.

Figura 2.12 (a) . Generación de llave y distribución desde AC



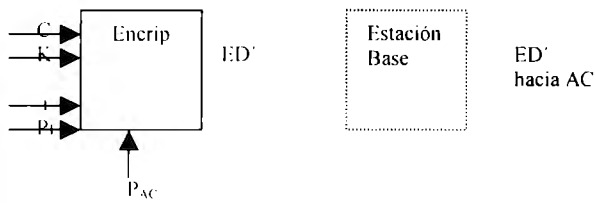
- Hash, I,K,Pi y obtiene h(I,K,Pi)
- Firma h(I,K,Pi) con S_{AC} para obtener el certificado
- Encripta C y K con P_i y obtiene ED
- Envía ED hacia BS

Fig. 2.12 (b) Proceso de recuperación de llave en el MS



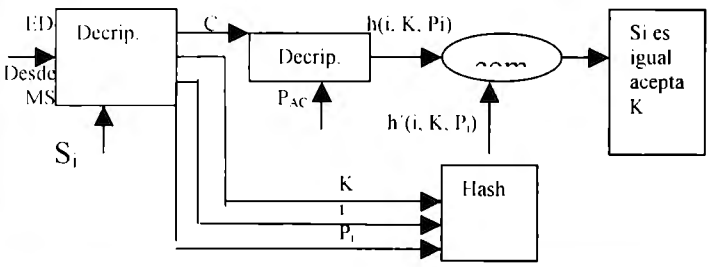
Decripta ED para obtener C y K empleando S_i
Decripta C pars obtener h(i, K,P_i) utilizando P_{AC}
recibe el hash i,K,P_i para obtener h'(i, K,P_i)
Compara la h calculada y la h recibida
Si las dos son iguales guarda K y C.

Fig. 2.12 (c) Proceso de transferencia de llave desde MS.



- Concatena i, K, P_i, y C
- Encripta el mensaje concatenado con P_{AC} para obtener ED'
- Envía ED' hacia el Centro de Autentificación.

Fig 2.12 (d) Proceso de recuperación de llave en el AC



- Decripta ED' para obtener i,K,Pi y C empleando S_{AC}
- Decripta C para obtener h(i, K,P_i) utilizando P_{AC}
- Aplica hash a i,K,P_i para obtener h'(i, K,P_i)
- Compara la h calculada y la h recibida
- Si las dos son iguales acepta K.

Comentarios

En resumen, este análisis cuenta con las siguientes características.

- Eliminación de base de datos de llaves secretas
- K se puede enviar por medios inseguros "aire" que no se hacía en los protocolos originales challenge-response.
- Periódicamente CA puede cambiar K. (incrementando seguridad)
- También CA puede cambiar las llaves de MS (S_i , P_i) Esto puede ser enviado hacia el MS similar que como lo hace para K en la figura 2.11 b.
- Es teóricamente posible cambiar llaves de AC (P_{AC} y S_{AC}). Prácticamente esto comienza una dificultad ya que todos los MS soportados por las PCS deben ser provistos con nuevos certificados firmados utilizando la nueva S_{AC}.
- Mejora la complejidad computacional de llave publica en tiempo real al momento de la autentificación en el portátil. Además, el MS no requiere poder de procesamiento alto. Pensando que la AC requiere poder de procesamiento alto para realizar los cálculos de llave pública en tiempo real, es importante notar la red puede ser equipada con sofisticados procesadores DSP de alta velocidad dedicados a estas tareas.

2.10 PROTOCOLO DE AUTENTIFICACIÓN PARA USUARIOS DE "ROAMING" EN REDES GSM.

En esta sección, se estudian y analizan los protocolos de autentificación de GSM para usuarios de "roaming", así como revisar la propuesta de un otro esquema con menos trafico de señalización y mejor tiempo de establecimiento de llamada [64].

2.10 .1 PROTOCOLO DE AUTENTIFICACIÓN DE GSM PARA USUARIOS DE ROAMING

El protocolo de autentificación de GSM para usuarios de roaming es realizado a través del mecanismo "challenge/response" el cual consiste de hacer una pregunta que solamente el equipo de usuario correcto (SIM) puede contestar. La respuesta regresada, calculada internamente en la tarjeta SIM del usuario, será comparada en el centro de autentificación [62,46,65,48]. De manera más precisa, un número aleatorio Rand es enviado y la respuesta esperada, llamando al resultado firmado (SRES) es regresado. El Rand es generado localmente por HLR/AuC y luego combinado con la llave (K_i) secreta del usuario a través del algoritmo (A_3) para conseguir el SRES. Cada momento que una estación móvil es autentificada, la red y el MS tienen que calcular la llave cifrada K_c la cual es empleada para estar cifrando y descifrando los datos transmitidos. Los cálculos de la llave K_c para cifrar es básicamente similar a la SRES, utilizando el algoritmo A8. La estación móvil esta escuchando continuamente la identidad del área de la localización (LAI) que es transmitida en el canal de difusión comparando la nueva LAI recibida con la ultima LAI (guardada en el SIM). Siempre que el LAI recibido sea diferente al LAI viejo guardado en su tarjeta SIM, el MS procede con un registro nuevo. Primero empieza el registro con conseguir acceso ha el canal de control dedicado independientemente (SDCCH) sobre el cual el BTS y el MS se comunican el uno con el otro.

Los siguientes pasos resumen el protocolo de autentificación.

PASO 1:
Por la comparación de la LAI recibida desde la red visitada y la LAI de la red local, el MS detecta que se ha introducido en un área de red nueva, así el MS transmite una solicitud de registro (actualización de la localización) hacia la estación base sobre el SDCCH, el BTS reenvía la solicitud de registro hacia la MSC la cual informa al correspondiente VLR acerca de esta solicitud la solicitud de registro incluye la identidad del subscriptor móvil temporal (TMSI) y la LAI.

PASO 2:
Analizando la TMSI, el VLR entiende que este usuario es una estación móvil de roaming (RMS). La red visitada no tiene la capacidad de autentificar a este RMS. Por lo tanto, el dominio local tiene que ser contactado para el proceso de autentificación. El TMSI contiene información para el código de país de dominio local como se muestra en la figura 2.12

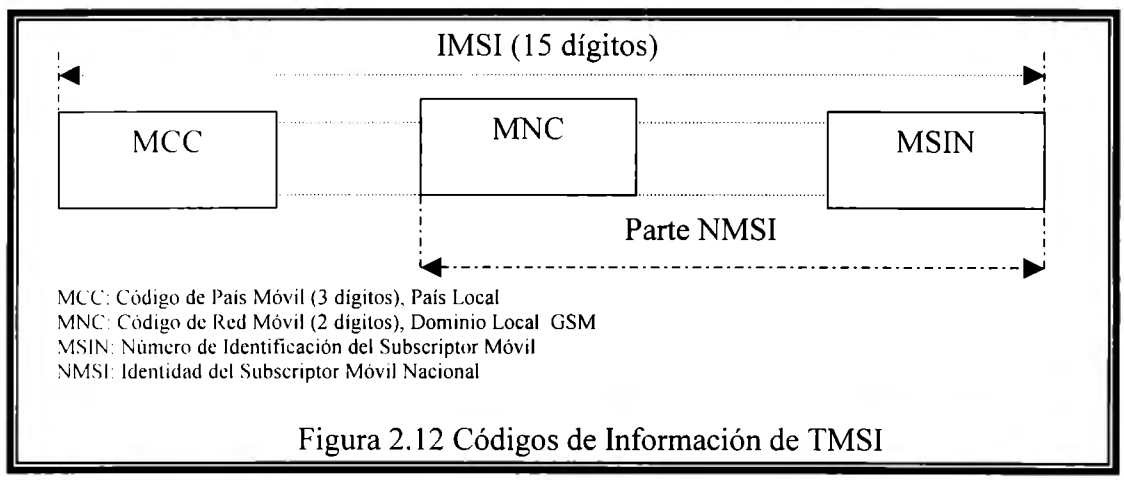


Figura 2.12 Códigos de Información de TMSI

PASO 3:
La VLR, a través de MSC hace una solicitud para el dominio local preguntando por la tripleta de autentificación (RAND, SRES, K_c) de este RMS.

PASO 4:
La VLR en el dominio local reenvía la solicitud hacia la AuC a través de HLR.

PASO 5:
La AuC calcula SRES y K_c. Aplicando la llave privada K_i de MS y un número RAND para los algoritmos A3 y A8. La VLR a través de MSC en el dominio local, envía la tripleta de autentificación (RAND, SRES y K_c) hacia la red visitada.

PASO 6:
Sobre la recepción de la tripleta, el VLR en la red visitada envía la RAND hacia el MS por el MSC y pide el MS para calcular la SRES y enviarla de regreso.

PASO 7:
La MS calcula la SRES y la K_c localmente empleando ese número RAND y la K_i a través de los algoritmos A3 y A8, luego envía SRES de Regreso hacia el VLR y mantiene K_c para uso posterior.

PASO 8:
La VLR una vez que recibe el SRES desde el MS compara este con el SRES proporcionado desde la AuC del dominio local si las dos son iguales, la MS pasa el proceso de autentificación.

La piedra angular en este concepto es la llave secreta, la cual nunca es transmitida en el aire o dada a ninguna otra parte. El intercambio de mensajes de señalización entre las entidades de red es ilustrado en la figura 2.13 Se puede derivar el número de mensajes intercambiados entre elementos de red y en particular los mensajes de "crossnetwork" (inter MSCS).
Estos mensajes de señalización agregan un sobreflujo extra para dominios locales así como la red visitada. Además de esto, los usuarios de "roaming" serán sacados del servicio (puede no hacer o recibir llamadas) en caso de falla de cualquiera de las bases de datos de la red local o de los enlaces de inter-redes en horas pico de tráfico, donde los "trunks / circuits" entre el local y las redes visitadas son completamente utilizados, los usuarios de roaming pueden no tener beneficio desde sus servicios PCS para los cuales ellos se suscriben.

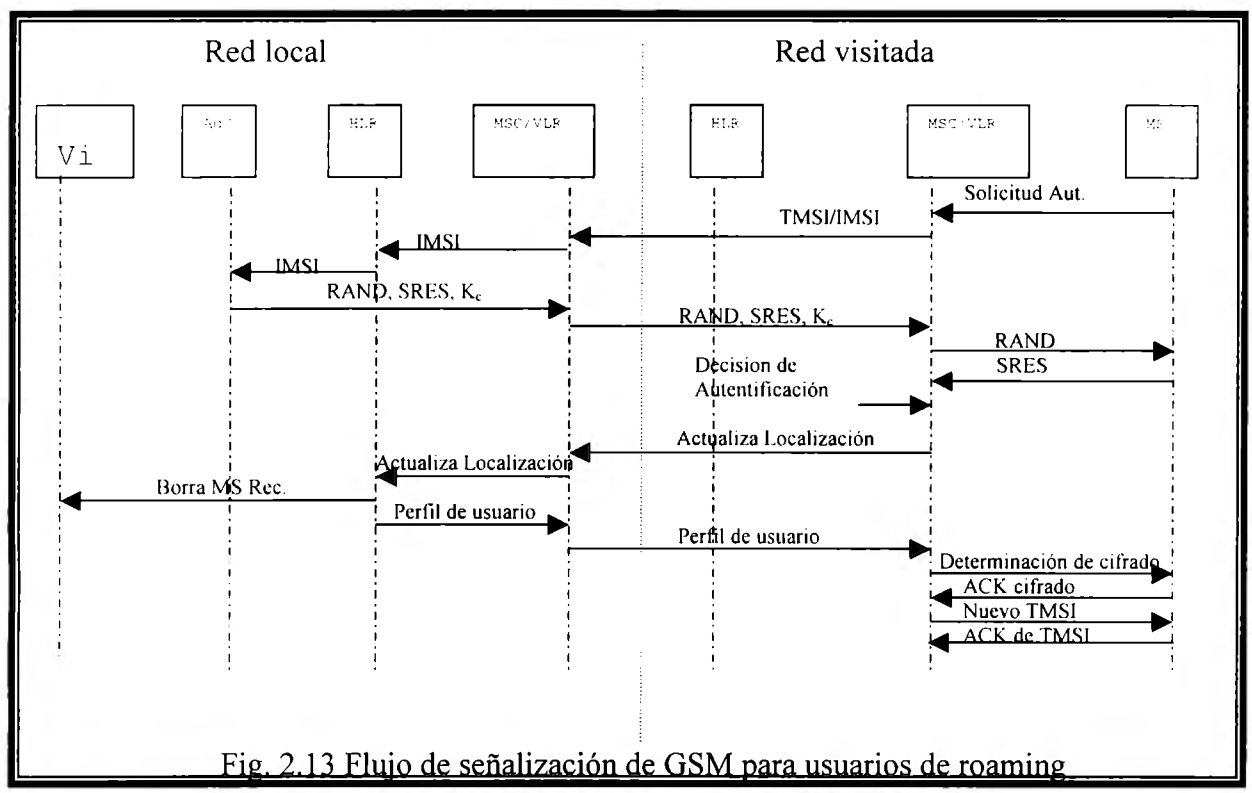


Fig. 2.13 Flujo de señalización de GSM para usuarios de roaming

2.10 .2 Protocolo de autentificación propuesto para usuarios de roaming

En el acercamiento de GSM, K_i existe solamente con la red local, la red visitada tiene que contactar la red local para cualquier tentativa de autentificación del usuario. Este contacto da una serie de mensajes de señalización importante y sobrecargando las entidades de red. Las tripletas

de autentificación (SRES, K_c, RAND) proporcionadas por la red local son enviadas hacia la red visitada a través de enlaces de red interna. Los mensajes de señalización de red cruzados incrementan los retardos de autentificación y eventualmente los tiempos de levantamiento de llamada [66,67]. Este esquema apunta a reducir los mensajes de señalización de red cruzados y para decrementar el tiempo de establecimiento de llamada. Esta basado en la idea de compartir la llave privada (K_i) del usuario. Con las redes visitadas, y el uso de un contador de registro de acontecimiento (COUNTM). El acercamiento de la propuesta es útil para usuarios con alta movilidad, así como, frecuentes movimientos entre sus dominios locales y otras redes visitadas dentro del mismo país. La red visitada podrá autentificar a estos usuarios de roaming. Sin ir detrás de su red local para colectar la tripleta de autentificación.

Se asume que el RMS ha guardado en la tarjeta SIM todos los códigos de red móvil (MNC) compartidos la llave privada K_i con la red local. El RMS recibe la (LAI) de la red visitada sobre el cruce de limite, del área geográfica de red local. En LAI se encuentra el MNC como se muestra en la figura 2.14.

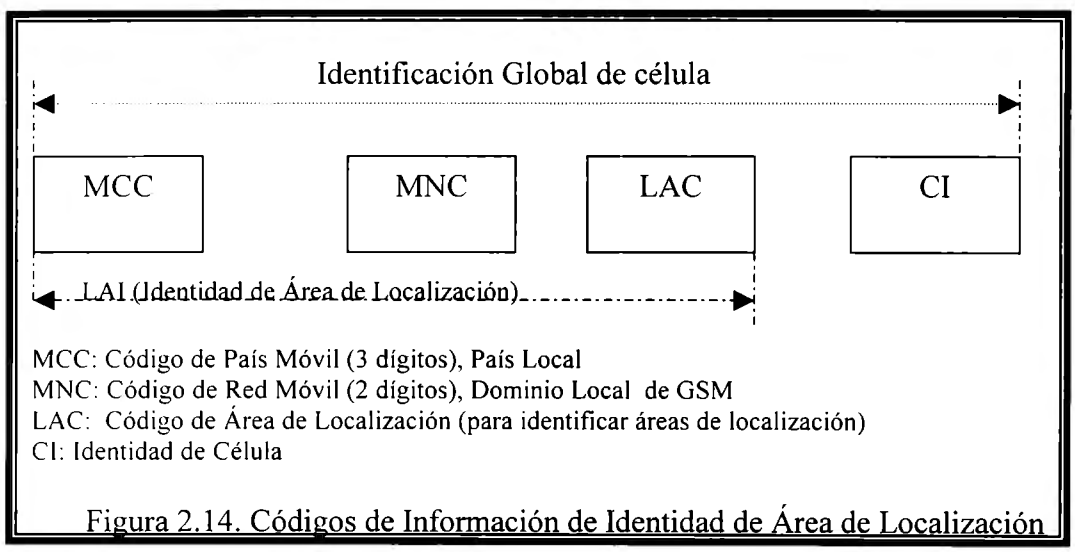


Figura 2.14. Códigos de Información de Identidad de Área de Localización

La tarea de información a la red acerca de las llaves de usuario que son compartidas, es llevado a cabo por un campo en el mensaje de solicitud de servicio llamado, compartiendo la llave (SHK). La SHK es puesta en "1" si la llave es compartida. El COUNTM será enviado a la red visitada para el primer registro solamente y para ser usado para los cálculos de la AUTHR. La red visitada continuara utilizando el COUNTM como un contador de registro de acontecimiento para el usuario de roaming y este será borrado sobre el registro del RMS. El ATHR es la salida del algoritmo A3 cuando TANDG y K_i son las entradas. RANDG es la concatenación de un número aleatorio RANDM generado por el MS y el COUNTM como se muestra en la figura 2.15.

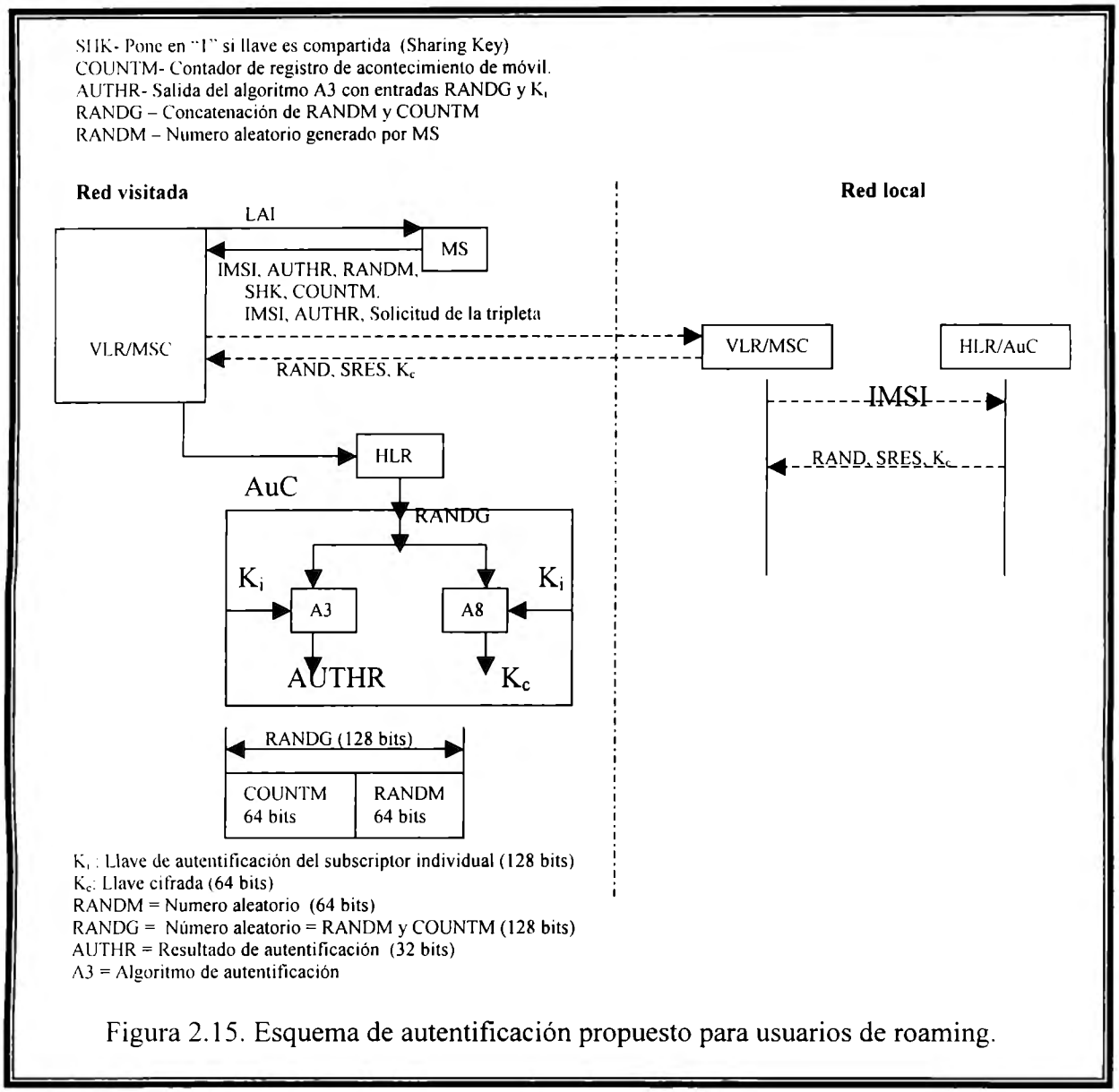


Figura 2.15. Esquema de autentificación propuesto para usuarios de roaming.

Los pasos del esquema son como se muestran a continuación:

PASO 1:
Por la comparación del LAI recibido desde la red visitada y LAI de la red local, EL MS detecta que este ha entrado a una nueva área de red, así, el MS transmite una solicitud de registro (actualización de localización) para la estación base sobre la SDCCH. La BTS reenvía la solicitud de registro hacia el MSC el cual informa a la correspondiente VLR acerca de esta solicitud. La solicitud de registro incluye el TMSI, LAI, AUTHR, RANDM, COUNTM, y SHK.

PASO 2:
Analizando el TMSI, El VLR entiende que este usuario es una estación móvil vagando, luego este checa la SHK, si esta es "1" o no. Si SHK es "1", la VLR interpreta como una solicitud del registro de un RMS que llave K se comparte, luego reenvía el conjunto completo hacia la HLR.

PASO 3:
La HLR guarda la CONTM, luego reenvía el mensaje de solicitud hacia el AuC.

PASO 4:
La AuC produce la RANDG, luego calcula AUTHR y K_c por aplicación de la llave K_i, del MS y el número RANDG en los algoritmos A3 y A8 respectivamente. Finalmente, la AuC compara las dos AUTHRs, si las dos son iguales la MS pasa el proceso de autentificación.

Cuando se identifica que la red visitada no esta compartiendo la llave del RMS, el campo SHK es puesto en "0". El VLR solicitara la tripleta de autentificación desde la red local para proceder como en el proceso de autentificación de GSM normal.

Los requerimientos de seguridad aún se mantienen ya que la llave privada nunca es transmitida en el aire y solamente es conocida por las redes (local y visitada) es esquema propuesto es construido en los mismos principios de seguridad de GSM. En GSM, se tiene que las SRES son calculadas en la AuC y en la tarjeta SIM, Luego la VLR compara ambas de estas. El usuario es autentico si ambas SRES son iguales. Para el esquema propuesto, una AUTHR es calculado en el MS y en la AuC, empleando el número aleatorio RANDG y K_i del usuario. Luego la red compara ambos. La MS es autentica si la AUTHR es igual. En las redes GSM, tenemos SRES, RAND, K_c, son enviadas hacia la red visitada, también SRES, RAND, son enviadas para el usuario en el aire, y ellos son considerados como llaves publicas. Para el protocolo propuesto, K_c nunca es enviado por el aire, la AUTHR, y RANDM propuesto, K_c nunca es enviado por el aire; la AUTHR y RANDM son llaves publicas solamente. K_i , K_c son todavía mantenidas como llaves privadas y nunca transmitidas por el aire.

Este protocolo genera menos mensajes de señalización que el propuesto por GSM. Los flujos de mensajes de señalización son ilustrados en la figura 2.16. La tabla 2.2 nos da una reseña del número de mensajes de señalización por solicitud de autentificación.

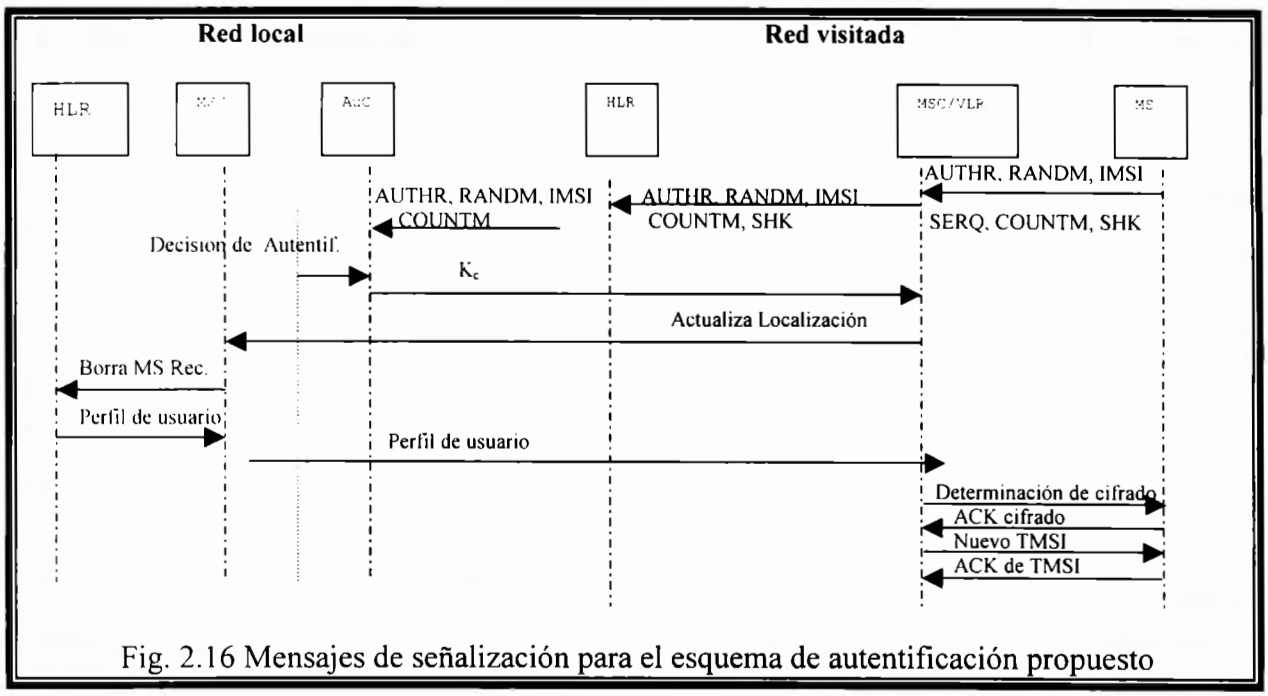


Fig. 2.16 Mensajes de señalización para el esquema de autentificación propuesto

Tabla 2.2

	Red vieja			Red visitada			Inter MSC
Actividad	**AuC**	**HLR**	**VLR/MSC**	**AuC**	**HLR**	**VLR/MSC**	
Registro	2	2	4	0	4	5	2
Origen de llamada	2	2	4	0	4	5	2
Terminación de llamada	2	2	4	0	4	5	2

2.11 ANÁLISIS DE SEÑALIZACIÓN DE TRÁFICO PARA AUTENTIFICACIÓN Y PROTOCOLOS PRIVADOS EN SISTEMAS DE COMUNICACIÓN PERSONAL.

Una revisión de los mecanismos de seguridad para sistemas de telefonía celular, los sistemas PCS y una descripción del flujo de señalización para autentificación y protocolos privados es presentada. [68] Vía un modelo de movilidad para sistemas celulares se hace un análisis de señalización de trafico para estos protocolos.

2.11.1. INTRODUCCIÓN

Los servicios de celular y PCS se han convertido en uno de los sectores de negocio de más rápido crecimiento de las industrias de las telecomunicaciones. Como la industria inalámbrica (de telefonía celular principalmente) ha crecido, la necesidad de seguridad también ha incrementado. Tanto para privacidad como para prevención de fraudes[16].

En 1991 el grupo de trabajo TR-45 dentro de la Asociación de Industrias de Telecomunicaciones (TIA) desarrollo el estándar IS-41 para operaciones íntersistemas en redes celulares. Mas tarde, en 1996 basado en la especificación del boletín de sistemas técnicos 51, (TSB-51) por TIA [69]. El IS-41 fue actualizado por la IS-41 revisión C. Ahí fueron definidos protocolos para autentificación de estaciones móviles, encripción de voz, y mensajes para sistemas de celular americanos.

Todos estos protocolos generaban una cierta cantidad de cargas de trafico de señalización, incluyendo aquellas que resultaban de la localización, "handoff", registro y otras transacciones asociadas con bases de datos de red. Los cálculos de estas cargas son importantes para determinar el grado de trafico de control sobre los canales de radio o a través de varios componentes de redes (bases de datos, switches, etc.)

2.11.2. EL MODELO DE REFERENCIA PARA REDES DE PCS Y CELULAR

Los operadores de red actuales de sistemas celular y PCS adoptan el esquema de "rehúso de frecuencia" para soportar un gran numero de subscriptores. Como resultados ambos sistemas de comunicación presentan arquitecturas similares. Bajo los modelos de referencia para celular y la

[16] La CTIA (Cellular Telecommunications Industry Association) estima que la industria de celular pierde mas de 2 billones de dolares por año en fraudes en los Estados Unidos de Norte america. [1] En 1997 los carriers de celular Brasileños perdieron alrededor de 7 millones por mes. [2]

mayoría de redes PCS hay una Estación Base (BS) instalada en cada célula, y Estaciones Móviles (MSs) dentro de una célula comunicándose con la BS a través de un enlace inalámbrico. Las MSs son terminales móviles utilizadas pos los suscriptores para enviar y recibir información. Las BSs determinan los protocolos de radio utilizados para las comunicaciones con el MS. Los BSs están conectados a un Centro de Conmutación Móvil (MSC) a través de troncales inalámbricas o líneas alámbricas. Un MSC ejecuta algunas funciones de conmutación típicas y coordina con la BS localizaciones, registros y llamadas perdidas. La MSC mantiene conexiones con otras MSCs y con la Rede de Telefonía Conmutada Pública (PSTN.) Los esquemas actuales para manejo de movilidad están basados en dos niveles de jerarquía de datos, el Registro de Localización Local (HLR) y el Registro de Localización de Visita (VLR) En general hay un HLR por cada red celular e información respecto a cada usuario, tal como, tipos de servicio suscritos, información de la factura, e información de localización son guardados en el perfil del usuario localizado en el HLR. Y Hay VLRs por cada MSC, y cada VLR guarda la información del área visitada por el MS (obtenida del HLR.) Los VLRs son empleados en conjunción con los HLRs para soportar el "roaming" automático. El Centro de Autentificación (AC) contiene las llaves secretas y otra información requerida para autentificar al MS y proporcionar de privacidad de voz.

Los esquemas actuales para el manejo de la movilidad están basados en una jerarquía de 2 niveles de datos, el registro de localización local (HLR) y registro de localización del visitante (VLR.)

2.11 .3. IS-41 REV. C.: AUTENTIFICACIÓN Y PROTOCOLOS DE PRIVACÍA.

El algoritmo para autentificación y la generación de mascara de privacía de voz y las llaves de encripción de mensajes de señalización empleados por IS-41 REV. C esta basado en técnicas criptográficas de llave privada en las cuales una llave secreta es solamente compartida entre el MS y la red.

Los protocolos de autentificación y privacidad en IS-41 utilizan los siguientes parámetros:

A). **A-Key**. Número de 64-bits asignado por el proveedor de servicio para el subscriptor sobre el servicio de inicialización. La A-key debe ser introducida dentro del MS y también conocida por la HLR de la red celular. La A-key nunca es usada o expuesta directamente en un proceso de autentificación.

B). **SSD** La SSD (Dato Secreto Compartido) es un número de 128-bits guardado en el MS (en memoria semi-permanente) y fácilmente disponible para la estación base. Es calculada por ambos el MSC y el MS desde un grupo de variables (MIN, ESN, A-key) siendo pasada por un algoritmo CAVE. La SSD es comparada entre el MSC y el MS para determinar la autenticidad de este. Hay que notar que el SSD autentifica al usuario en ambos ambientes "local" y "roaming".

C). **CAVE** (Encripción de Voz para Autentificación Celular) Presenta un algoritmo de autentificación por la combinación del RAND desde MSC celular con información de MS. Si el resultado que es calculado por el MS coincide con el resultado producido por el MSC, entonces el MS será considerado a ser autentico.

D). **RAND.** Este es un número aleatorio de 32-bits sostenido en la estación MS. El valor del RAND es difundido por el MSC, en el canal de control, y actualizado periódicamente.

E). **ESN.** El ESN (Número serial electrónico), es un número binario de 32 bits que únicamente identifica al MS para cualquier red celular. Debe ser fabricado cuidadosamente y no debe ser fácilmente alterable en el campo. La modificación de ESN requerirá un recurso especial que no es normalmente disponible para los suscriptores.

F). **MIN.** Número de Identificación del móvil es un número binario de 34-bits el cual es derivado de un número de teléfono de directorio de 10 dígitos. Un MIN es un numero del Plan de Numeración Norte Americano (NAMP) que es guardado en el MS al momento de la construcción y no puede ser cambiada.

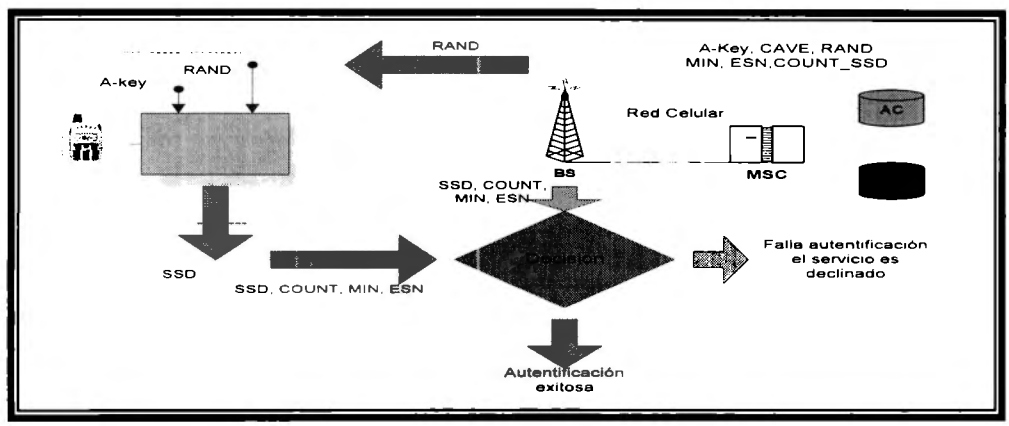
G). **COUNT**. Es un contador de llamada de 64-bits y sostenido en el MS. El IS-41 emplea un COUNT para la protección contra la duplicación o clonación general del MS. El COUNT es incrementado en el MS sobre un comando desde la red, generalmente durante una llamada. La red celular también mantiene el contador. Más adelante, durante una tentativa subsiguiente del acceso el MS envía su contador de llamada de regreso a la red. Si múltiples MSs están compartiendo una identidad, la red acumulará una cuenta que exceda probablemente el del utilizador legítimo. Una vez que un clon es sospechado, el personal de la red puede intervenir.

2.11.3.1 El SSD compartido y el SSD no compartido.

La HLR y el MS comparten tres elementos que crean el dato secreto: MIN, ESN, y A-Key. La red utiliza dos tipos de esquemas para esto:
El SSD compartido y el SSD no compartido. Si el SSD es compartido, esto significa que el HLR enviará este al MSC donde el MS esta vagando. Si el SSD es no compartido, el HLR mantendrá el dato para el mismo.
El "challenge" para el procedimiento de autentificación.

Los procesos de autentificación consisten de un diálogo de "challenge-response" entre el MSC o HLR y el MS. Con SSD compartido, el challenge es publicado por el HLR. Existen dos tipos de challenge:

El challenge único
El challenge global [70, 71]



En la figura 2.17 Muestra del proceso para la autentificación.

2.11.3.2 Flujo del mensaje

En esta sección se muestra un detalle del flujo del mensaje de los protocolos debido al registro, al inicio de llamada y terminación de llamada, para ambos SSD compartido y SSD no compartido.

Flujo del mensaje debido al registro cuando SSD es compartido.

El MS, basado en la señal transmitida por la BS, determina que ha introducido una nueva área de registro RA (nueva MSC) y esa autentificación es requerida para tener acceso nuevamente. La figura 2.18 ilustra el flujo de los mensajes y los pasos de abajo describen este procedimiento:

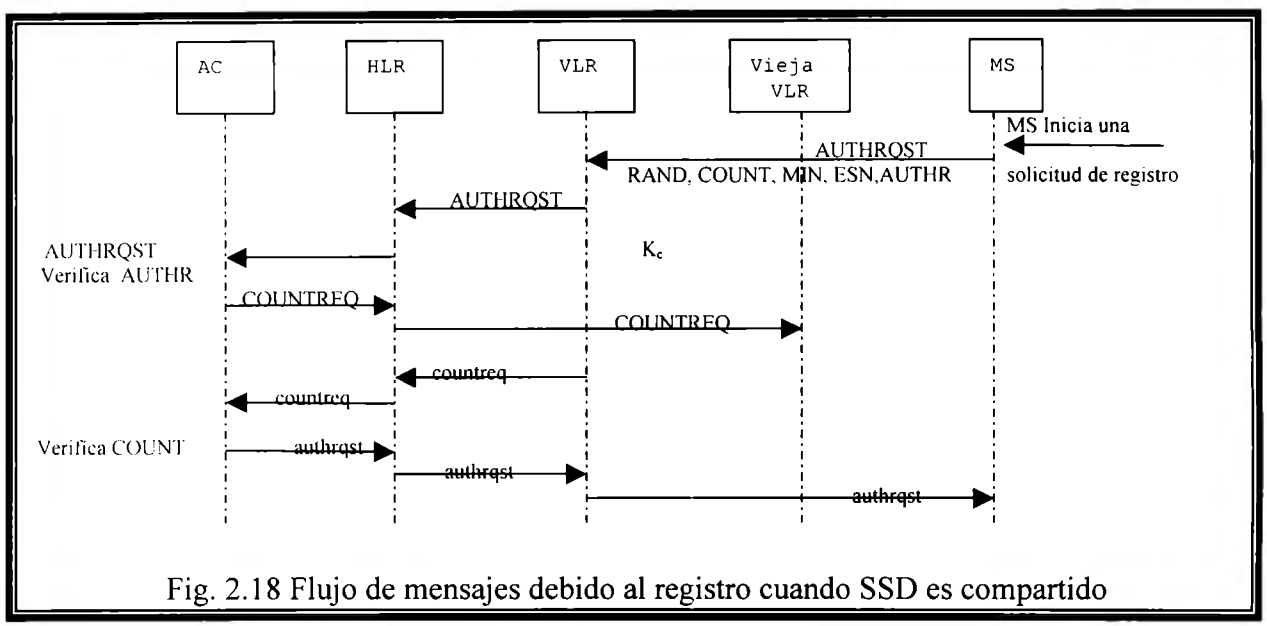


Fig. 2.18 Flujo de mensajes debido al registro cuando SSD es compartido

1) El MS determina que ha entrado una nueva RA y una autentificación es requerida para tener acceso. Se ejecuta el algoritmo CAVE empleando el SSD y sus ESN, MIN y RAND obtenidos de la MSC visitada en ese momento. El algoritmo produce un AUTHR (Resultado de Autentificación de Registro).
2) El MS solicita registro con la nueva MSC proveyendo AUTHR, su ESN, MIN, RAND y COUNT.
3) El nuevo MSC remite la solicitud de autentificación en un mensaje AUTHRQST para la VLR que sirven al nuevo RA.
4) El VLR remite la solicitud a la HLR junto con todos los parámetros recibidos
5) En el turno de la HLR remite la solicitud de autentificación a la AC.
6) La AC extrae de su base de datos el SSD asociado con el MIN y, utilizando el SSD extraído, ejecuta el algoritmo CAVE con los parámetros adicionales recibidos desde la HLR, (MIN, ESN, RAND) produciendo el resultado de autentificación.
7) En este punto la AC ha completado la verificación del AUTHR generado por el MS. Sin embargo, no tiene el actual valor del COUNT y por lo tanto no puede verificar qué es provista por el microteléfono en su petición del registro. La AC necesita conseguir el valor actual desde la VLR de la MSC visitada previamente.

8) La AC envía un mensaje COUNTREQ a HLR, dirigido a la VLR visitada previamente. Solicitando el valor actual del COUNT.
9) La HLR remite la solicitud al VLR para el microteléfono al cual está señalando, del cual el que está para el RA del MS que acaba de moverse realmente.
10) La VLR responde a la solicitud enviando COUNT en un mensaje countreq.
11) La HLR remite este a la AC
12) La AC verifica el count y envía un AUTHR en un mensaje authrqst para el sistema solicitante.

Flujo de mensaje debido a la creación de la llamada cuando SSD es compartido.

La figura 2.19 muestra el flujo de mensaje de señalización cuando un MS en el sistema visitado inicia una llamada.

1) El MS ejecuta el algoritmo CAVE utilizando el SSD y sus ESN, MIN y RAND obtenidos de la MSC visitada en aquel momento. El algoritmo genera AUTHR, SMEKEY y VPMASK, y envía RAND, AUTHR, COUNT, ESN, MIN y los dígitos marcados hacia la MSC visitada.
2) La MSC visitada envía un mensaje de solicitud de autentificación AUTHRQST para la porción VLR.
3) La VLR ejecuta el algoritmo CAVE y genera AUTHR, VPMASK y SMEKEY.
4) Después de verificar AUTHR y COUNT, incluye el resultado de la verificación junto con VPMSK y SMEKEY en un mensaje de respuesta authrqrst para el sistema visitado.

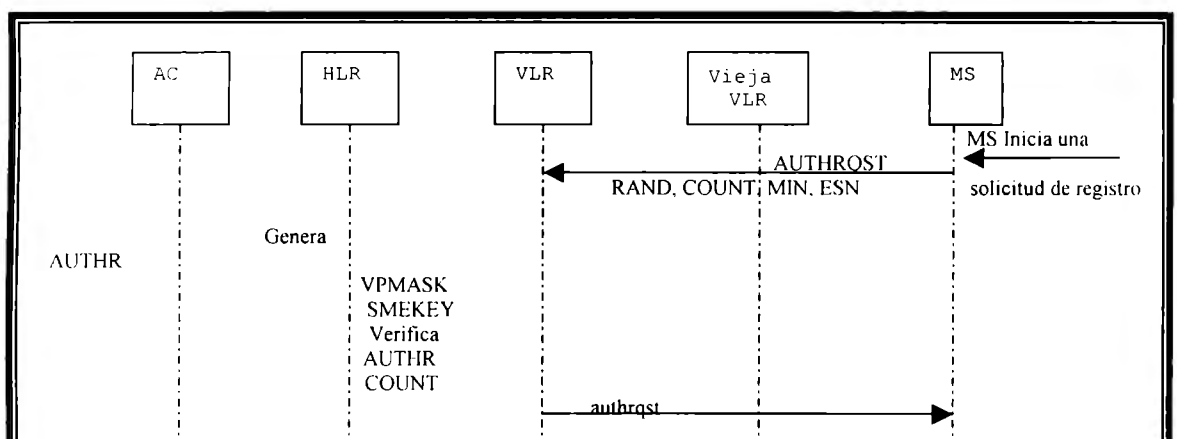


Fig. 2.19 Flujo de mensajes debido a la creación de llamada cuando SSD es compartido

Flujo de mensajes debido a la terminación de la llamada cuando SSD es compartido.

Después de la localización del MS es alcanzada por el IS-41 el flujo de mensaje para la autentificación y privacía es similar al de creación de llamada. Una vez que el MS llamado es autentificado con éxito, un canal de voz es establecido para la llamada entre los dos MSs.

Utilizando las figuras 2.18 y 2.19 podemos contar el numero de mensajes para cada una de las bases de datos (AC, HLR, y VLR, VLR nueva) para el registro, creación y terminación, estos resultados se resumen en la tabla 2.3.

Tabla 2.3. Mensajes de autentificación y privacía cuando SSD es compartido.

Procedimiento	AC	HLR	VLR	Old VLR
Registro	2	4	2	1
Origen	0	0	1	0
Terminación	0	0	1	0

Flujo del mensaje debido al registro cuando SSD no es compartido.

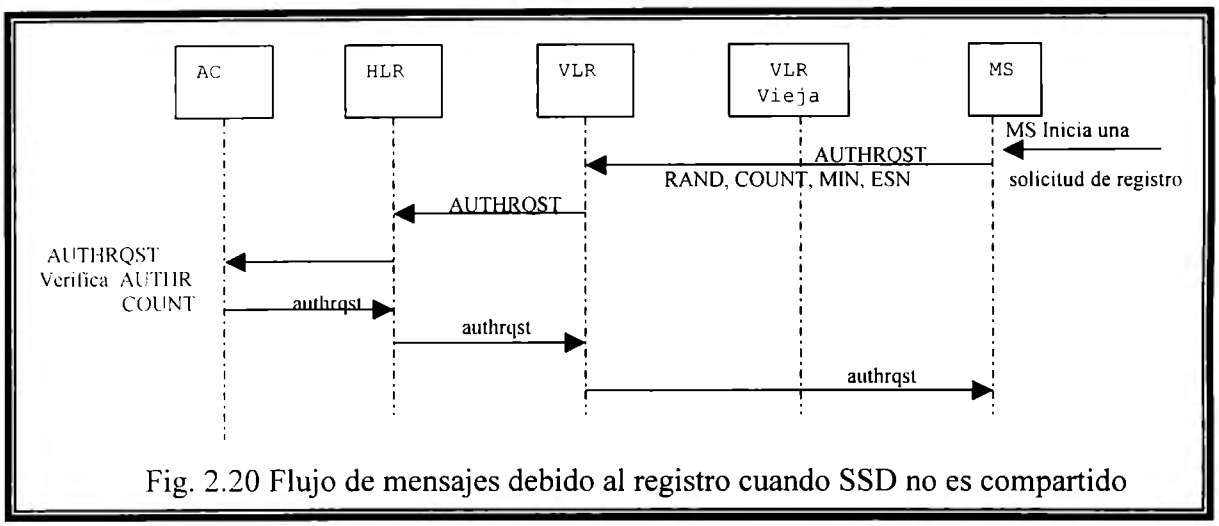


Fig. 2.20 Flujo de mensajes debido al registro cuando SSD no es compartido

Flujo de mensaje debido a la creación de la llamada cuando SSD no es compartido.

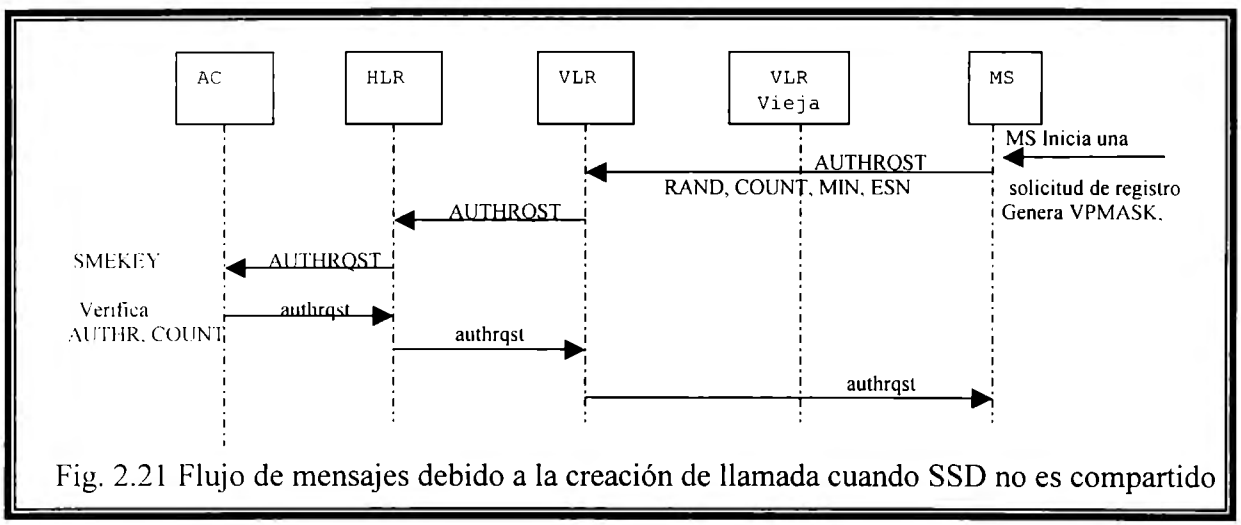


Fig. 2.21 Flujo de mensajes debido a la creación de llamada cuando SSD no es compartido

Flujo de mensajes debido a la terminación de la llamada cuando SSD no es compartido.

Después de que la localización del MS es alcanzada por el IS-41 el flujo de mensaje para la autentificación y privacía es similar al de creación de llamada. Una vez que el MS llamado es autentificado con éxito, un canal de voz es establecido para la llamada entre los dos MSs.

Utilizando las figuras 2.20 y 2.21 se puede contar él numero de mensajes para cada una de las bases de datos (AC, HLR, y VLR, VLR nueva) para el registro, creación y terminación, estos resultados se resumen en la tabla 2.4.

Tabla 2.4 Mensajes de autentificación y privacia cuando SSD no es compartido.

Procedimiento	AC	HLR	VLR	Old VLR
Registro	1	2	2	0
Origen	1	2	2	0
Terminación	1	2	2	0

2.11.4. MODELO DE TRÁFICO

Los modelos tradicionales de tráfico han sido desarrollados para redes alambricas. Estos modelos predicen él trafico agregado que ira a través de los conmutadores telefónicos. Como tal, ellos no incluyen la movilidad de los suscriptores y por lo tanto necesita modificaciones para que sea aplicable para el modelo de trafico celular. Tres modelos son desarrollados para estimar este trafico. Los modelos se ocupan del comportamiento promedio de un suscriptor como los acercamientos comunes para modelar los movimientos humanos [66].

2.11.4.1 Modelo de Markovian

Es el modelo del caminar-aleatoriamente describiendo comportamiento del movimiento individual donde un suscriptor sigue estando dentro de una región o se mueve la voluntad a la región adyacente según una distribución de la probabilidad de la transición.

2.11.4.2 Modelo de Gravity

Ha sido utilizado ampliamente en investigación de la transportación para modelar el comportamiento del movimiento humano. Hay muchas variaciones y han sido aplicados a regiones de tamaños variables, desde modelos de ciudades hasta modelos nacionales e internacionales.

2.11.4.3 Modelo de fluido [72,73]

Conceptualiza el flujo del trafico como el flujo de un fluido y acostumbra utilizar un modelo macroscópico del comportamiento del movimiento. Este modelo asume que los subscriptores que llevan microteléfonos se mueven a una velocidad promedio "v", la gente es esta uniformemente distribuida en una área "A" y la dirección del viaje de cada suscriptor relativo al limite es uniformemente distribuida en (0, 2π.) Si ρ es la densidad de los suscriptores por kilómetro cuadrado y L es la longitud del perímetro del área A, entonces el numero promedio de estaciones móviles que dejan A por segundo es dado por:

MSs dejando un área A/s = $\rho vL/3600\,\pi$

2.11.5. ANÁLISIS DEL TRAFICO GENERADO.

Utilizando el modelo basado en flujo en la ecuación 1 podemos analizar y calcular el trafico de señalización debido a los protocolos de privacía y autentificación de IS-41 en cualquier Red Celular Regional (RCN). Haciendo un marco de referencia general para calcular este trafico, los pasos de abajo describen como podemos calcular este trafico:

1. Sea τ el que denota él numero de áreas de registro (Ras) en una RCN. Observemos que una RA tiene una MSC con una VLR, y hay una sencilla HLR y un solo servidor CA en una RCN.
2. Sea σ la que denota el tamaño de cada RA en Km^2.
3. Calcular la longitud del borde (L) de la RA. Para este caso asumimos que cada RA es hexagonal, entonces $L = 3.72 \times \sqrt{\sigma}$.
4. Sea ρ la densidad de los suscriptores en términos del numero de MS´s por RA.
5. Calcular el numero total de suscriptores (incluso el numero total de estaciones móviles) utilizando $\tau \times \sigma \times \rho$.

REFERENCIAS DEL CAPÍTULO

[1] Thomas Y.C. Woo and Simon S. Lam "Authentication for Distributed Systems" IEEE, Vol. 25 No. 1 January 1992 pp. 39-52 University of Texas at Austin.

[2] D.E. Denning, Cryptography and Data Security, Addison-Wesley Publishing Co., Reading, Mass., 1982.

[3] G.J. Simmons, "Symetric and Asymetric Encription," ACM Computing Surveys, Vol. 11, No. 4, Dec. 1979, pp. 305-330.

[4] Data encryption Standard, FIPS Pub. 46, National Bureau of Standards, Washington, D.C., Jan. 1977.
[5] R.L. Rivest, A. shamir, and L. Adleman, "a Method for Obtaining Digital Signatures and Public-Key Cryptosystems," Comm. ACM, Vol. 21, No. 2 Feb. 1978, pp. 120-126.

[6] W. Diffie and M.E. Hellman, "Privacy and Authentication: An Introducction to Cryptography," Proc. IEEE, Vol. 67, No. 3, Mar. 1979, pp. 397-427.

[7] J. Linn, "Practical Authentication for Distributed Computing," Proc. IEEE Symp. Research in security and Privacy. IEEE CS Press. Los Alamitos, calif., Order No. 2060, 1990, pp. 31-40.

[8] R.M. Needham and M.D. Schroeder, "Using Encryption for Authentication in Large Networks of computers," Comm. ACM, Vol. 21, No. 12, Dec. 1978. pp. 993-999.

[9] U. Feige, A. Fiat, and Shamir, "Zero-Knowledge Proofs of Identity," Proc. ACM Symp. Theory of Computing, ACM Press, New York, 1987, pp. 210-217.

[10] M. Burrows, M. Abadi, and R.M. Needham, "A logic of Authentication," ACM Trans Computer Systems, Vol. 8, No. 1, Feb. 1990, pp. 18-36.

[11] D. Dolev and A.C. Yao, "On the Security of Public Key Protocols," IEEE Trans. Information Theory, Vol. IT-30, No. 2, Mar. 1993, pp. 198-208.

[12] M. Abadi et al., "Authentication and Delegation with Smart Cards," tech. Report 67, Systems research Center, Digital Eqduipment Corp., Palo alto calif., Oct. 1990.

[13] Tomas Y.C Woo y Simon S. Lam, " A Methodology for verifying Authentication Protocolos," Technical Report, department of Computer Sciences, the University of Texas at Austin 1994.

[14] M. Abadi and R. Needham. "Good engineering practice for security in cryptographic protocols". Manuscript, November 1993.

[15] R. Bird, I Gopal, A. Herzberg, P.A. Janson, S. Kutten, R. Molva, and M. Yung. "Systematic Design of a Family of Attack-Resistant Authentication Protocols". IEEE Journal on Selected Areas in Communications, 11 (5): 679-693, June 1993.

[16] T.Y.C. Woo and S.S. Lam. "A semantic model for authentication protocols". In proceedings of 14th IEEE Symposium on Research in Security and Privacy, pages 178-194, Oakland, California, May 24-26 19993.

[17] T.Y.C. Woo and S.S. Lam. "Verifying authentication protocols: Metodology and example. In proceedings of International Conference on Network Protocols, pages 36-45, San Francisco, California, October 19-22 1993.

[18] J. Linn. Generic Security Service Application Program Interface, September 1993. RFC 1508.

[19] W. Mao and C. Boyd " Methodical use of cryptographic transformations in authentication protocols" IEE Proc.- Comput. Digt. Tech, Vol 142, No. 4, July 1995.

[20] Otway, D. and Rees, O. "Efficient and timely mutual authentication", ACM Oper. Syst. Rev., Junuary 1987, 21, (1), pp. 8-10

[21] Miller, S.P., Neuman, C., Schiller, J.I., and Saltzer, J.H. "Kerberos authentication and autorization system", Project Athena Technical Plan Section E.2.1, 1987.

[22] Kohl J. and Neuman C.:"The Kerberos network authentication service (v5)", Internet Archive RFC 1510, September 1993.

[23] Burrows M., Abadi, M., and Needham , R. " A logic of authentication", SRC Technical Report 39, Digital Equipment Corporation. February 1989.

[24] ISO/IEC. CD 11770-2: 1993 "Key management. Part 2: Key management mechanisms using symmetric techniques".

[25] Needham, R M., and Schroeder, M.D. "Using encryption for authentication in large networks of computers", Commun. ACM, 1978, 21, (12), pp. 993-999.

[26] Denning, D.E. and Sacco G.M. " Timestamps in key distribution protocols", Commun. ACM, 1981, 24, (8), pp. 533-536.

[27] Ford, M. Identity Authentication and 'E-Commerce' 30.10.1998, [referred 9.11.2000], http://www.law.warwick.ac.uk/jilt/98-3/ford.html

[28] Branchaud, M. A Survey of Public Key Infrastructures March 1997, [referred 9.11.2000], <http://home.xcert.com/ marcnarc/PKI/thesis/>

[29] Diffie, W. and Hellman, M. New Directions in Cryptography, IEEE Transactions on Information Theory November 1976, pp. 644-654

[30] Abelson, H. et al. The Risks of Key Recovery, Key Escrow, and Trusted Third Party Encryption 20.8.1998, [referred 9.11.2000], <http://www.cdt.org/crypto/risks98/>

[31] Gerk, E. Overview of Certification Systems: X.509, CA, PGP, and SKIP 17.4.1997, [referred 9.11.2000], <http://www.mcg.org.br/cert.htm>

[32] Ellison, C. et al. RFC 2693 - SPKI Certificate Theory September 1999, [referred 9.11.2000], ftp://ftp.ietf.org/rfc/rfc2693.txt

[33] Puhakainen, P. Electronic Commerce: Market Estimates and Security Considerations Licentiate's thesis. Helsinki University of Technology. Espoo Finland. July 2000, [re-ferred 9.11.2000], <http://www.certall.fi/finnish/content/businessarea/lic4.pdf>

[34] The MeT Initiative. MeT Overview White Paper Version 1.0, 2 of October 2000, [referred 9.11.2000], <http://www.mobiletransaction.org/techinfo.html>

[35] SSH Communications Security Introduction to cryptography 2000, [referred 9.11.2000], <http://www.ssh.com/tech/crypto/intro.html>

[36] Wireless Application Forum, Ltd. Wireless Application Protocol, Wireless Transport Layer Security Specification 18 February 2000, [referred 9.11.2000], http://www.wapforum.org

[37] Wireless Application Forum, Ltd. Wireless Application Protocol, Identity Module Specification 18 February 2000, [referred 9.11.2000], <http://www.wapforum.org>

[38] RSA Security RSA SecurID, Web Portfolio. How RSA Se-curID Agents Can Secure Your Website 2000, [referred9.11.2000],<http://www.rsasecurity.com/products/securid/whitepapers/web/Web_Agent_Solution_WP.pdf>

[39] Sonera SmartTrust Ltd. SmartTrust SIM Security Client 2000, [referred 9.11.2000], <http://www.smarttrust.com/products/sim_security_client.html>

[40] L. Jianwei, W. Yumin, " Authentication of mobile users in Personal Communications System", IEEE 1996, pp 1239-1241.

[41] V. O. K. Li, X. X. Qiu, " Personal Communication Systems (PCS)," Proc. IEEE, Vol. 83, No. 9, pp. 1210-1213.

[42] B. Jabbari, "Intelligent Network concepts in mobile communications, " IEEE Commun. Mag., pp. 64-69, Feb. 1992.

[43] K. Kohiyama, et al, "Advanced personal communications system", IEEE VTC'40, pp. 161-166, 1990.

[44] P. E. Wirth, "Telegraphic Implications of Database Architectures in Mobile and Personal Communications," IEEE Commun, Mag., pp 54-59, June 1995.

[45] M. Rahnema, "Overview of the GSM System and Protocol Architecture", IEEE Commun. Mag., pp. 92-100, April 1993.

[46] D. Brown, "Techniques for privacy and authentication in personal communication systems", IEEE Personal Commun., pp. 6-10, Augost 1995.

[47] S. Chokhani, "Toward a national public key infrastructure", IEEE Commun. Mag., pp. 70-74, Sept. 1994.

[48] Refik Molva, Didier Samfat, and Gene TsudiK, "Authentication of mobile users," IEEE network, pp. 26-34, March-April 1994.

[49] Zheng Zhibin, Zhang Naitong, "Dynamic authentication protocol for PCS" International Conference on Communication Technology ICCT, October 22-24 1998, Beijing, China. Pp. 1-5

[50] H. Y. Yu and L. Harn, "Authentication protocols for personal communications Systems, ACM Comp. Commun. Review, pp. 256-261, September, 1995.

[51] Liu Jianwei and Wang Yumin, " A protocol for authentication of mobile user based on Kryptonight, Acta Electronica Sinica, pp. 257-262, No. 1, January 1998.

[52] J.S. Stach, E. K. Park, Z. Su , " An Enhanced authentication protocol for personal communication systems", IEEE, 1998. pp. 128-132.

[53] ETSI / TC, "Recommendation GSM 03.20, security related network functions", Version 3.3.2 (1991).

[54] Shiuh-Pying Shieh, Wen-Her Yang, and Hun-Min Sun, "An authentication protocol without trusted third part", IEEE, Communications Letters, Vol 1. No.3 May 1997.

[55] A. Shamir, "Identity-based cryptosystems and signature schemes," in Proc. Crypto-84, santa Barbara, CA, 1984, pp. 47-53.

[56] S. Tsujii, T. Itho, and K. Kurosawa, "ID-based cryptosystem using discrete logarithm problem", Electron. Lett., vol. 23, pp. 1318-1320, November 1987.

[57] E. Okamoto and K. Tanaka, "Identity-based information security management system for personal computer networks", IEEE J. Select. Areas Commun., vol. 7 pp. 290-294, February 1989.

[58] J. Hastad, "On using RSA with low exponent in a public key networ", in Lecture Notes in Computer Science: Advances in Cryptology CRYPTO'85 Proc., pp. 403-408.

[59] M.I. Samarakoon, B. Honary, " Novel authentication and key Agreement protocol for low processing power and systems resource requirements in portable communications sustems", IEE, Savoy Place, London WC2R OBL,1999.

[60] RFC 1321, "The MD5 Message- Digest Algorithm"

[61] B. Schneier, " Applied Cryptography", Published by John Wiley & Sons, Inc.

[62] Michael. J. Beller, Li-Fung. Chang, and Yacov Yacobi, "Privacy and authentication on a portable communications system", IEEE Globecom'91 Conf. , Phoenix, pp. 1922-1927, December 1976.

[63] M. J. Beller, and Y. Yacobi, "Fully fledged two way public key authentication and key agreement for low-cost terminals", Electronics letters. Vol. 19, No. 11, pp. 999-1001, May 1993.

[64] *Khalid Al-Tawil, Ali* Akrami, "A new authentication protocol for roaming users in GSM networks", IEEE 1999.

[65] Shesadri Mohan, "Privacy and authentication protocols for PCS", IEEE Personal Communications, pp. 34-38, October 1996.

[66] Derek Lam, Donald C. Cox and Jennifer Widom, "Teletraffic Modeling for Personal Communication Services", IEEE Communications Magazine, pp. 79-87, Feb. 1997.

[67] Thomas F. La porta, Malathi Veeraraghavan, and Richard W. Buskens, "Comparison of signaling Loads for PCS systems", IEEE/ACM transactions on Networking, vol. 4, no. 6, pp. 840-855, December 1996.

[68] Herbert Luna Galiano, Lee Luang Ling, "Análisis of Signaling Traffic for Authentication and Privacy Protocols in Cellular and PCS systems", IEEE, 1998. pp. 329-334.

[69] Telecommunications Industries Association. Appendix A of IS-54 Rev. B, February 1992.

[70] Luna, H., Rochol, J., "Security in PCS Systems" SBRC97 Brazilian Symp. on Comp. Net., may 1997.

[71] Expocomm Newsletter Publication of E. J. Krause of Brazil, Sao Paulo, January 1998.

[72] Pollini, G., Goldman D., "Signaling System Performance- Evaluation for PCS". IEEE Transactions On Vehicular Technology, pp. 131-138, February 1996.

[73]Mohan S., "Two user location strategies for PCS", IEEE Personal Communications, pp. 42-50. First quarter 1994.

CAPÍTULO 3

3. INTRODUCCIÓN

3.1 PLANTEAMIENTO DE LOS PROBLEMAS GENERALES SOBRE LOS PROTOCOLOS DE AUTENTIFICACIÓN PARA PCN-PCS

Recientemente, varios protocolos de autentificación para PCSs han sido propuestos por los organismos estandarizadores [5, 6, 7, 8] e investigadores independientes [9, 10, 11, 12, 13] entre muchos otros de los ya comentados en él capitulo anterior. Con diversas consideraciones en mente y técnicas utilizadas, cada uno tiene sus propias ventajas y desventajas en diversas aplicaciones [14, 11, 15, 13, 16, 17]. Pero todos con el mismo fin de contrarrestar distintas amenazas de seguridad. Sin embargo, la fuerza de estos protocolos se debilita generalmente en un ambiente de "roaming" donde la apertura de seguridad de una red visitada podría conducir a daños persistentes a los suscriptores que la visitan. La identidad del suscriptor no está bien protegida en varios de los protocolos, y los mecanismos apropiados que solucionan conflictos en cuentas de roaming no se utilizan tampoco. Para ofrecer una propuesta de solución alternativa a estos problemas, se proponen protocolos de autentificación en este trabajo de tesis con nuevas características de seguridad que no se han explorado completamente antes.

En sistemas de comunicaciones personales, tener acceso abierto a la radio expone el contexto de la comunicación sobre conexiones inalámbricas entre una unidad móvil y la red fija. Tal apertura también da a un intruso la oportunidad de enmascararse como suscriptor legítimo para hacer llamadas libres. Para proporcionar protección apropiada a estas conexiones inalámbricas, necesitan ser proporcionadas características de seguridad, tales como confidencialidad, integridad o disponibilidad.

En principio, algunas de estas características se pueden alcanzar con los protocolos de autentificación que verifican las identidades de entidades en ambos puntos finales de la conexión inalámbrica y establecen una llave de sesión secreta entre ellos para la comunicación secreta

siguiente. Aunque, los protocolos en redes alámbricas con características similares han estado disponibles, no sería apropiado aplicarlas directamente en el ambiente de PCNs-PCSs, debido a algunos requisitos específicos en el ambiente inalámbrico. Por ejemplo, las consideraciones en complejidad del hardware, potencia de batería, y retardo de la validación han limitado al dispositivo móvil para poder realizar los cómputos, ya que requieren de costoso hardware o son de alto consumo de energía.

El servicio de los sistemas de comunicación personales es proporcionado por múltiples redes regionales, cada una funcionando bajo diferente administración. Un suscriptor podría vagar entre varias redes. Para la mayoría de los sistemas, el suscriptor y su red local comparten una llave de autentificación con la cual pueden aprobarse el uno al otro durante el proceso de autentificación. Para una situación de movimiento, en lugar de la llave de autentificación en sí misma, algunos parámetros de seguridad derivados de la llave de autentificación se envían de la red local del "roamer" a la red visitada, y así este puede realizar el proceso de autentificación. Para reducir al mínimo el retardo causado por las interacciones con la red local durante el proceso de autentificación, uno u otro de varios conjuntos de parámetros de seguridad se generan y se transfieren en lotes a la red visitada antes de la autentificación, o los mismos valores de los parámetros de seguridad se utilizan en varias ocasiones en varios casos de autentificación para cortar cierto tráfico. Sin la transferencia del último secreto (la llave de autentificación) a la red visitada, se reduce el riesgo de exponer la llave de autentificación, que causa daño serio al servicio. En caso de un hueco de seguridad, se espera que la seguridad del servicio se recupere después de que expiren los parámetros comprometidos de dicha seguridad. Tal seguridad puede también ser recuperada desechando estos parámetros de seguridad si se detecta ta' compromiso. Sin embargo, estos parámetros de seguridad imponen carga adicional de la misma a la red visitada para su almacenamiento y manejo. Ya que hay muchas redes en los PCSs, cada una funciona bajo diversa administración con diverso nivel de protección, algunas redes son más vulnerables que otras a los ataques de intrusos o de internos. Una vez que estos parámetros de seguridad sean comprometidos, por un intruso o un interno, el fraude sucederá en cualquier momento. A veces, el daño causado es más serio y persistente que como fue pensado inicialmente. Por ejemplo, si la tripleta (Kc, R, S) en GSM [8], KS en DECT [7], o el SSD en USDC [6] son conocidos por el atacante, este podría utilizar esto para "impersonate"[17] una red legal (estación-base) para establecer una conexión con el correspondiente suscriptor, y por lo tanto poder trazar cierta información privada de la conversación, por ejemplo, la identidad de las partes implicadas en la conversación. El atacante puede también entonces fingir ser la otra parte de la llamada para recopilar la información adicional hasta que se detecta tal personificación. Este ataque se puede lanzar en varias ocasiones en GSM porque la unidad móvil no se diseña para detectar parámetros de seguridad utilizados. En USDC, aunque, un nuevo SSD puede ser reestablecido a través del protocolo de actualización del SSD. Desgraciadamente, puede ser invocado solamente por la red del segmento. Si un intruso intenta enmascarase como red legal, él no invocará al protocolo de la actualización del SSD en todos. En DECT, un protocolo independiente se proporciona al suscriptor para verificar la red del segmento. La invocación de este protocolo es opcional. Si un suscriptor sospecha que la red esta siendo personificada, él puede invocar este protocolo con el costo de un retardo extra o ancho de banda adicional. Incluso con el método de llave-pública, cierta información sensible de un suscriptor-específico se podría también encontrar en una red visitada, por ejemplo, la llave común η de RCE y MU en MSR+DH [10]. Con el conocimiento de η, el atacante puede personificar siempre al suscriptor legal dentro de esta red específica.

[17] acceder al sistema utilizando la identidad de otro

Bajo algunas situaciones, una llave de sesión vieja puede también ser derivada sin tener que romper una unidad móvil o un proveedor de servicio. Con esta llave de sesión comprometida junto con otra información registrada, un atacante puede hacer llamadas fraudulentas o enmascararse como red legítima del segmento para establecer una conexión falsa con el suscriptor.

El presente capítulo contiene los protocolos que hemos propuesto en el desarrollo de esta tesis, y que han sido publicados en conferencias y congresos internacionales[18] [1, 2, 3, 4]

Para el resto del documento, primero determinaremos las características de la seguridad deseadas y los requisitos de la implementación de los protocolos de autentificación para PCSs. Y posteriormente se describen a detalle en cada uno de los protocolos propuestos.

3.1.1 CARACTERÍSTICAS DESEADAS DE LA SEGURIDAD Y REQUISITOS DE IMPLEMENTACION.

Seguridad.

Establecimiento de la llave de sesión: Las señales de radio transmitidas sobre el aire en sistemas inalámbricos y celulares actuales se pueden interceptar fácilmente por los exploradores (scanners) comercialmente disponibles. En los sistemas digitales avanzados, este problema todavía existe. Para proteger el contenido de la comunicación contra escuchantes no autorizados, los mensajes se deben transmitir en ciphertext. Por lo tanto, durante el proceso de autentificación, un secreto común se debe convenir entre el suscriptor y la red. Esta llave de sesión se puede utilizar repetidamente en algunas situaciones según lo mostrado en [13] y opcionalmente en DECT [7]. Pero debido a las preocupaciones de la seguridad, la mayoría de los protocolos requieren una llave nueva para cada sesión.

Confidencialidad del mensaje: Un mensaje transmitido después del levantamiento del enlace, incluyendo datos y voz, debe protegerse de externos. Esto puede ser alcanzado encriptando dicho mensaje con una llave de sesión secreta común establecida por el subscriptor y la red durante la fase de autentificación.

Confidencialidad de ID del llamante: En el sistema de telefonía tradicional, un suscriptor está conectado con la oficina local a través de una línea fija. Esta línea identifica automáticamente al suscriptor (el número de teléfono.). Sin embargo, en el ambiente inalámbrico, sin tal asociación física, un suscriptor tiene que proporcionar de alguna manera su identidad a la red para la verificación necesaria. Como una identidad del suscriptor, es decir, su localización actual, puede ser de valor especial para algunas personas [18], la identidad real del suscriptor no se debe exponer al exterior. Algunas veces, la identidad del roamer se puede incluso ocultar de las redes visitadas. Desgraciadamente, tal confidencialidad no es soportada rigurosamente en algunos estándares actuales.

---

[18] IASTED (International Conference on Wireless and Optical Communications (WOC 2001) Junio 2001, Intelligent Systems and Control (ISC 2001) November, 2001). en Banff, Canadá y Tampa, Florida, U.S.A. Y el X Congreso Internacional de Computo CIC 2001, noviembre, 2001, Ciudad de México, D. F.

Autentificación mutua: En sistemas celulares anteriores, una petición de llamada hecha por un roamer se concede mientras que la autentificación todavía está en curso. Para el momento en que salga el resultado, varias llamadas fraudulentas pudieron haber sido terminadas ya. Tal retardo es debido a la carencia de comunicación apropiada entre-carriers y ha causado billones de dólares en pérdidas a los estos mismos [19]. Sin embargo con un establecimiento de acuerdo inter-carrier, el proceso de la validación puede ser terminado antes de que se conceda la primera llamada. La modificación de números de serie y escuchar sin autorización en señales de radio todavía deja la puerta abierta para que los atacantes cometan fraude. Pero, con la aparición de nuevos sistemas digitales, las técnicas criptográficas modernas se pueden ahora utilizar para eliminar tal fraude causado por enmascaramiento. Un problema similar es la personificación de una red por el intruso como ya se ha mencionado, lo cual también causa problemas serios. Por lo tanto, es importante para un suscriptor y la red que se autentifiquen mutuamente en el proceso de autentificación.

No-negación del servicio: Para el proveedor de servicio, es deseable que un suscriptor no pueda negar la cuenta incurrida de los servicios que él solicitó. Semejantemente, el suscriptor no debe ser incorrectamente cargado debido a ningún error de factura o apertura de la seguridad en la red. Teóricamente, ambas metas se pueden alcanzar con el uso de las firmas digitales [20]. Aunque generalmente no se recomienda esto en ningún estándar para cantidades grandes de cálculos implicados.

Requisitos de la implementación

Una consideración importante para proporcionar autentificación apropiada sobre conexiones inalámbricas, son los gastos indirectos de cómputo en las unidades móviles. Debido a las consideraciones sobre complejidad del hardware, la potencia de batería, y el retardo del cómputo, en algunos dispositivos móviles, por ejemplo, teléfonos celulares, no se pueden realizar las operaciones complicadas que requieren hardware costoso o bastante tiempo. Tales limitaciones generalmente excluyen del uso de técnicas criptográficas de llave-pública (por el alto consumo de tiempo) las cuales pueden proporcionar al servicio deseado de la no-negación. Probablemente, esta es la razón por la cual la confidencialidad fuerte del suscriber-ID no sé a utilizado en la mayoría de los estándares actuales. Los Datos Celulares de Paquete Digital (CDPD) [5], MSR+DH [10], y el propuesto por Aziz y Diffie [9] son ejemplos que utilizan técnicas de llave-pública entre otros. Estos protocolos requieren mucho más cálculos en comparación con los que utilicen técnicas criptográficas de llave secreta. Otra preocupación con propuestas de llave-pública es el revocado de los certificados, que requieren mecanismos complicados para su manejo. Los algoritmos criptográficos convencionales de una-llave proporcionan operaciones rápidas en la encripción y desencripción, y por lo tanto se utilizan para la comunicación secreta, si una llave secreta común es convenida entre las partes que se comunican. Incluso en el proceso de autentificación de los actuales estándares que establece una llave de sesión común para una conexión inalámbrica, todavía se prefieren.

Otra consideración a tomar en cuenta es el retardo de la validación. Durante el proceso de validación, además del cómputo realizado por las partes participantes, se intercambian mensajes entre un suscriptor y la red o entre una red visitada y la red local del suscriptor. Estas interacciones, causan retardo en la validación. La reducción de tales interacciones es importante tomar en cuenta al diseñar protocolos de autentificación.

3.2 PROTOCOLOS.

3.2.1. PROTOCOLO DE AUTENTIFICACIÓN DE USUARIOS MÓVILES PARA REDES DE COMUNICACIÓN PERSONAL

Como se ha venido comentando, las redes de comunicaciones móviles requieren de funciones de seguridad adicionales, en contraste con aquellas funciones existentes en las redes fijas. En efecto. un nuevo problema que involucra la movilidad, se refiere a la situación donde los usuarios tienen la capacidad de acceder a los recursos de red desde múltiples puntos, los cuales pueden estar separados por distancias geográficas significativas, involucrando esto la necesidad de transferir información sobre diferentes redes que pueden presentar diferencias en las administraciones del servicio. Por tal motivo, como estos puntos de acceso no están necesariamente bajo el control de una única autoridad administrativa, es necesario disponer de un conjunto de mecanismos "inter-dominios" que permitan a los usuarios realizar operaciones seguras, entre y en los dominios. ofreciendo, mediante estos mecanismos una garantía en las diferentes transacciones que realizan. Para lograr todo esto, es necesario contar con los protocolos adecuados. En el primer articulo sugerimos un protocolo general para la autentificación de usuarios, proponiendo un nuevo esquema de autentificación de usuario mediante llave pública, basado en la modificación del sistema de firma digital, que emplea certificados y curvas elípticas, como una alternativa conveniente, para generar un protocolo que permita su aplicación en redes de comunicaciones móviles. Superando de esta manera, deficiencias del protocolo de autentificación de llave secreta: teniendo una complejidad computacional baja y ofreciendo una alta seguridad.

3.2.1.1 Revisión de los métodos y protocolos propuestos

La movilidad de usuarios es una característica que puede ser ofrecida en diferentes ambientes de red. tales como; WLAN, celular móvil, infrarrojo entre otros. En el caso de las redes celulares móviles, la arquitectura GSM [8, 8′,8′′] es la primera en suministrar servicios de seguridad tales como: autentificación de usuarios, confidencialidad del trafico, y distribución de llaves. El principal problema del método de autentificación que emplean GSM y CDPD [5]) se refiere a confiar en el supuesto de que la "red fija" sea segura. En la figura 3.1 es ilustrada la diferencia entre enlace seguro y seguridad "punto-a-punto", y en la figura 3.2 mostramos la capa de enlace que por si mismo es solamente un salto típico mas en las redes inalámbricas con respecto a las redes alámbricas.

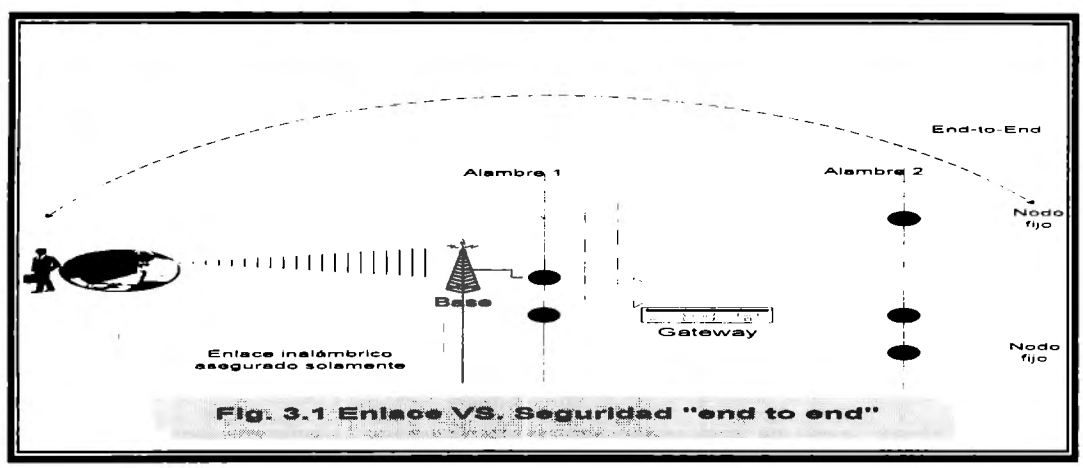


Fig. 3.1 Enlace VS. Seguridad "end to end"

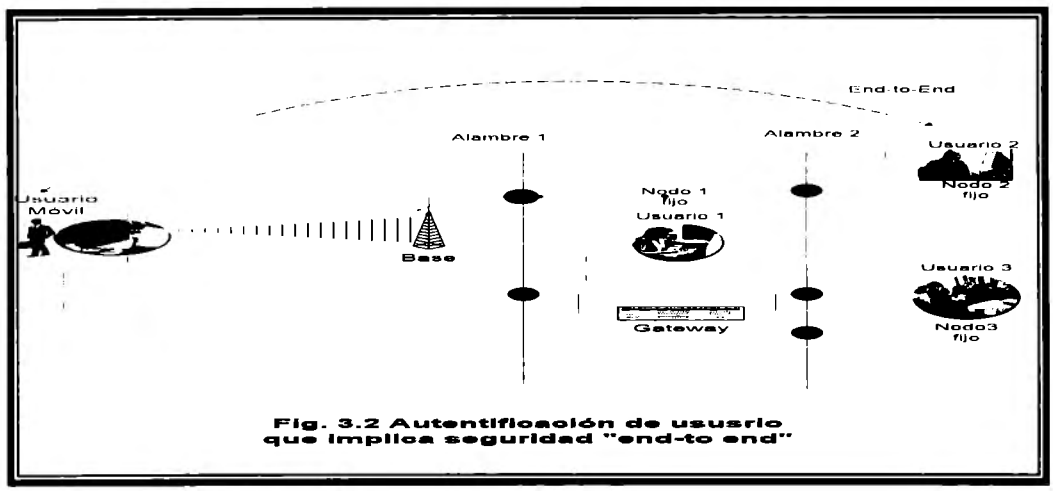


Fig. 3.2 Autentificación de usuario que implica seguridad "end-to end"

Los mensajes son transmitidos de forma "clear text" entre los Centros de Conmutación Móvil (CCM[19]). Donde todos los CCM confían en la seguridad de la comunicación entre ellos. Sin embargo, la misma suposición no puede ser hecha para ambientes de red heterogéneos, gestionados por diferentes autoridades administrativas. De tal manera que, una arquitectura de seguridad con un mínimo de suposiciones, es necesaria en términos de la información que es intermediada.

Por otro lado, es importante considerar la forma de distribuir la información de autentificación de usuario en una red de PCS. En algunos casos, el dominio de origen o local, debe generar un conjunto de "pares de autentificación" mientras que, el dominio externo o visitado utilizará estos en los flujos de autentificación sucesivos con el usuario final. Esta solución, es ineficiente en términos del uso o consumo de anchos de banda y el proceso de generación de los protocolos desde el origen. Además, es probable que solamente un pequeño numero de los pares mencionados sea comunicado, lo cual provocaría que el dominio visitado solicitara retransmisiones al dominio origen para poder contar con un "batch" fresco.

Otros ambientes pueden ser adaptados para soportar movilidad de usuarios. Esto es, una red alámbrica puede ser equipada de tal forma que pueda permitir accesos universales para ofrecer un servicio de valor agregado; tal como, las TPU[20] o UMTS[21][21,22]. A diferencia de GSM, TPU no esta en un estado maduro donde soluciones especificas de seguridad hayan sido propuestas.

El protocolo de autentificación de usuarios de llave secreta

Actualmente, muchos estándares de sistemas móviles, emplean este tipo de protocolos. [23] En estas redes los usuarios son autentificados mediante el envío de un numero aleatorio, el cual es generado por una estación base; El usuario calcula una respuesta empleando este número, y la

[19] CCM = MSC (Mobile Switching Centers)
[20] Telecomunicaciones Personales Universales ó UPT (Universal Personal Telecommunication)
[21] Universal Mobile Telecommunication System (UMTS)

envía de regreso a la red. El procedimiento en el cual este método es llamado "challenge-response" se muestra en la figura 3.3.

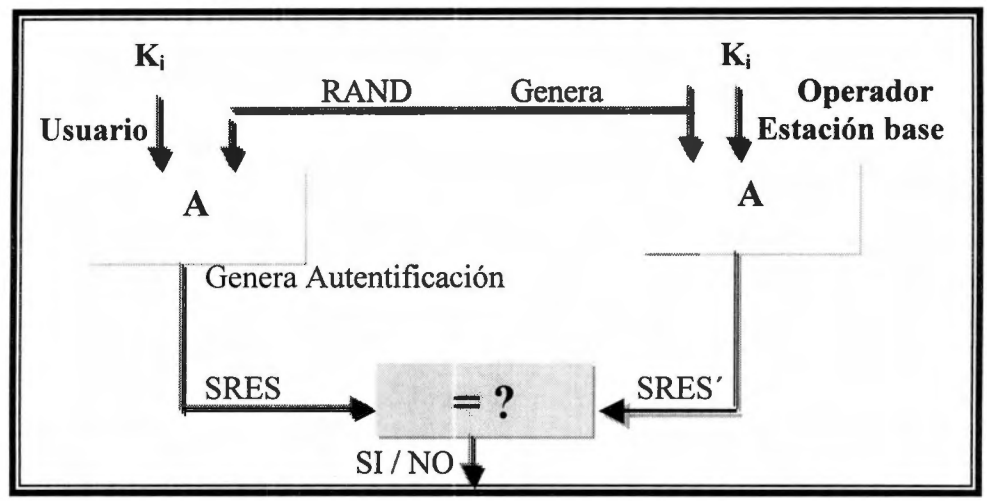


Fig.3.3 El protocolo de autentificación de usuarios de llave secreta

Donde, se puede observar que en ambos procesos el sistema hace uso de una clave secreta (Ki), la cual sirve como entrada al algoritmo correspondiente (A). Esta Ki solamente es conocida por el usuario y el operador, estas claves son diferentes para el algoritmo de autentificación y generación de claves, por el uso de un número secreto aleatorio (RAND) el cual es generado por el operador. En otros términos, las técnicas de llave secreta o simétrica tienen fundamentos de complejidad diversa, pero todas usan una misma llave K_i que es conocida por el remitente de los mensajes y por el receptor, y mediante la cual sé encripta y desencripta el mensaje que se quiere proteger; pero el principal inconveniente estriba en la necesidad de que todas las partes conozcan K_i. Esta llave es distribuida mediante una transacción separada y diferente a la transmisión del mensaje encriptado. Es aquí, precisamente, donde se halla el punto vulnerable del mecanismo: la distribución de la llave. Si la llave es interceptada se pone en peligro todo el sistema de seguridad. La llave debe ser transmitida por un canal seguro para poder asegurar la eficacia del sistema criptográfico. Más aún, si el usuario (estación móvil) quiere autentificarse así mismo hacia la red, la red debería conocer la llave secreta de autentificación del usuario, K_i por adelantado. Y una vez que K_i es descubierta el protocolo podría ser roto. Eso es muy inseguro. Pero es todavía ampliamente utilizado hoy en día, por su alta velocidad y fácil implementación

El protocolo de autentificación de llave pública

En 1993, M.J. Beller et al propusieron 3 protocolos de autentificación de llave pública para redes de comunicaciones portátiles [24]. Estos protocolos fueron presentados en el articulo de Liu Jianwei, et al [25]. Y lo reproducimos en esta parte por tratarse de algo de interés; Sus características de seguridad y complejidad computacional son mostradas en la tabla 3.1.

Tabla 3.1

Protocolo público	Autentificación de red	Privacidad del usuario descubierta hacia la red	Multiplicidad modular
MSR	NO	SI	1
IMSR	SI	SI	2
MSR+DH	SI	NO	212

De la tabla 3.1, podemos ver que aun cuando los protocolos MSR y el IMSR tienen una complejidad computacional baja, la privacidad del usuario puede ser descubierta del lado de la red. Si el individuo que descubrió esto es una tercera parte no autorizada (intruso), puede obtener información de la ubicación del usuario mediante infiltración de la identidad de los mismos hacia la red. Por consecuencia, su seguridad es reducida. El protocolo MSR+DH tiene alta seguridad mientras que la privacidad de usuarios no es descubierta hacia la red, pero este necesita demasiados cálculos, lo cual hace al protocolo muy impractico.

3.2.1.2 Suposiciones iniciales

En este primer protocolo asumimos que, cuando se acceda a la red desde el dominio de origen, el usuario móvil es autentificado mediante un mecanismo basado en servidor autentificador tradicional. Los usuarios de cualquier dominio de red son registrados con ese servidor autentificador de dominios (AS). El AS de un dominio puede ser duplicado o subdividido dentro del dominio pero el grupo de todas las subdivisiones y duplicados de AS representan un solo nivel de dominio de autoridad. Una característica importante de los ambientes móviles es la velocidad a la cual los usuarios se mueven entre dominios en la red.

Más aún, tomando en cuenta que los sistemas de comunicación móvil, comparados con las redes fijas, tienen diferentes propiedades. Tales como:

- Estaciones móviles se mueven desde células hacia otras células, Esto requiere que la velocidad de autentificación debería satisfacer los requerimientos de una comunicación en tiempo real.
- El acceso ilegitimo o no permitido a la red es un asunto de alta inquietud. Por que esto puede afectar el correcto desempeño de la red. En algunos sistemas, la autentificación de la red a los usuarios no es considerada.
- Los recursos computacionales son asimétricos. Del lado del usuario, un microprocesador de 8 bits es generalmente utilizado. Pero del lado de la red, una gran escala de cómputo puede ser adoptada.

Utilizando estas propiedades, proponemos un nuevo protocolo de autentificación, mismo que emplea los siguientes criterios de diseño.

3.2.1.3 Criterios de diseño

El protocolo que se presenta en primer término en esta tesis, debe tener éxito aún en presencia de agentes maliciosos (intrusos.) Se asume que este intruso puede estar dentro del sistema y por lo tanto puede interceptar e introducir nuevos mensajes en el sistema utilizando información de mensajes que ha visto previamente. Es por esto, que la solución debe tomar en cuenta los siguientes criterios de diseño.

- Separación de dominios. Dominio especifico, información secreta o susceptible tal como la llave secreta de usuario o password no deberán ser propagadas desde el dominio de origen hacia un dominio externo o entre dominios visitados.

- Transparencia para usuarios. La autentificación en dominios visitados debe tener un impacto mínimo en la interfase de usuario con respecto a la autentificación en el dominio local.

- Identidad de usuario confidencial. Esto es a menudo deseable para mantener tanto, los movimientos, como los actuales paraderos de los usuarios móviles, secretos. Por esta razón, toda la información de identificación de usuarios debe ser protegida desde la divulgación.

- Encabezados reducidos. La distancia entre el dominio local y el visitado puede ser muy grande, por lo tanto, él numero de mensajes intercambiados entre él domino de origen y el remoto con fines de autentificación debe ser mantenido reducido.

3.2.1.4 Fundamentos del protocolo

De acuerdo con los criterios de diseño, así como las propiedades de los sistemas de comunicación móvil se crea el siguiente fundamento del protocolo, el cual es descrito a continuación:

Las siguientes notaciones son empleadas para este protocolo:

$Cert_A$	Certificado de Autentificación
α	Base del problema de logaritmos discretos D-H (Diffie-Hellman)
m_S	Modulo de D-H
A_s	Llave de sesión dependiente de la localidad solo utilizable en el dominio R)
A_S	Llave de sesión compartida entre AS_R y KCC
K_s	Llave de sesión de corto uso, a ser compartida entre AS_R y usuario
m_A	Modulo A_S
p_A	Numero primo para generar las llaves A_S
q_A	Numero primo para generar la llave A_S
n_p	La llave pública del KCC (Centro de Conmutación de Llaves)
m_n	Modulo de la PK (Llave pública) del KCC
iddR	Identification del dominio remoto
idA, idB	Identification de usuarios A y B
$m\alpha$,	Primo grande para firma digital y /o parámetro de curva elíptica
NR	Nonce generado por el visitante del dominio
AS_R	Servidor de Autentificación del dominio visitado
Zp	Campo multiplicativo Zp (residuos mod p)

Procedimiento para KEA[22] (Key Exchange Algorithm)

En el esquema de Diffie-Hellman, Alice combina la llave pública de Bob con su propia llave privada para crear una llave de sesión. Después "Bob combina su llave privada con la llave pública de Alice para crear la misma llave de sesión.

En KEA se hace esto de manera diferente. Ambos, Alice y Bob, tienen unas llaves privada y pública de larga duración, pero también generan llaves privadas y públicas de un solo uso para una sesión específica. Alice combina su llave privada de larga duración con la llave pública de sesión de Bob, y su llave privada de sesión con la llave pública de larga duración de Bob.

Como ejemplo de uso de esquemas basados en el Problema de Logaritmo Discreto (PLD), en el protocolo Diffie-Hellman básico[23], dos partes A y B quieren compartir un secreto. A genera un primo p y generador $\alpha \in Z^*_p$; $(2 \leq \alpha \leq p-2)$, genera un número aleatorio (secreto) x; $1 \leq x \leq p-2$, calcula $\alpha^x \bmod p$, y envía a B el mensaje $(p, \alpha, \alpha^x \bmod p)$ Una vez que B recibió el mensaje, elige un número aleatorio y; $1 \leq y \leq p-2$, y envía el mensaje $(\alpha^y \bmod p)$ hacia A. B luego calcula el secreto compartido haciendo $K = (\alpha^x)^y \bmod p$. Por otra parte, A recibe el mensaje de B, y calcula el secreto compartido de la siguiente forma, $K = (\alpha^y)^x \bmod p$.
Aunque el problema real a que se enfrenta un adversario es encontrar $\alpha^{xy} \bmod p$, dado que conoce el primo p, un generador α de Z*p, y elementos $\alpha^x \bmod p$ y $\alpha^y \bmod p$, se cree que la resolución de este problema tiene la misma complejidad que la del PLD.

Entonces, en el caso del protocolo Diffie-Hellman, la intención de un oponente es encontrar los valores aleatorios privados, es decir x e y, o bien xy. Sí un atacante logra esto, el cálculo del secreto compartido es eficiente.

Además, se puede generalizar la definición del GPLD para uso con cualquier grupo G y elementos $\alpha, \beta \in G$, donde el problema es encontrar un entero x tal que $\alpha^x = \beta$, si es que tal entero existe. Aquí no se requiere que G sea cíclico, ni tampoco se requiere que α sea un generador de G. Se cree que este problema es mucho más difícil de resolver.

Entonces, otros grupos son de interés para criptografía, como el grupo de puntos de una curva elíptica definida sobre un campo finito. Luego, podemos instanciar el problema con el grupo de las curvas elípticas sobre un campo finito.

Definamos ahora, el problema de los logaritmos discretos de curva elíptica (ECPLD). Sea E una curva elíptica definida sobre un campo finito F_q, y sea $G \in E(F_q)$ un punto sobre E de orden n (número primo y grande). El ECPLD es, dados E, G, y un múltiplo escalar Q de G, determinar un entero l tal que $Q = lG$.

Cabe hacer notar, la similitud de las definiciones para la ECPLD y PLD, hace que todos los criptosistemas basados en el PLD puedan ser adaptados utilizando curvas elípticas. Así, se tienen variantes de los protocolos y esquemas populares convertidos a curvas elípticas, y es entonces podemos tener ECDSA, EC Diffie-Hellman, encriptación EC ElGamal, etc. Siendo necesarias

---

[22] KEA (Algoritmo de Intercambio de Llaves)

[23] El protocolo básico no provee ni autenticación de las partes involucradas en la comunicación, ni autenticación de los secretos generados.

algunas modificaciones técnicas para adaptarlos al grupo de las curvas elípticas, pero los principios subyacentes son los mismos que para los otros sistemas basados en el PLD.

Ahora, describiremos el procedimiento general que incluye precálculos, formación de usuario y seguridad (nivel) de usuario. Es importante distinguir que los procedimientos de firma digital serán implementados en ambientes de comunicación usuario-a-usuario. A continuación, presentamos las funciones para generación de llave, generación de firma y verificación de firma empleando DSA.

Generar Llave

1. Seleccionar q primo: $2^{159} < q < 2^{160}$
2. Seleccionar un número primo p con la propiedad: q/p-1

   Observación: DSS indica que p es primo: $2^{159+64t} < p < 2^{160+64t}$: $0 \leq t \leq 8$, si t =8 entonces p es primo de 1024 bits.

3. Seleccionar $h \in Z^*_p$ y calcular $g = h^{(p-1)/q} \bmod p$, y repetir hasta que g = 1.
(g es el generador del único grupo cíclico con orden q en Z*p)

4. Seleccionar en forma aleatoria un entero x en el intervalo [1, q-1]

5. Calcular $y = g^x \bmod p$

6. Siendo (p, q, g, y) la llave pública, x la llave secreta.

Generar Firma.

Para firmar un mensaje, cada entidad tiene que hacer:
Sea m el mensaje

1. Seleccionar k ε Z: k ε[1, q-1], k aleatorio
2. Calcular $r = (g^k \bmod p) \bmod q$
3. Calcular $k^{-1} \bmod q$
4. Calcular $s = k^{-1} \{h(m)+xr\} \bmod q$, h es la función SHA-1[24]
5. Si s = 0 ir al paso 1
6. Se firma el mensaje es el par (r, s)

Verificar Firma.

Para la verificación de firmas (r, s), el Rx debe:

1. Obtener una copia autentica de la llave pública de Tx (p,q, g, y)
2. Calcular $w = s^{-1} \bmod q$ y h(m)
3. Calcular $u_1 = h(m)\, w \bmod q$, $u_2 = rw \bmod q$
4. Calcular $v = (g^{u_1}\, y^{u_2} \bmod p) \bmod q$
5. Aceptar la firma cuando y solamente cuando v = r.

[24] SHA (Secure Hash Algorithm)

Para el ambiente de comunicación usuario-a- estación base, debemos prestar atención a las siguientes construcciones.

1. A y B eligen sendas llaves aleatorias R_A y R_B respectivamente y forman un D-H α^{Rx} mod $m\alpha$, donde α y $m\alpha$ pueden ser públicos.
Los módulos $m\alpha$ pueden ser menores que todos los demás módulos. Él módulo de sesión y el módulo de la llave pública de certificados. $m\alpha$ se puede formar por elementos del KCC.

2. Se forma un campo infoA = (id_A, $A_S\alpha$, *(idd_R), id_B, $B_S\alpha$, *(idd_R), TE,$Cert_A$A)
*(idd_R) significa que podría ser una función encriptada.

2 A).- Formación del mensaje

$S = \{ \{Cert_A A\}, \{Ex, DSA_S \{S, info_x, D\text{-}H\ m_n\}\}\}$

3. Seguridad del protocolo

La idea del protocolo es que el esquema D-H garantiza la seguridad (confidencialidad) El DSA (Digital Signature Algorithm) garantiza la autentificación.

En términos de enmascaramiento, un enemigo puede agregarse al mensaje generado por la entidad origen, pero debido a que la información es firmada con la llave secreta de sesión, así como el certificado de origen de la persona, resulta prácticamente imposible enmascarar todos estos pasos.

Complejidad del protocolo.

En general, los algoritmos basados en criptografía de llave pública son computacionalmente hablando alrededor de 1000 veces más costosos que los algoritmos de llave simétrica. Sin embargo, con el desarrollo de la tecnología, y con el surgimiento de circuitos integrados de propósito especial, la criptografía de llave pública está viendo una oportunidad de implementación. Más aún, al emplear CCE, en el protocolo propuesto, se cumple con cierta eficiencia; baja sobrecarga en cálculos, tamaño de clave y ancho de banda.. (Más detalles ver el anexo B.)

Cabe hacer notar que él cálculo de la complejidad, no se llevó a cabo de una manera formal, sin embargo, de manera general y sobre la base de los algoritmos similares empleados en trabajos previos se confía en la complejidad para estos algoritmos y se considera que la complejidad computacional del protocolo en cuestión es del orden:

O ($\log^n (R_A R_B)$), quedando para trabajos futuros comprobar esta hipótesis.

Interpretación y ejecución del protocolo

Las bases o fundamentos del protocolo son descritos en la figura 3.4. Después de inicializar a la red y los usuarios móviles es habilitado el usuario móvil para establecer una residencia temporal en el dominio visitado por la solicitud de las transferencias de las referencias de autentificación dependiente de la localidad, desde el servidor de autentificación de su dominio local para aparecer en el dominio visitado.

Ahora nos turnaremos a los detalles del protocolo.

1. Precálculos

En esta parte, necesitamos inicialización para la red y para los usuarios; en el primer caso la red no es autentificada para estos usuarios, el centro de autentificación solamente distribuye los siguientes parámetros; $m\alpha$, n_p, idd_R, y α hacia la estación base. En el segundo caso, la KCC selecciona una identidad para A y B, y las transmite a ellos, con $m\alpha$, α, id_A, y id_B. Los usuarios arbitrariamente seleccionan un correspondiente numero aleatorio R_A y R_B como sus llaves secretas y calculan la correspondiente llave pública ($P_A = \alpha^{R_A} \bmod m\alpha$, $P_B = \alpha^{R_B} \bmod m\alpha$). Después el usuario transmite P_i hacia el centro de autentificación. El centro genera el certificado y es transmitido hacia el usuario. El usuario recibe este, y lo guarda[25] R_A, P_A, $m\alpha$, α, id_A , $Cert_A$ R_B, P_B, id_B, $Cert_B$, momentáneamente.

2. El usuario comienza por formar un campo de información A, el cual incluye; id_A, $A_S\alpha$, idd_R, id_B, $B_S\alpha$, Tiempo de expiración (TE), $Cert_A A$. Posteriormente forma el mensaje como se muestra a continuación:

$S = \{ \{Cert_A A\}, \{Ex, DSA_S \{S, info_x, D\text{-}H\ m_n\}\}\}$, donde, el usuario emplea su certificado, calcula la autentificación $DSA_S \{S, info_x, D\text{-}H\ m_n\}$ con una llave dependiente de la localidad calculada para el dominio visitado (una llave que el usuario puede emplear solamente en el dominio visitado). Como es conocido, existen muchos caminos para construir este tipo de llaves. Por ejemplo, $K_{sesión}$ podría ser calculada por $E_{AS}[H(AS_r, U)]$ donde H es una función hash. En nuestro caso, emplearemos asumimos que ya contamos con esta llave. El empleo de encripción en la formación de estas llaves, es esencial para la satisfacción de la restricción de la separación de dominios previniendo la derivación de la clave compartida entre el usuario y el servidor de autentificación del dominio local desde la llave dependiente de la localidad. Posteriormente, se forma el mensaje y es enviado a la estación base.

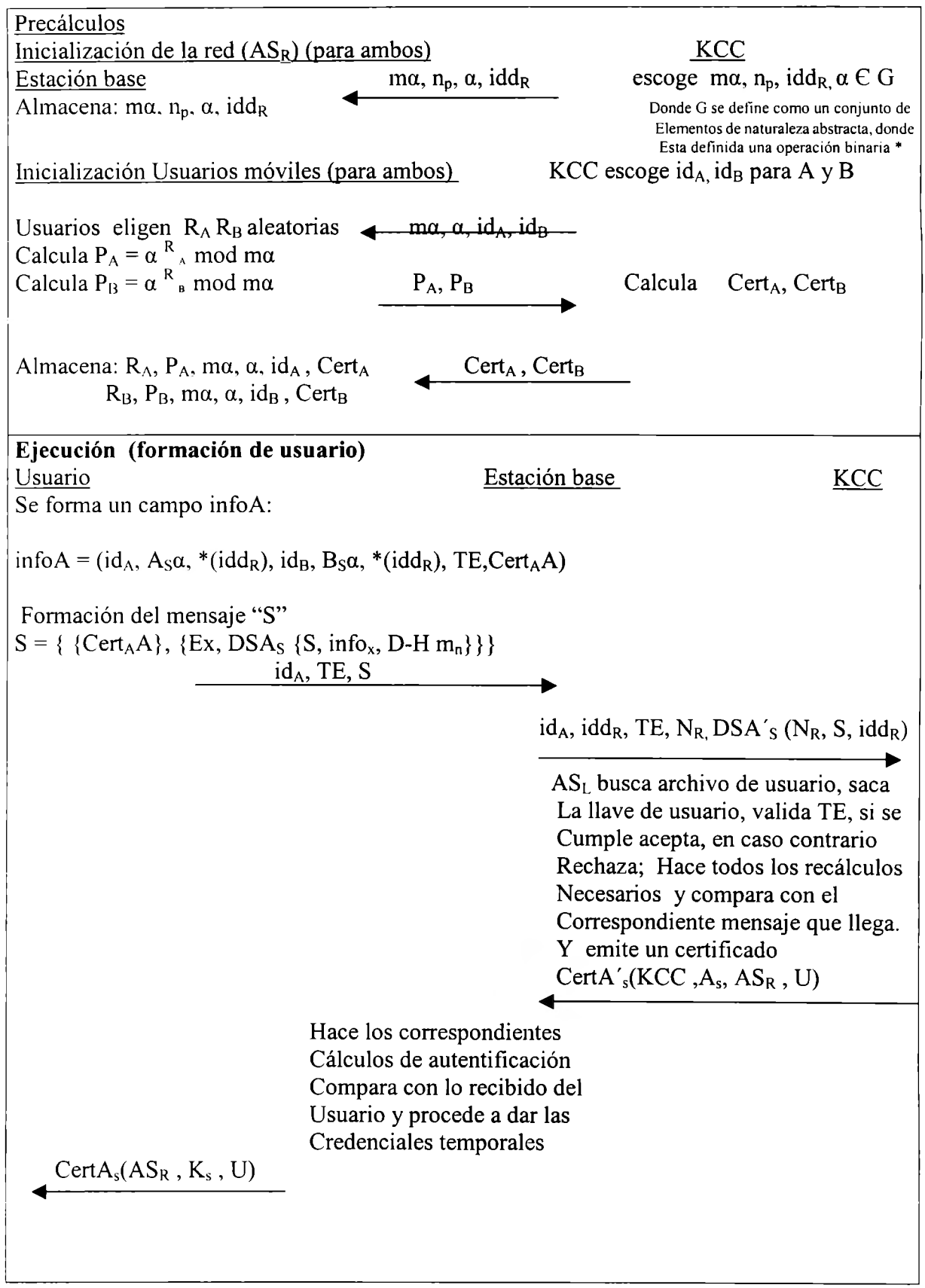
3. Cuando el dominio visitado recibe el mensaje inicial y nota que no tiene ningún medio de autentificar a este usuario, porque él reconoce que el usuario no es nativo. Entonces el dominio visitado necesita solicitar una prueba de la identidad del usuario al dominio local. Esta petición debe también autentificar el dominio del visitante al KCC, y el usuario al KCC.

4. Cuando el KCC recibe el mensaje, este procede a buscar el archivo del usuario y obtiene la llave compartida entre ellos (usuario y servidor de autentificación del dominio local), valida TE; Si este es verdadero acepta, de otra manera rechaza. Empleando los parámetros dados, es necesario hacer algunos cálculos nuevamente para compararlos con el correspondiente "token" que llego en el flujo pasado. Una comparación satisfactoria en el último paso autentifica al usuario y al dominio visitado al KCC. Entonces en este punto, el KCC pública un certificado que confirma la identidad del usuario y permite al usuario operar en el reino del dominio visitado.

5. Cuando la estación base recibe el mensaje desde el KCC, procede de la siguiente manera: Hace los recálculos de autentificación correspondientes, compara estos con los correspondientes "token" recibidos desde el usuario (primer flujo.) Una comparación en el paso anterior es decisiva, completa el ciclo del protocolo por la autentificación al dominio visitado a ambos KCC y usuario simultáneamente. Entonces el dominio visitado procede a instalar a los usuarios

[25] Se asume que esta información es almacenada en algún lugar seguro.

temporalmente. El flujo pasado puede ser utilizado para establecer una llave de sesión de trabajo entre el dominio visitado y el usuario.



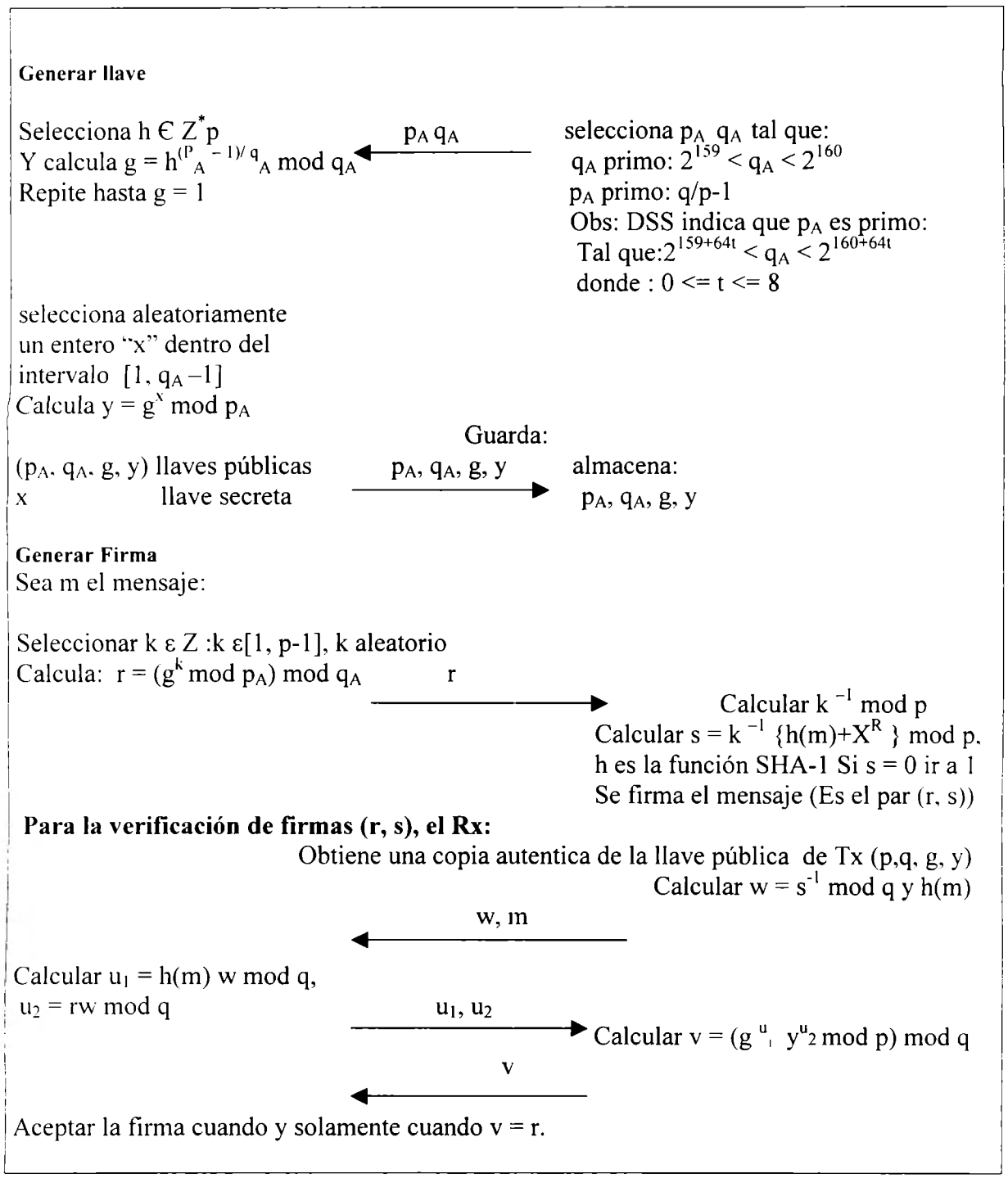


Figura 3.4 Esquema del flujo del protocolo

3.2.1.5 Comentarios

En esta primer fase presentamos un protocolo de autentificación para usuarios móviles en un ambiente "inter-dominios" con algunas ventajas para autentificación de usuarios empleando una modificación del esquema de autentificación del estándar de llave pública basado en un sistema de firma digital el cual empleaba certificados para los procedimientos entre los usuarios y estaciones base; la complejidad del algoritmo de firma digital confía en la complejidad para tal

clase de criptosistemas empleando curvas elípticas. En siguiente articulo presentamos una variación del protocolo propuesto considerando una situación donde los usuarios definen nuevos procedimientos de la seguridad a nombre de las autoridades basadas en certificados distribuidos.

3.2.2 PROTOCOLO DE AUTENTIFICACIÓN DE USUARIOS MÓVILES EMPLEANDO CRIPTOGRAFÍA DE LLAVE PÚBLICA DISTRIBUIDA.

En los últimos años, como ya se ha venido comentando, los Sistemas de Comunicaciones Personales han dado como resultado la aparición de sistemas que permiten que la comunicación se lleve a cabo desde cualquier lugar, y en cualquier momento, debido al constante crecimiento en las necesidades de comunicación de la sociedad. Estos sistemas, dada su naturaleza (movilidad) presentan nuevos problemas que afectan la confiabilidad y seguridad en la comunicación.

En este segundo protocolo se presenta una variación del primer protocolo propuesto, considerando la situación donde los usuarios definen nuevos procedimientos de seguridad a favor de una distribución de certificados de autentificación basada en autoridades.

En este caso se describe un protocolo para una completa autentificación distribuida de usuario móvil empleando criptografía de llave pública. Este protocolo de autentificación de usuario móvil esta pensado para usuarios con alto grado de movilidad; donde la autentificación, aún cuando, empleare certificados no sería del todo satisfactoria debido a la movilidad experimentada por el usuario. Se pueden evaluar dos direcciones:

1. Distribuir el certificador (Centro de Conmutación de Llaves Distribuido) DKCC (crear esquema de certificación distribuida)
2. Proponer y estudiar un sistema de tipo Distribución de Autentificación de Llave Pública (DPKA) En este se podrían evitar los certificados.

El primer caso, ofrece un interés particular en vista que existen esquemas ya probados en trabajos previos.
Se mencionan ambas direcciones, por que es de nuestro interés definir la ruta a seguir para posibles trabajos futuros.

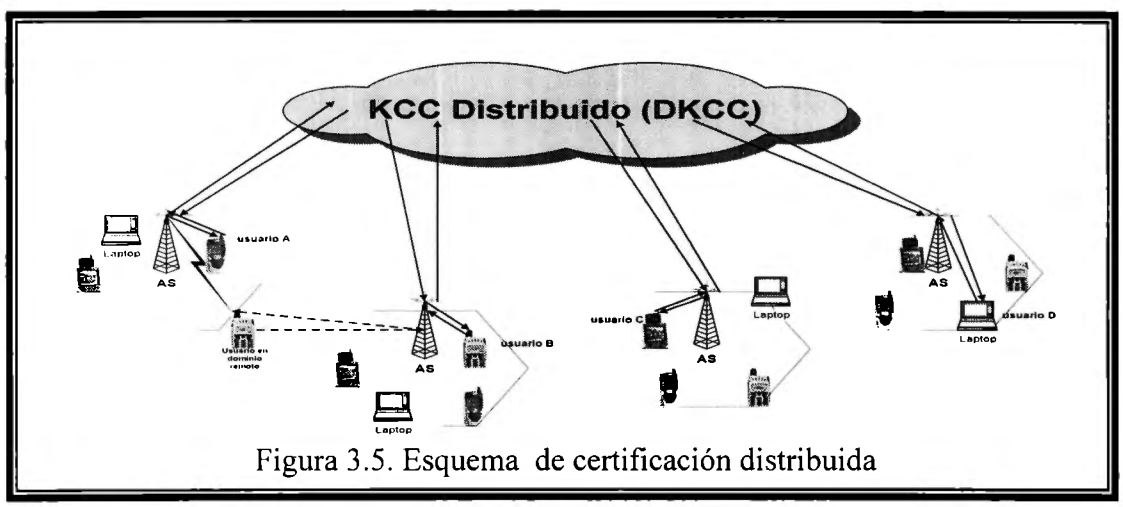


Figura 3.5. Esquema de certificación distribuida

La descripción del primer protocolo, fue dada en términos de una amplia red con KCC singular. Claramente para una amplia red, un singular KCC no puede esperar servir a todos los nodos de la red. En estos casos una jerarquía de KCCs es empleada, estas son descritas en detalle en el ITU-T antes CCITT x.509[42] y el PEM RFCs. Cuando una jerarquía de KCC es empleada, el protocolo es modificado de la siguiente manera. Los mensajes incluirán no solo el certificado en cuestión, sino también enviará un certificado de paso, el cual permitirá al móvil verificar el certificado de la base y del KCC distribuido. En la figura 3.6, se puede notar como esta determinada la estructura interna del DKCC basándonos en la idea de una estructura jerárquica. En la capa mas alta se encuentra el GKCC (KCC-global), este distribuye cadenas de datos a los RKCC (KCC-Regional) del segundo nivel, los RKKC se encargan de distribuir cadenas de datos a los AKCC (KCC- de Área) del tercer nivel y estos se encargan de distribuir cadenas de datos a los LCKCC (KCC-Local) del cuarto nivel. Estos últimos son los encargados de distribuir cadenas de datos a cada uno de los controladores de acceso y a los usuarios.

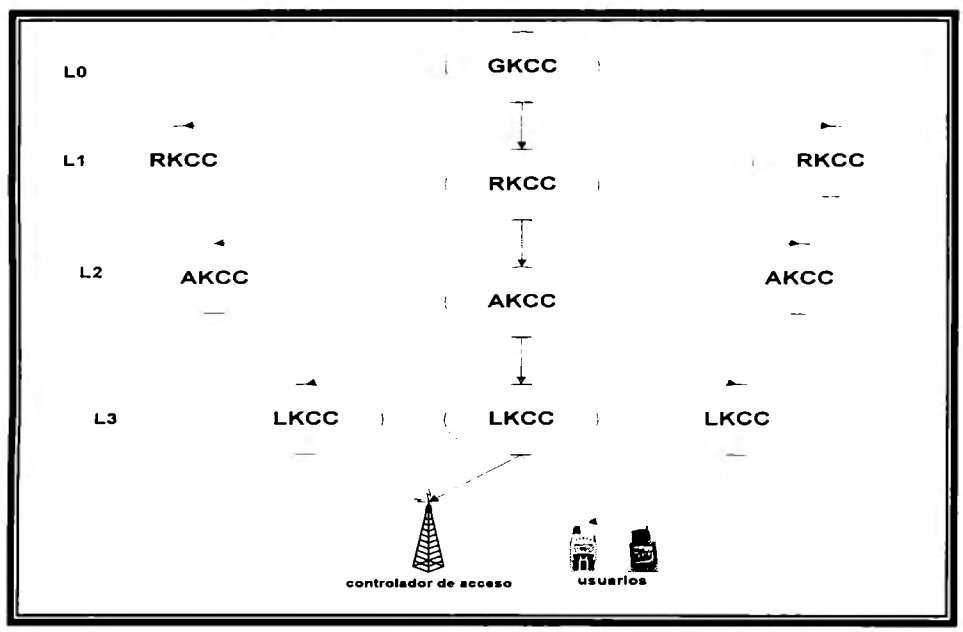


Fig. 3.6 Estructura interna del DKCC de 4 capas

3.2.2.1 Introducción

Los sistemas de comunicación móviles, dada su naturaleza (movilidad) presentan nuevos problemas que afectan la seguridad en la comunicación. La movilidad de los dispositivos y los usuarios permite el acceso a las redes de comunicación en diferentes puntos. Estos puntos de acceso no requieren, necesariamente, de un control de acceso mediante una autoridad centralizada; en este punto surge el problema de autentificación de las partes en ambientes inter-dominios. Las técnicas criptológicas, las políticas y mecanismos de seguridad ofrecen una solución a estos problemas de "comunicación confidencial y autentica"

La escalabilidad de las infraestructuras de redes seguras esta comenzando un crecimiento relacionado con el explosivo desarrollo de los continuos crecimientos de redes internacionales. El número de nuevos usuarios y aplicaciones de autentificación requeridos continuaran

incrementándose con una gran rapidez en un futuro cercano. La proliferación del comercio electrónico basado en web [26, 27], por ejemplo; agregará a la red decenas de millones de transacciones diariamente entre grandes cantidades de comerciantes y consumidores distribuidos, considerablemente separados geográficamente; se requieren esquemas de autentificación efectivos, los cuales puedan ser escalados para manipular millones de usuarios dentro de un dominio de confianza; tal como todos los usuarios de un banco grande.

Tal como Newman et al [28], hizo notar esquemas tradicionales tales como Kerberos presentan un esquema atractivo en cuestión de seguridad en la forma de Centros de Distribución de Llaves (KDC), los cuales comparten llaves simétricas con cada usuario en el dominio. En el caso de revelar el contenido de un KDC, todas las llaves simétricas serán divulgadas para un tercero no autorizado, y tendrán que ser revocadas. Recuperando de esto el demandar el reestablecimiento de una nueva forma de compartir llaves con todos los usuarios en el dominio. Dicha recuperación es muy costosa en términos de tiempo, esfuerzo y recursos económicos.

El papel de las entidades intermediarias de confianza en la autentificación

Los sistemas basados en llave pública confían en Autoridades Certificadoras como entidades intermediarias de confianza cuando autentifican usuarios y servidores. Centros de conmutación de llaves son otra forma de intermediarios confiables. Si la petición no es satisfecha, los usuarios pueden experimentar retardos significativos en la autentificación, o ser forzados a aceptar un incremento riesgoso de credenciales fraudulentas.

El esquema que se pretende resolver con Autentificación de Llave Pública (PKA) es como se plantea a continuación:

En esquemas tradicionales (tales como Kerberos), el KDC emite a los clientes credenciales con una vida relativamente corta (un "Ticket Granting Ticket" TGT) el cual debe ser presentado después ante un Servidor de Concesión de Acreditaciones (TGS) centralizado para obtener una acreditación para un servidor particular. El KDC es cargado, por la necesidad de una constante renovación de estos TGTs de corta-vida, y los TGS deben estar involucrando cada momento al cliente que desea establecer contacto con un nuevo servidor. Típicamente, el tiempo-de vida-corta del TGT provee la única protección contra credenciales revocadas; la replica del TGS no será informada cuando el KDC revoca los privilegios de un usuario.

Ahora bien, una solución al problema anterior con PKA es de la siguiente manera:

En el esquema de llave pública, una Autoridad Certificadora (CA) emite credenciales con una duración relativamente extendida. Cuando ambos, usuarios y servidores tienen dichos certificados, ellos pueden autentificarse entre sí, sin una referencia adicional al CA. Sin embargo, precisamente por que estos certificados son de larga-duración, son requeridos algunos métodos para informar a los servidores de certificados revocados. Esto puede ser hecho por el requerimiento de servidores que revisen la actual validez de los certificados con la CA en cada uso de certificado, o por una distribución de listas de certificados de revocación (CRL´s) a todos los servidores periódicamente.

Por lo tanto, tales, factores controlan los cargos en los servicios centralizados en ambos casos. En Kerberos, el cargo excesivo en los KDC / TGS es determinado por él numero de veces que

los clientes quieren autentificarse con los servidores. En un esquema de llave pública, esto esta determinado por la frecuencia con la que los clientes o servidores deben estar en contacto con una CA para obtener una lista de certificados revocados.

Al repartir a los usuarios dentro de los servicios del dominio[26], mediante diferentes servidores se obtiene un método para implementar escalabilidad. Esto debe ser acoplado con un intermediario para autentificación en condiciones de interconexión de dominios. (Ver articulo de A. Torres et al [29]). La autentificación de interconexión de dominios se verifica a través del establecimiento bilateral de llaves inter-dominios por administradores de estos. Esto ultimo se realiza con la finalidad de registrar los Servidores de Concesión de Acreditaciones de un dominio como un usuario en el otro. Esta división de dominios determina reinos que pueden, en teoría, ser organizados jerárquicamente. Acuerdos bilaterales deben ser establecidos a priori y unos cuantos reinos son registrados al cruce en practica. Con una cadena de certificados permite al usuario una verificación del certificado emitido por una CA diferente de la propia.

Recientemente han surgido numerosas propuestas para incorporar criptografía de llave pública dentro de Kerberos [28,30,31,32]. Estas propuestas se concentran en varios aspectos de Kerberos, tales como seguridad y portabilidad. El KDC centralizado permanece en todas las propuestas. El trabajo realizado en esta parte de la tesis, extendido y generalizado desde el protocolo Netbill de seguridad y transacción [33], busca ambas direcciones, la escalabilidad y seguridad concerniente por el paso del centralizado KCC enteramente, además de enfocarlo a un ambiente móvil. Esta extensión propuesta para este esquema será referida como " KCC basado en llave pública para autentificación distribuida" o "DKCC" en el resto del documento.

Neuman et al, proponen para resolver el problema de seguridad, el empleo de criptografía de llave pública en la autentificación inicial entre los clientes y el KCC. Registrándose solamente las llaves públicas con el KCC, los clientes no tendrán que regenerar un nuevo intercambio secreto en el caso de un compromiso KCC. Solamente los servidores de aplicación, los cuales continuaran usando criptografía convencional, tendrán que reestablecer nuevas llaves simétricas. Limitando el uso de criptografía de llave pública para la autentificación inicial es justificado en el desempeño de territorios. Intercambio de llave de una vez es realizado durante la autentificación inicial, todos los intercambios de mensajes subsecuentes, incluyendo aquellos para el servidor de certificados seguros desde el servidor de concesión de credenciales, pueden ser logrados empleando él computo más eficiente en métodos de llave simétrica.

La extensión del DKCC propuesto requiere del empleo de operaciones de llave pública cada momento que un servicio de certificado es requerido. Sin embargo, estas operaciones son distribuidas entre los usuarios y servidores, prefiriendo a concentrarlos en el KCC. Más aún, esto ofrece una solución más completa al problema de seguridad de autentificación completamente distribuida entre los usuarios del esquema y servidores empleando intermediarios de criptografía de llave pública, que ni los usuarios ni los servidores necesiten mantener llaves simétricas con el KCC. Incluso, no hay una amplia KCC centralizada a ser comprometida. Solamente la CA queda como intermediario de confianza.

---

[26] Conjunto de dispositivos de comunicación y usuarios, generalmente, coincide con una organización o área especifica.

3.2.2.2. Fundamentos del Protocolo

De acuerdo con los criterios de diseño del primer protocolo y extendiéndolos de manera apropiada al ambiente distribuido, así como las propiedades de los sistemas de comunicación móvil se crea el siguiente fundamento del protocolo, el cual es descrito a continuación:

Notación

La siguiente notación será empleada para denotar las partes, operaciones, y llaves variables involucradas en el intercambio de mensajes descritos en este protocolo:

$Cert_A$	Certificado de Autentificación
α	Base del problema de logaritmos discretos D-H
m_S	Modulo de D-H
A_s	Llave de sesión dependiente de la localidad solo utilizable en el dominio (R)
K_s	Llave de sesión de corto uso
m_A	Modulo A_S
p_A	Numero primo para generar las llaves A_S
q_A	Numero primo para generar la llave AS
n_p	La llave pública del KCC
m_n	Modulo de la PK del KCC
iddR	Identificación del dominio remoto
idA, idB	Identificación de usuarios A y B
$m\alpha$,	Primo grande para firma digital y /o parámetro de curva elíptica
NR	Nonce generado por el visitante del dominio
AS_R	Servidor de Autentificación del dominio visitado
U	Usuario
S	Servidor
CA	Autoridad Certificadora
K_A	Llave simétrica aleatoria de una sola vez
K_{S-U}	Llave simétrica compartida por S y U
K_{PubS}	Llave pública de S
K_{PrivU}	Llave privada de U
$\{M\}K_x$	Mensaje encriptado empleando llave K
$\{M\}\ K_{PubS}$	Mensaje encriptado empleando llave pública de S
$K_{PrivU}\ \{M\}$	Mensaje firmado empleando llave privada de U
T_{exp}	Tiempo expiración
Ti_{auten}	Tiempo inicial de autentificación
C_{S-U}	Certificado para la sesión entre S y U

3.2.2.3. Consideraciones de Celular / inalámbrico. En una elevada dinámica de ambiente inalámbrico donde los usuarios frecuentemente cruzan los limites de dominios en medio de las comunicaciones, esto es crucial para la transferencia de estados necesarios entre dominios en forma transparente al usuario. El mismo problema también ocurre cuando los usuarios emigran entre diferentes células dentro del mismo dominio. Sin embargo en el último caso, la autentificación no es el asunto.

En GSM [8′,8′′], por ejemplo, se hacen previsiones para una rápida transferencia de autentificación de información de usuarios entre dominios. Por consecuencia, en GSM el dominio de "host" suministra a los dominios externos con un conjunto de pares de "challenge / response"

En el esquema del protocolo básico, una comparación en él último paso es crucial. Esto completa el ciclo del protocolo para autentificación del Servidor de Autentificación remoto (AS_r) a ambos, KCC y el usuario (U) simultáneamente. Luego, AS_r procede a instalar las credenciales temporales del usuario en la base de datos subscriptor / usuario.

Él último flujo de estos, es estrictamente opcional. Puede ser utilizado para presentar "sing-on" para usuarios basados en la recepción de información desde KCC, esto es, para establecer una sesión de llaves de trabajo entre AS_r y U. El certificado es calculado bajo la nueva adquisición de A_s y contiene una fuerte sesión de llave (K_s) que el usuario puede utilizar inmediatamente.

Para el subsiguiente acceso a red en el mismo dominio, el usuario externo puede ser autentificado vía ordinaria (single sing-on protocol) usando la misma A_s. [27] Ver figura 3.7

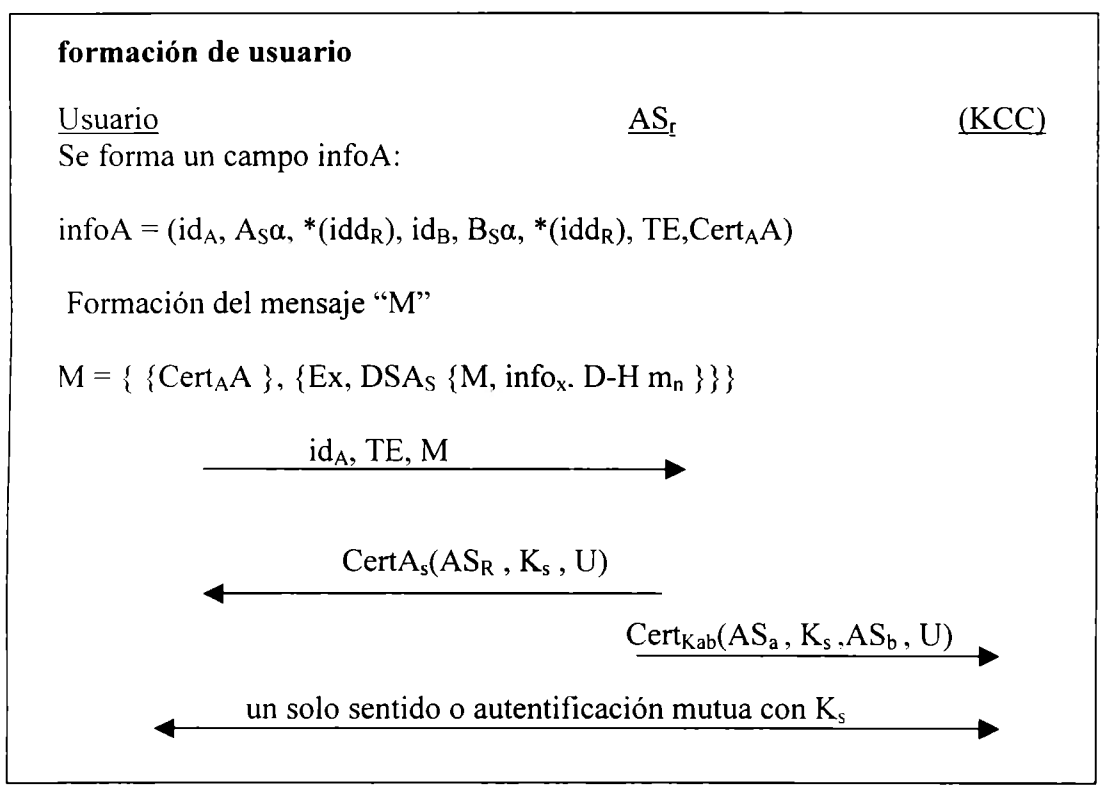


Figura 3.7 Rápida entrega (hand-over) de información de autentificación.

Cada par es bueno para una autentificación de un tiempo de usuario. En cualquier tiempo el usuario se mueve desde dominios externos A hacia el dominio B, AS_a permite reenviar el par no empleado "challenge/response" a AS_b. Esto permite a AS_b autentificar al usuario (o más bien al usuario (SIM) inmediatamente sin considerar contactar al dominio Local. De esta manera, cuando todo el grupo de la tripleta de autentificación haya sido agotado, no habrá forma para los

[27] Esencialmente, esto significa que solamente los flujos 1 y 4 de esta parte del protocolo serán ejecutados

dominios visitados (AS_b) de autentificar a las visitas de los usuarios sin contactar el dominio local nuevamente.

La figura 3.6 describe un protocolo de rápida entrega (hand-over) que evita una nueva visita del dominio (AS_b) para contactar al KCC durante una comunicación. Los dos primeros flujos representan una forma normal de firma de usuario en dominios externos. Estos flujos son idénticos a los flujos 1 y 4 en el protocolo básico. Posteriormente, el usuario cruza los limites dentro de dominio externo B adyacente. En lugar de contactar inmediatamente al Servidor de Autentificación local o al KCC (AS_L, la cual esta potencialmente muy lejos), AS_a reenvía a AS_b un certificado que contenga la misma llave K_s que fue distribuida para el usuario en el flujo 2 de la ultima firma de red en A. Conociendo K_s permite a AS_b para autentificar al usuario inmediatamente y directamente.[28]

Sin embargo, este protocolo es solamente útil por medidas temporales. Esto supone que el próximo tiempo de usuario atienda al acceso de red en el dominio B.

Ahora si tomamos que sobre la base del protocolo básico, un procedimiento de autentificación normalmente ejecutado comenzaría con el intercambio de los siguientes 5 mensajes:

1. U ——————→ AS: SSA (Solicitud de Servicio de Autentificación)

2. U :RSA (Respuesta de Servicio de Autentificación) ←—————— AS

3. U ——————→ SCC: SSCC (Solicitud de Servicio de Concesión de Certificado)

4. U :RSCC (Respuesta de Servicio de Concesión de Certificado) ←—————— SCC

5. U ——————→ S: AP_SOL (Solicitud de Servicio de Aplicación)

Al referirnos a la autentificación distribuida de llave pública, se podrían eliminar los pasos 3 y 4 con algún esquema fortificado de llaves.

El usuario primero obtiene una CCA (Concesión de Certificado de Acreditación) desde el servicio de autentificación (AS) del KCC, con el cual un intercambio de llave secreta ha sido previamente establecido. Usando este CCA, el usuario se comunica con el SCC para una segura sesión de llave compartida entre el mismo y el servidor con el cual desea comunicarse. El usuario puede entonces proceder a solicitar el deseado servicio desde el servidor de aplicación en el paso 5. Esta secuencia del intercambio del mensaje es preservado en la extensión propuesta por Neuman et al [28]. El usuario continuara recibiendo y empleando CCA´s convencionales y servicio de acreditaciones.

En el protocolo propuesto, el DKCC se comunica directamente con el servidor de aplicación, el usuario estará posibilitado para solicitar los certificados de los servidores y establecer un certificado de sesión sin necesariamente contactar al intermediario centralizado. Un servidor DKCC-habilitador permitirá solicitudes de servicio de certificados tan buenos como la sesión de

[28] El certificado reenviado a AS_b debe tener un tiempo de vida muy corto.

certificados emitidos en la capacidad del SCC, a pesar de ser solamente para sesiones con ellos mismos:

1. U ——————————▶ S: SCS (Solicitud del Certificado del Servidor)

2. U: PCS (Provisión del Certificado del Servidor) ◀—————————— S

3. U ——————————▶ S: DKSCC_SOL (Llave Pública basada en solicitud de SCC)

4. U: DKSCC_RESP (Llave Pública basada en respuesta de SCC) ◀—————— S

5. U ——————————▶ S: AP_SOL (Solicitud del Servidor de Aplicación)

De lo anterior se puede observar que el protocolo DKCC puede ser ejecutado de una manera completamente distribuida. Los primeros 2 pasos requieren operaciones no criptográficas, ya que la información que esta siendo transferida es toda información pública y la integridad de los certificados es verificada en los subsecuentes pasos. Otra vez, la criptografía de llave pública esta limitada para autentificación inicial solamente, lo cual ocurre en el paso 3 de este protocolo. Mientras el usuario genera y envía una llave pública basada en solicitud del SCC en el paso 3, se recibe un servicio de acreditación convencional en el paso 4 y proceden operaciones normales desde el paso 5 en adelante.

3.2.2.4 Obteniendo los certificados de llave pública de los servidores

El cliente inicia el intercambio de autentificación por la solicitud desde el servidor de certificados de llave pública del servicio de aplicación. Esto es necesario para la construcción de los subsecuentes mensajes certificados solicitados (DKSCC_SOL) requieren la encripción de datos usando los servidores de llave pública. El mensaje SCS simplemente consiste de la identidad (nombre principal y dominio) del servidor:

SCS : S

Este y todos los subsecuentes mensajes de autentificación DKCC son dirigidos a los servicios asignados al puerto de aplicación.

En respuesta a la solicitud, el servidor regresa este certificado o cadena de certificados, los cuales pueden ser transmitidos vía un canal desprotegido:

PCS : s-cert

Si el cliente tiene capacidades de almacenamiento en memoria inmediata del certificado, los 2 pasos de arriba pueden ser omitidos para intentos de autentificación subsecuentes con un servidor.
El cliente es responsable de verificar que el certificado ha sido firmado por una CA confiable o cadena de CA's, y este no ha sido posteriormente revocado. Esto puede también ser logrado comprobando contra una copia actualizada regularmente de la lista de certificados revocados (CRL), o realizando un pregunta a la autoridad certificadora (CA). Alternativamente, el cliente puede escoger asegurar los certificados de los servidores directamente desde la CA. En todo caso,

el cliente debe tener establecida comunicación segura con la CA, por ejemplo empleando DKCC con las llaves públicas bien conocidas de los CA´s

3.2.2.5. Autentificación cliente / servidor (estación base) utilizando criptografía de llave pública.

Llave pública basada en solicitudes SCC

Una vez que el cliente ha obtenido y verificado el certificado de llave pública, puede proceder a generar la solicitud de servicio de certificado. El mensaje contiene información similar a la de una solicitud de certificado convencional, pero es enviada al servidor directamente, en vez del KCC como en el Kerberos tradicional. Este mensaje es firmado digitalmente con la llave privada del cliente, y encriptado con la llave pública del servidor. Por lo tanto, el servidor y solamente el servidor puede determinar y autentificar la identidad del cliente. Recíprocamente, el cliente es asegurado de la identidad del servidor por que solamente el servidor con la correspondiente llave privada puede desencriptar la DKSCC_SOL y construir una respuesta valida.

Los campos críticos de los mensajes DKSCC_SOL son como se muestran:

DKSCC_SOL : S, $\{Ti_{auten}, K_A,$ auten-data $\}$ K_{pubS}
Donde:
auten-data = U, u-cert, $\{ K_A, S, K_{pubS}, Ti_{auten}\}$ K_{PrivU}

La identidad del servidor, S, es el único campo transmitido en claro. Todos los otros campos son encriptados para cualquiera de los dos; seguridad o razones de privacidad. La integridad del campo S es garantizada por esta inclusión en el campo de autorización, el cual es digitalmente firmado. La identidad del servidor debe estar en claro, por eso el oyente de los procesos recibiendo los mensajes DKSCC_SOL sabe para cual usuario el mensaje es intentado en el momento que múltiples usuarios son atendidos desde el mismo puerto de servicio.

El regreso del mensaje es encriptado empleando la llave pública del servidor, K_{pubS}. (en la practica esto significa, encriptar el mensaje con una llave simétrica la cual en su turno es encriptada empleando K_{pubS}) Hay tres distintos elementos en esta porción de encripción, especialmente tiempo inicial de autentificación Ti_{auten}, la llave aleatoria K_A, y el campo de autorización digitalmente firmado " auten-data"

El campo Ti_{auten}, índica el tiempo de la solicitud de autentificación inicial, el cual es el tiempo en el que la solicitud del mensaje actual esta siendo generado. Desde que el cliente genera este tiempo de expiración, el servidor tendrá para verificar el tiempo transcurrido entre este tiempo de expiración y cuando se recibió este mensaje. Por la negación de un servicio de solicitud de un certificado que ocurrió en el pasado "muy lejano", esto es, después de la desviación de reloj aceptable, el servidor puede prevenir ataques de replica. Este campo no debe ser enviado en claro ya que cualquier retrazo observable indicaría a un atacante potencial que el cliente podría tener problemas de sincronización de reloj, lo cual puede ser aprovechado.

El campo K_A, es una llave de una sola vez aleatoria generada por el cliente. En el Kerberos tradicional, esta llave aleatoria es generada por la KDC para el cliente para usar en comunicación con el TGS en el intercambio TGT_REQ / TGT_REP. El cliente ahora genera

esta llave aleatoria. Esta llave no es la actual llave de sesión, pero es usada bastante por el servidor para encriptar la respuesta, la cual contiene el certificado de servicio y la misma llave de sesión.
La generación de esta llave de sesión impone un cargo al cliente para tener un apropiado generador de numero aleatorio. De cualquier modo, el cliente tiene acceso a la llave privada de usuario, la cual puede ser empleada todo el tiempo con otras fuentes disponibles de entropía para dar lugar a sembrar una categoría de generador de números seudo aleatorios. Esto debería ser notado, si comparamos que Kerberos tradicional requiere de la generación por parte del cliente de un nonce, pero el nonce puede ser basado también en un generador de numero aleatorio o en tiempos de expiración. La inclusión de estas llaves aleatorias de una sola vez en el mensaje elimina la necesidad de un nonce separado.

El tercer y final elemento, "auten-data", es en esencia un campo de autorización digitalmente firmado con la llave privada del cliente. Este campo contiene la información necesaria para autentificar la identidad de clientes y para revisar la integridad del mensaje.

La construcción del "auten-data" merece especial atención. Para representar el campo "auth-data" se podría emplear la construcción "Signed-Data" como una especificación en los PKCS, Sintaxis Estándar de Mensajes criptográficos (PKCS # 7) [14]. El contenedor de "Signed-Data" incluye no solamente un lugar propietario para el contenido a ser firmado, sino además lugares propietarios para la identidad U del cliente, y los certificados del cliente "u-cert" (todo el tiempo otros campos de soporte necesitan para operaciones de llave pública.) Estos son por consecuencia, no necesarios para duplicar explícitamente estos campos en otra parte en el mensaje solicitado.

La identidad de clientes y campos de certificado, donde parte de la construcción de "auten-data", no son por ellas mismas firmadas de por si. Los campos actualmente sujetos a los procesos de firmado es la llave aleatoria K_A, la identidad del servidor S, la llave pública del servidor K_{pubS}, y el tiempo de expiración Ti_{auten}. La K_A es firmada para autenticidad, mientras la identidad del servidor es incluida para prevenir ataques de replica mencionados por Denning-Sacco [35]. La llave pública del servidor (o cualquier otro identificador el cual únicamente enlaza el certificado del servidor a la llave utilizada en la encripción de este mensaje DKSCC_SOL) servirá para desviar o prevenir ataques "man-in-the-middle". Un candidato alternativo para este campo podría ser una combinación de "emisor de certificado" y los campos específicos del emisor "número serial de certificado" en un certificado X .509. El servidor, S, por la comparación de este único identificador con el mismo campo encontrado en estas propias copias certificadas, detectarán cualquier conato por un servidor S* previo para denegar los auten-data firmados del cliente en un intento de obtener un certificado en U's llamados por S. Finalmente, los tiempos de expiración Ti_{auten} son incluidos para prevenir ataques repetidos.
Hay que notar que la identidad de los clientes y estos certificados pueden solamente ser encontrados dentro de la porción encriptada del mensaje. Lo mismo es cierto para los mensajes DKSCC_RESP a ser descritos en la siguiente sección. Esto esta en contraste con los mensajes convencionales SCC_REQ y SCC_REP, donde la identidad del cliente es transmitida en "cleartext".

Respuesta a una solicitud de DKSCC

El servidor, en esta capacidad de servidor de concesión de acreditaciones SCC, responde con un mensaje DKSCC_RESP muy similar al mensaje SCC_REP:

$$DKSCC\_RESP: T_{SU'} \{U, S, K_{S-U'}, Ti_{auten}\} K_A$$

Este mensaje, como el SCC_REP, consiste del certificado de servicio T_{SU} y una parte encriptada. El certificado es tan solo un certificado convencional, idéntico al emitido por el SCC tradicional:

$$T_{SUs} = S, \{ K_{S-U}, U, Ti_{auten} \} K_{s'}$$

Aquí, el certificado es encriptado usando $K_{s'}$, una llave simétrica conocida únicamente por el servidor. Esto previene al cliente de modificaciones del certificado. En Kerberos tradicional, la llave simétrica es compartida entre el servidor y el TGS. Claro que, el servidor y el SCC son la misma entidad en el DKCC. Esto refuerza el hecho de que el servidor puede solamente emitir certificados de sesión a clientes para comunicarse con ellos mismos, y no con cualquier otra aplicación del servidor. Esto contrasta con el Kerberos tradicional, donde el TGS centralizado puede emitir ticket de sesión para cualquier aplicación de servicio que fue registrada así misma con el TGS.

Moviéndonos hacia la parte encriptada del mensaje, se puede notar que la identidad del cliente, U, no es transmitida en grandes distancias en "plaintext", como es el caso en Kerberos tradicional. El propósito de este par de cambios es para ofrecer un alto grado de protección de la privacidad del cliente. Aún cuando estas propuestas no evitan observadores de red, esto evita rastreos de solicitudes de sesión entre el cliente identificable y el par de servidores, como es el caso con Kerberos V5.

La llave de sesión K_{S-U}, es también incluida en la respuesta encriptada. El cliente utilizara esta llave de sesión para construir el exacto autentificador, como en el Kerberos tradicional.
Es importante notar que la llave de encripción empleada en DKSCC_RESP es la llave aleatoria K_A, extraída del mensaje DKSCC_SOL. Esta llave realiza la misma función que la llave de sesión extraída desde el TGT en el Kerberos tradicional. Mas aún, ya que K_A es una llave generada una sola vez por el cliente, esto puede y sirve un segundo rol como el nonce.

Empleando el servicio de certificado

El servicio de certificado recibido por el cliente es simplemente un servicio convencional de certificado. Por lo que, el cliente genera la solicitud de servicio de aplicación como antes:

$$AP\_SOL: T_{SU} \{U, T_{expl}\} K_{S-U}$$

Todas las operaciones desde este punto pueden proceder por operaciones normales de Kerberos. Autentificación directa entre cliente y servidor.

3.2.2.6 Comentarios

En resumen, el protocolo presentado maneja la idea de distribuir la autoridad de certificados considerando de la situación donde los usuarios definen nuevos procedimientos de seguridad, evitando el hecho de diseminar la información privada. Se plantea la idea de emplear criptografía de llave pública para una completa autentificación distribuida de usuario móvil

3.2.3 PROTOCOLO DE AUTENTIFICACIÓN DE USUARIOS MÓVILES EMPLEANDO AUTENTIFICACIÓN MUTUA Y LLAVES PÚBLICAS DISTRIBUIDAS

En la primer propuesta [1] presentamos un protocolo de autentificación de usuario móvil en un ambiente que consideraba "inter-dominios" con algunas ventajas para la autentificación de los usuarios empleando una modificación del esquema bien conocido de firma digital con llave pública; para evaluar la seguridad este protocolo emplea los certificados para los procedimientos entre los usuarios y las estaciones base. En una segunda propuesta [2] presentamos una variación del protocolo mencionado anteriormente que considera la situación de autoridades de certificación distribuidas. Mientras que en un tercer planteamiento [3] describimos un protocolo de usuario móvil empleando autentificación mutua y criptografía de llave pública distribuida. Este protocolo se piensa para los usuarios con alto grado de movilidad; donde el procedimiento de la autentificación, no emplearía los certificados debido a la movilidad experimentada por el usuario. Para alcanzar esta idea desarrollamos un método que combine otros protocolos de la autentificación del usuario.

3.2.3.1. Introducción

En la figura 3.8 se ilustra el ambiente donde se ejecuta el protocolo que proponemos. En un ambiente inalámbrico altamente dinámico donde los usuarios frecuentemente cruzan él limite de dominio en el centro de las comunicaciones, esto es crucial para la transferencia de los estados necesarios entre los dominios en forma transparente al usuario. El mismo problema también sucede cuando los usuarios emigran entre diversas células dentro del mismo dominio. Sin embargo, en el último caso, la autentificación no es la cuestión.

Nuestro protocolo utiliza las funciones unidireccionales f y g, las cuales son difíciles de invertir, y además tienen las siguientes características; $g(f(x), f(y)) = f(g(x, y))$.

En esta propuesta, desarrollamos un método que combina otros protocolos de la autentificación del usuario. Los protocolos han mostrado algunas eficiencias discutidas previamente en la literatura; básicamente el protocolo de la autenticidad mutua (MAP) desarrollado por C. Cavaiani y J. Alves-Foss [36], los trabajos de W. A. Wulf, A. Yasinac, K.S. Oliver y R. Peri [37] y los protocolos de Bellovin y de Merrit (A-EKE) [38,39], en la mejora de los KEA-Diffie Hellman que utilizan dos protocolos.

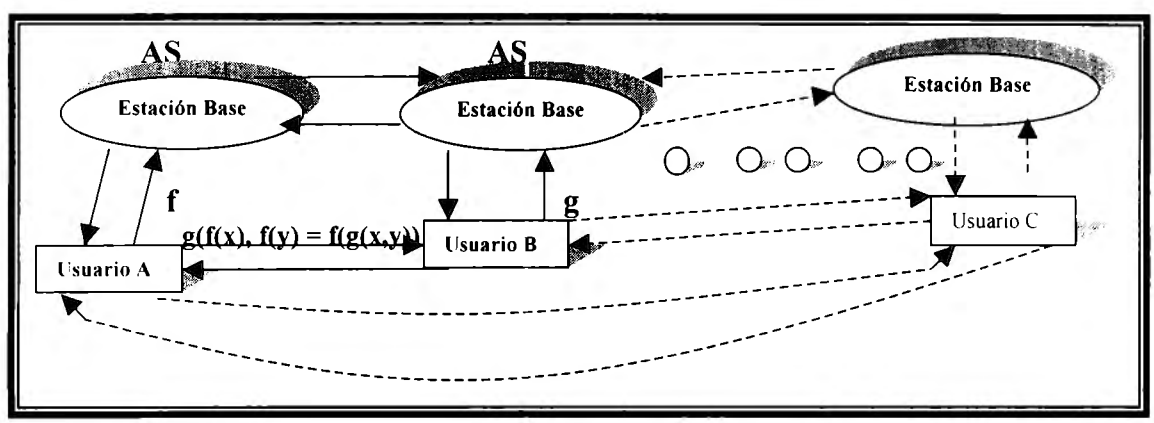


Figura 3.8 Ambiente de ejecución del protocolo

3.2.3.2 Revisión de protocolos y métodos existentes.

A continuación, comentaremos otros de los protocolos de autentificación mutua existentes, y diferentes a los ya revisados en él capitulo 2, mismos que tomamos de base para nuestras propuestas.

Protocolo de Autentificación Mutua Básico.

El esquema de Diffie-Hellman [40] (protocolo básico de KEA) evita que un ataque de escucha no autorizada comprometa la llave común invirtiendo la función para determinar la llave k. Sin embargo, éste todavía no es inmune para a un intruso de "spoofing" a través de un ataque "man-in the- middle."

Este esquema para derivaciones de llaves limita el impacto de un compromiso del criptosistema. Si un no criptosistema de Diffie-Hellman está quebrado, o si se compromete la llave privada, entonces todas las llaves primarias y tráfico de mensaje protegido bajo sistema de llave privada son comprometidos. Sin embargo, si es comprometida una derivación de Diffie-Hellman, sólo el tráfico protegido bajo una llave primaria se compromete, y la autentificación es inafectada. Una de las desventajas de este sistema, es que utiliza llaves públicas, otro, se relaciona con la movilidad y la autentificación inicial del usuario a la estación base. En [1] se explora tal situación.

Para clarificar el procedimiento que empleamos, describiremos otros dos esquemas de autentificación, reproducimos parte de estos para distinguir; sus características principales.

Esquema de Autentificación de la Universidad de Virginia.

Un esquema desarrollado por W. A. Wulf, A. Yasinac, K. S. Oliver, y R. Peri de la universidad de Virginia (UofV) [37] trata la debilidad del protocolo básico. El protocolo de UofV toma la propuesta interesante de pasar funciones unidireccionales múltiples para protegerse contra el enmascaramiento, conjetura de palabra de paso, y ataques de replica.

El protocolo de UofV propone un método para proporcionar autentificación del usuario sin conocimiento compartido entre los usuarios. Esto proporciona un nivel creciente de confianza en la autentificación del usuario. Este método es ilustrado por el ejemplo siguiente:

Cuando Alicia solicita servicios de Bob, Alicia debe contestar a una pregunta que solamente Alicia sepa. Esto significa que ningún otro (ni siquiera Bob) conocen la respuesta, pero Bob debe poder verificar que la respuesta recibida de Alicia es correcta.

Este método quita la necesidad del uso tradicional de una llave secreta o de la palabra de paso. Más aún, el problema del almacenaje de palabras de paso o las llaves secretas también se elimina.

El UofV se basa en dos características principales:

- Funciones unidireccionales f y g, difíciles de invertir
- Estas funciones f y g tienen las propiedades conmutativas: $g(f(x), f(y)) = f(g(x, y))$. Aunque las funciones f y g no pueden ser hechas públicas, los autores del protocolo de

UofV argumentan que las formas de f y g se pueden hacer públicas. Esto permite a cada usuario elegir una función apropiada para f en privado y no requiera f ser pasado sobre un canal o utilizar un servicio de autentificación de una tercera parte.

Intercambio de llave de encripción aumentado.

Otro esquema que proporciona un método excelente para la autentificación mutua es un protocolo desarrollado por S. M. Bellovin y M. Merritt [38,39] de los laboratorios de la AT&T Bell. El protocolo de intercambio de la llave de encripción aumentado (A-EKE) permite a dos partes que comparten una palabra de paso para comunicarse sin exponer la palabra de paso compartida. El protocolo mutuo de autentificación utiliza funciones hash unidireccionales conjuntamente con la encripción de la llave pública. El A-EKE presenta un mecanismo en el cual un host principal no tenga que salvar palabras de paso en texto claro, y lo protege contra enmascaramiento, diccionario, y ataques hombre-en-el-medio. Bellovin y Merritt [39] presentaron su protocolo como una combinación nueva de la criptografía simétrica y asimétrica que permite a las dos partes que comparten una palabra de paso común intercambiar información confidencial y autentica sobre una red sin garantía. Desde entonces, han modificado su protocolo para incluir el uso de las funciones hash [38]

3.2.3.3. Autentificación mutua con protocolo de llaves distribuidas para los usuarios móviles.

En este protocolo también se parte de las mismas suposiciones iniciales y criterios de diseño plantados para el primer protocolo.

Protocolo de Autentificación Propuesto.

Se propone un protocolo de autentificación de usuario móvil empleando la autentificación mutua y llaves públicas distribuidas. Para proporcionar un nivel de seguridad del sistema y para satisfacer las metas básicas siguientes del diseño:

- Autentificación mutua entre los usuarios o entre el usuario y la estación base.
- Mínima información transferida para la autentificación mutua.
- Llaves autentificadas de palabras de paso (PAKs.)
- Protección de palabra de paso durante el tránsito sobre una red sin garantía.
- Fácil de utilizar y protocolo eficiente para el uso seguro sobre una Red de Comunicación Personal.

Fundamentos del protocolo

De Acuerdo con los criterios del diseño y las características de los sistemas móviles de la comunicación, creamos el siguiente fundamento del protocolo, que a continuación será descrito:

Las siguientes notaciones son empleadas en este protocolo.

Atributos del usuario.

α	valor no repetido
β	valor no repetido
id	Identificación de los usuarios (password) A y B
R_A	Número primo grande
R_B	Número primo grande
$P_A(\alpha)$	Valor de la función del logaritmo discreto
P_B	Función de un solo sentido
k	Llave de sesión
h(id)	Función hash de un solo sentido del Password
h	valor del hash
m_S	Modulo de D-H
A_S	Llave de sesión secreta
m_A	Modulo de A_S
idd_R	Identificación del dominio remoto
$m\delta$,	primo grande para firma digital o un parámetro de curva elíptica
N_R	Nonce generado por el dominio del visitante
AS_R	Servidor de Autentificación del dominio del visitante

Ejecución y protocolo básico.

El protocolo básico se representa en el cuadro 3.8 Ahora nos abocaremos a los detalles del protocolo:

1. En esta parte, "A" solicita el acceso a "B", utiliza la función hash (identificada con el símbolo "h") para aplicarla a su identificación para producir el h(id). "A" también genera un valor no-repetido "α" y el número primo grande R_A. Después, calcula un valor funcional del logaritmo discreto $P_A(\alpha) = \alpha^{R_A} \bmod n$, encripta $P_A(\alpha)$ por el "hashing" con el h(id) para producir h(id) $[P_A(\alpha)]$. Esto evita que un intruso obtenga el valor de la función $P_A(\alpha)$ que será utilizado más adelante para generar una llave de sesión "k". "A" también encripta la α con el h(id) para producir el h(id) $[\alpha]$. Entonces "A" envía h(id)$[P_A(\alpha)]$ y el h(id) $[\alpha]$ a B.

2. Cuando "B" recibe h(id)$[P_A(\alpha)]$ y h(id)$[\alpha]$, selecciona una función unidireccional P_B que sea conocido solamente para "B", genera un valor β no-repetido, y un número primero grande R_B. "B" recupera el valor del h(id) que se guardo, y desencripta h(id)$[P_A(\alpha)]$ y el h(id) $[\alpha]$ para obtener $P_A(\alpha)$ y α respectivamente. Luego "B" toma el valor α de "A" y el valor β, y genera los valores de la función: $P_B(\alpha) = \alpha^{R_B} \bmod n$, $P_B(\beta) = \beta^{R_B} \bmod n$, y $P_B(\alpha,\beta) = (\alpha\ \beta)^{R_B} \bmod n$. Después B genera una llave de sesión k, tomando los módulos n del producto $P_B(\beta)$ y $P_A(\alpha)$, $k = (P_A(\alpha) * P_B(\beta)) \bmod n$. B utiliza la llave de sesión k, para encriptar los valores α y β. Estos valores serán utilizados para verificar que ninguna intrusión ha ocurrido entre A y B. B utiliza h(id) para encriptar las funciones: $P_B(\alpha)$, $P_B(\beta)$, y $P_B(\alpha,\beta)$. Después "B" envía a "A" h(id)$[P_B(\alpha), P_B(\beta), P_B(\alpha,\beta)]$ y k $[\alpha,\beta]$.

3. A desencripta $P_B(\alpha)$, $P_B(\beta)$, y $P_B(\alpha,\beta)$ y genera la llave de sesión k, por $k = (P_A(\alpha) * P_B(\beta)) \bmod n$. A desencripta k $[\alpha,\beta]$ y verifica que α sea igual que la enviada por B para asegurar que ninguna intrusión ha ocurrido. Entonces A genera tres valores de la función: $P_A(\beta) = \beta^{R_A} \bmod n$, $P_A(\alpha,\beta) = (\alpha\beta)^{R_A} \bmod n$ y $P_A(P_B(\alpha,\beta) = (P_B(\alpha,\beta))^{R_A} \bmod n$ (el encriptado adicional de valores

funcionales no es necesario después de este punto). A encripta β con la llave de sesión k. Entonces A envía $P_A(\beta)$, $P_A(\alpha, \beta)$, $P_A(P_B(\alpha,\beta))$ y k[β] para B.

4. B desencripta k[β] para verificar si es el mismo β que él envió A y que no ha ocurrido ninguna intrusión. B calcula $P_B(P_A(\alpha), P_A(\beta))$ y compara el valor a $P_A(P_B(\alpha,\beta))$ y si tiene éxito la comparación, entonces A es autorizado para tener acceso al sistema. B genera la función $P_B(P_A(\alpha,\beta))$ y la envía a A.

5. A calcula $P_A(P_B(\alpha), P_B(\beta))$ y compara el valor con $P_B(P_A(\alpha,\beta))$ para autentificar a B, y se aseguran que ella no esté teniendo un ataque de "spoof".

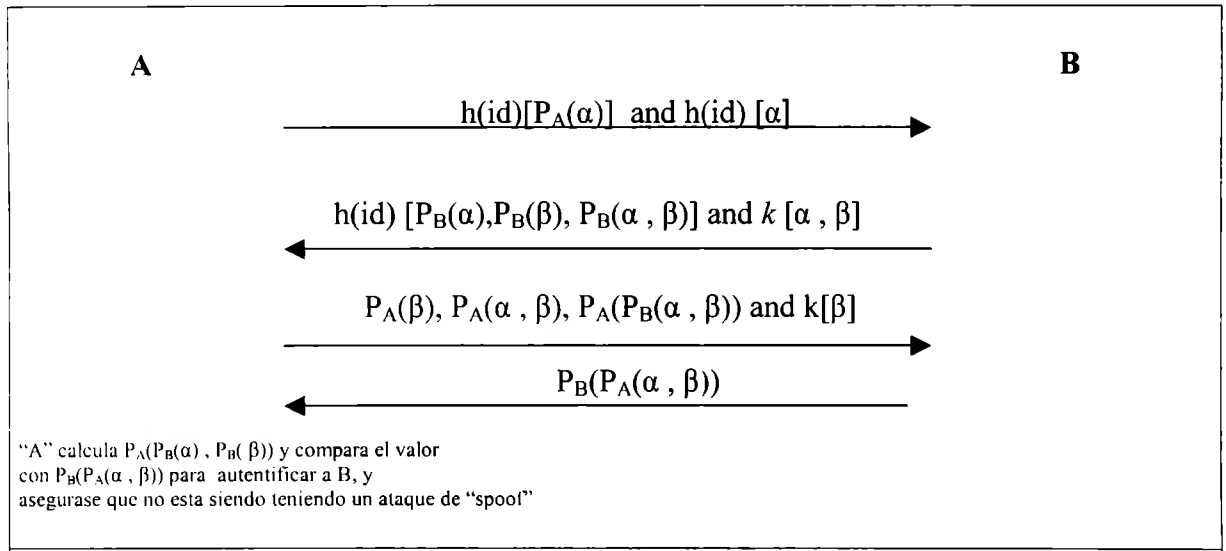
6. El usuario comienza por formar un campo de información X, el cual incluye; id_A, $A_S\delta$, idd_R, id_B, $B_S\delta$, Tiempo de expiración (TE), $P_B(P_A(\alpha,\beta))$. Posteriormente forma el mensaje como se muestra a continuación:

$S = \{\{P_B(P_A(\alpha,\beta))\}, \{Ex, DSA_S \{S, info_x, D\text{-}H\ m_n\}\}\}$, donde, el usuario emplea su $P_B(P_A(\alpha,\beta))$. calcula la autentificación DSA_S {S, $info_x$, D-H m_n } con una llave dependiente de la localidad calculada para el dominio visitado (una llave que el usuario puede emplear solamente en el dominio visitado). Como es conocido, existen muchos caminos para construir este tipo de llaves. En nuestro caso, emplearemos encripción criptográfica. El empleo de encripción en la formación de estas llaves, es esencial en orden de la satisfacción de la restricción de la separación de dominios previniendo la derivación de la clave compartida entre el usuario y el servidor de autentificación del dominio local desde la llave dependiente de la localidad. Posteriormente, se forma el mensaje y es enviado a la estación base.

7. Cuando el dominio del visitante recibe el mensaje inicial que no se tiene ningún medio de autentificar a este usuario, porque no se reconoce que el usuario sea nativo. Entonces el dominio visitado necesita solicitar una prueba de la identidad del usuario al dominio local. Esta petición debe también autentificar el dominio del visitante a la estación base, y el usuario a la estación base.

8. Cuando la estación base recibe el mensaje, esta procede a buscar el archivo del usuario y obtiene la llave compartida entre ellos (usuario y servidor de autentificación del dominio local), valida TE; Si este es verdadero acepta, de otra manera rechaza. Empleando los parámetros dados, es necesario hacer algunos recálculos para compararlos con el correspondiente "token" que llego en el flujo pasado. Una comparación satisfactoria en el último paso autentifica al usuario y al dominio del visitante con la estación base. Entonces en este punto, la estación base pública un $P_B(P_A(\alpha,\beta))$ que confirma la identidad del usuario y permite al usuario operar en el reino del dominio visitado.

9. Cuando la estación base recibe el mensaje, procede de la siguiente manera:
Hace los recálculos de autentificación correspondientes, compara estos con los correspondientes "token" recibidos desde el usuario (primer flujo.) Una comparación en el paso anterior es decisiva, completa el ciclo del protocolo por la autentificación al dominio del visitante a ambos la estación base y usuario simultáneamente. Entonces el dominio visitado procede a instalar a los usuarios temporalmente.
El flujo pasado puede ser utilizado para establecer una llave de sesión de trabajo entre el dominio visitado y el usuario.



Execution (user formation)

Usuario — Estación Base (BS)

Se forma un campo infoX:

$infoX = (id_A, A_S\delta, *(idd_R), id_B, B_S\delta, *(idd_R), TE, P_B(P_A(\alpha , \beta)))$

Formación del mensaje "S"

$S = \{ \{ P_B(P_A(\alpha , \beta))\}, \{Ex, DSA_S \{S, info_x, D\text{-}H\ m_n \}\}\}$

id_A, TE, S →

id_A, idd_R, TE, N_R, DSA_S (N_R, S, idd_R) →

busca archivo de usuario, obtiene
la llave de usuario compartida entre
ellos, valida TE, si se Cumple acepta,
en caso contrario Rechaza,
Hace todos los recálculos necesarios y
compara con él correspondiente mensaje
que llega. Y emite un $P_B(P_A(\alpha , \beta))$

← $P_B(P_A(\alpha , \beta))$ (BS ,A_s, AS_R , U)

La estación base hace los correspondientes
Cálculos de autentificación y compara estos
Con los correspondientes recibidos desde
El usuario (primer flujo).
Entonces el dominio visitado
Procede a instalar temporalmente al usuario

← $P_B(P_A(\alpha , \beta))$ (AS_R , K_s , U)

3.2.3.4. Comentarios

El protocolo propuesto se basa, en la combinación de otros ya existentes. De hecho, el Protocolo de Autentificación Mutua [36] y DPKP [2] que consideramos conveniente volver a retomar el modelo de estos trabajos para aplicarlo en el área de nuestro interés, como la autentificación del usuario y la seguridad para los ambientes inalámbricos. Este protocolo podría proporcionar un método para una autentificación mutua y asegurar tráfico del mensaje entre muchos usuarios móviles. Proporciona pasos mínimos para los usuarios al implementarse y podría ser puesto en ejecución a través de diversas plataformas y de sistemas operativos. Como fue mostrado, estas combinaciones podrían mejorar el funcionamiento y el comportamiento en ambientes múltiples.

Un escenario que este esquema no cubrió es el uso de la llave de autentificación de Password como fue hecho por A. Torres et al [41] para puentear el CA en B2B.

REFERENCIAS DEL CAPÍTULO

[1] Gustavo A. Santana. T, David Higuera R, "A Mobile User Authentication Protocol for Personal Communication Networks", paper accepted for presentation at the IASTED (International Conference on Wireless and Optical Communications (WOC 2001)). June 27 to June 29, 2001, in Banff, Canada.

[2] Gustavo A. Santana. T, Arturo Torres D., David Higuera R., " Protocolo de Autentificación de Usuario Móvil para ATM Inalámbrico" X Congreso Internacional de Computo CIC 2001, 12-16 Noviembre de 2001, Ciudad de México, D.F.

[3] Gustavo A. Santana. T, David Higuera R., "Protocolo de Autentificación de Usuario Móvil empleando Criptografía de Llave Pública Distribuida" X Congreso Internacional de Computo CIC 2001, 12-16 Noviembre de 2001, Ciudad de México, D.F.

[4] Gustavo A. Santana. T, David Higuera R, "A Mobile User Authentication Protocol Using Mutual Authentication for Personal Communication Networks", paper accepted for presentation at the IASTED (International Conference on Intelligent Systems and Control (ISC 2001)). November 19-22, 2001 Tampa, Florida, USA.

[5] Cellular Digital Packet data (CDPD) System Specification, Release 1.0, July 19, 1993.

[6] EIA/TIA-IS-54-B

[7] ETSI, ETS 300 175-7, October 1992.

[8] ETSI/TC Recommendation GSM 03.20, *Security Related Network Functio*n, version 3.3.2, Jan. 1991.

[8´] M. Rahnema, Overview of the GSM System and Protocol Architecture, IEEE Communications Magazine, April 1993.

[8´´] B. Mallinder An Overview of the GSM System, Proceedings of Digital Cellular Radio Conference, October 1988.

[9] A. Aziz and W. Diffie, "Privacy and Authentication for Wireless Local Area Network", IEEE Personal Communications, First Quarter, 1994, pp. 25-31.

[10] M. J. Beller, L. Cheng, Y. Yacobi, "Privacy and Authentication on a Portable Communication System", *IEEE J. on Selected Areas in Communication*s, Vol. 11, No. 6, pp. 821-829, Aug. 1993.

[11] U. Carlsen, "Optimal Privacy and Authetication on a portable Communications Syste*m", Operating Systems Revie*w, June, 1994.

[12] H.Y. Lin and L. Harn, "Authentications in Wireless Communications," *Proc. of GLOBECOM '9*3, pp. 550-554, Nov. 29-Dec. 2, 1993.

[13] R. Molva, D. Samfat, and G. Tsudik, "Authentication of Mobile Users", *IEEE Networ*k, pp. 26-34, March/April, 1994.

[14] Dan Brown, "Security Planning for Personal Communications", *Proc. of 1st ACM Conference on Computer and Communications Security*, pp. 107-111, Nov. 1993.

[15] J.C. Cook, and R.L. Brewster, "Cryptographic Security Techniques for Digital Mobile Phones", *IEEE International Conference on Selected Topics in Wireless Communications,* pp. 425-428, 1992.

[16] K. Vedder, "Security Aspects of Mobile Communication", Computer Security and Industrial Cryptography – State of the Art and Evolution, East Course, Springer-Verlag, May, 1991, pp. 193-210.

[17] M. Walker, "Security in Mobile and Cordless Telecommunications", Computer Systems and Software Engineering, Proceedings of CompEuro 1992, 1992, pp. 493-496.

[18] M. Spreitzer and M. Theimer, "Scalable, Secure, Mobile Computing with Location Information", *Communications of the ACM*, Vol. 36, Iss. 7, p. 27, July, 1993.

[19] S. Eckelman, "Minimizing Fraud", Telephone Engineering and Management, Vol. 94, No, 18, pp. 62-64, Sept. 1990.

[20] R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-key Cryptosystem", *Comm. of ACM*, Vol. 21, No. 12, 1978, pp. 120-126.

[21] European Telecommunications Standards Institute, Universal Personal Telecommunications, ETSI NA7 WP1, November 1992.

[22] European Telecommunications Standards Institute, Universal Personal Telecommunications, ETSI NA7 TS 02.03, January 1992.

[23] B. Preneel, et al, "Computer security and industrial cryptography" in Lecture Notes in Computer Science, pp193- 210, Leuven, Belguim, May 1991.

[24] M.J. Beller, et al, "Privacy and authentication on a portable communication network", IEEE Journal on Selected Areas in Communication, vol.11, August, 821-829, 1993.

[25] Liu Jianwei, Wang Yumin, "A User Authentication Protocol for Digital Mobile Communication Network", IEEE 1995.

[26] R. Lichota, G. Hammonds, S. Brackin. Verifying Cryptographic for electronic Commerce. 2nd USENIX Wokshop on Electronic Commerce, November 1996.

[27] M. Sirbu, J.D. Tygar. Netbill: An Internet Commerce System Optimized for Network Delivered Services. IEEE CompCon Conference, March 1995.

[28] B. C. Neuman, B. Tung, J. Wray, J. Trostle. Public Key Criptography for Initial Authentication in Kerberos. Internet Draft, October 1996. (ftp://ietf.org/internet-drafts/draft-cat-kerberos-pk-init-02.txt)

[29] G.A. Santana Torrellas and A. Torres Domínguez (Mexico) Bridge Certification Authorities: Connecting B2B Public Key Infrastructure with Password-Authentication-Keys. Thirteenth IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2001) August 21-24, 2001 Anaheim, California, USA

[30] D. Davis. Kerberos Plus RSA for World Wide Web Security. In Proceedings of the USENIX Workshop on Electronic Commerce, July 1995.

[31] B. C. Neuman, T. Ts'o. Kerberos: An Authentication Service for Computer Networks. IEEE Communications, 32(9): 33-38, September 1994.

[32] M. Sirbu, J.C. Chuang. Public key Based Ticket Grating Service in Kerberos. Internet Draft, may 1996 (ftp://ietf.org/internet-drafts/draft-sirbu-kerb-ext-00.txt)

[33] B. Cox, J.D. Tygar, M. Sirbu. NetBill Security and Transaction Protocol. In Proceedings of the USENIX Workshop on Electronic Commerce, July 1995.

[34] RSA Laboratories. PKCS # 7: Cryptographic Message Syntax Standard. Version 1.5, November 1993.

[35] D. E. Denning, G.M. Sacco. Timestamps in Key Distribution Protocols. Communication of the ACM, 24(8): 533-536, August 1981.

[36] Charles Cavaiani and Jim Alves-Foss, A Mutual Authenticating Protocol with Key Distribution in Client/Server Environment, ACM 12 June, 2001

[37] William A.Wulf, Alec Yasinac, Katie S. Oliver and Ramesh Peri. A Technique for Remote Authentication, University of Virginia, Charlottesville, VA. Available via anonymous FTP from ftp.research.att.com.

[38] Steven M. Bellovin and Michael Merritt. Augmented Encrypted Key Exchange: A Password-Based Protocol Secure Against Dictionary Attacks and Password File Compromise, AT&T Bell Laboratories. Available via anonymous FTP from idea.sec.dsi.unimi.it.

[39] Steven M. Bellovin and Michael Merritt. Encrypted Key Exchange: Password-Based Protocol Secure Against Dictionary Attacks. In *Proceeding IEEE Computer Society Symposium on Research in Security and Privacy*, May 1992: pp.72-84. Available via anonymous FTP fromftp.research.att.com .

[40] Whitfield Diffie and Martin E. Hellman, New directions in Cryptography, *IEEE Transactions on Information Theory*, vol. IT-22, no. 6, 1976:pp.644-654.

[41] G.A. Santana Torrellas and A. Torres Domínguez (Mexico) Bridge Certification Authorities: Connecting B2B Public Key Infrastructure with Password-Authentication-Keys. Thirteenth IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2001) August 21-24, 2001 Anaheim, California, USA

[42] CCITT Recommendation X.509, "The Directory-Authentication Framework", 1998

CAPÍTULO 4

4. APLICACIONES

En el ambiente inalámbrico, existen varias propuestas y estándares que actualmente están funcionando para proporcionar servicios para los PCS, sin embargo, varias de estas aplicaciones tienen deficiencias de seguridad y es aquí donde consideramos que es un área de oportunidad para la aplicación de nuestros protocolos de autentificación propuestos en este trabajo de tesis. A continuación se presenta una propuesta de "Protocolo de Autentificación de Usuario Móvil para ATM Inalámbrico", mismo que fue trabajado en conjunto con Arturo Torres, y presentado en el congreso internacional CIC 2001 [1] Posteriormente, damos una descripción general de varias aplicaciones, donde consideramos podrían ser aplicados nuestros protocolos.

4.1 PROTOCOLO DE AUTENTIFICACIÓN DE USUARIO MÓVIL PARA ATM INALÁMBRICO

Como se ha comentado anteriormente el desarrollo experimentado en la última década por los sistemas de comunicaciones personales ha permitido aumentar la cobertura y los servicios de comunicaciones en condiciones de movilidad. Sin embargo, ésta movilidad extendida genera vulnerabilidad en la seguridad de las comunicaciones.

En ésta propuesta, se considerará una posibilidad para fortalecer el esquema utilizando una autentificación por "password" con intercambio de llaves "Password-Authentication Key Exchange" verificando los requisitos de seguridad para redes "Wireless-ATM". Se propone una estrategia de autentificación que garantice a las partes la integridad de la identidad como medida de seguridad.

El protocolo PAK es el primer protocolo "Diffie-Hellman" de "Password-Authentication Key Exchange" [2] en proveer la seguridad formal en contra de adversarios pasivos y activos de manera explícita. El protocolo PAK es probablemente seguro ya que existe una posibilidad de que el atacante adivine el password generado por una tercera entidad y utilizarlo para enmascarar

a una de las partes involucradas. Pero al incorporarlo en el protocolo desarrollado en la primer propuesta presentada en él capitulo 3 [3] se fortalecerá el esquema para la aplicación de redes ATM inalámbricas.

Introducción

Desde la aparición de los sistemas de comunicaciones personales, en la década de los años 90s, se ha logrado una mayor integración de servicios de voz, datos y video. Como resultado; la demanda y necesidad de velocidades de transmisión son cada vez mayores, así mismo, la seguridad y los mecanismos de administración de la movilidad son necesidades que demandan especial atención y cuidado.

Lo que buscamos son características para desarrollar un protocolo que se realiza sobre un medio inalámbrico manteniendo una calidad de servicio aceptable. Por lo que existen muchas razones para utilizar ATM en la Red de Comunicación Personal, tales como: Ancho de banda flexible; selección de tipo de servicio para un rango de aplicaciones; multiplexado eficiente sobre tráfico; provisión de ancho de banda necesario para los servicios sobre un medio inalámbrico; mejoramiento de la confiabilidad con técnicas de QoS, etc. Una red ATM esta diseñada para proveer comunicaciones a altas velocidades para los usuarios punto a punto. Además, tiene una técnica de multiplexado de paquetes orientado a circuitos virtuales de tamaño fijo. Añadiendo una característica como inalámbrico es lo que nos permite realizar conexiones ATM entre usuarios móviles con muy buenos resultados.

Suposiciones iniciales

Asumimos las mismas suposiciones iniciales del primer protocolo.

Criterios de diseño

Al extender el protocolo [3] con el protocolo PAK (Password-Authentification Key Exchange) con autentificación explícita se toma en consideración lo siguiente:

- Generando intercambio de llaves

En el mundo ideal, el intercambio de llaves entre dos usuarios honestos es seguro, ¿pero en la red? Lo que necesitamos del protocolo de intercambio seguro (Secure key Exchange) es que, aunque sea dada parcialmente información sobre las llaves, el adversario no deberá de ser capaz de hacer algo en el mundo real que no pudiera hacer en el mundo ideal. En otras palabras, Si un protocolo de mas alto nivel es seguro en el sistema ideal con llaves generadas a través del Centro de Conmutación de Llaves (KCC), ese protocolo también deberá ser seguro si usamos el esquema de "Intercambio seguro de llaves"

El tema importante en cuestión de intercambio de llaves es que la autentificación es alcanzada basándose en una cuestión de passwords.

Ahora. ¿cómo se puede describir de manera correcta nuestro sistema ideal en el mundo de intercambio de Password-Authenticated Key? Para ello utilizamos las siguientes consideraciones de modelo:

Si dos usuarios completan el intercambio de llaves de manera exitosa, entonces un adversario no podrá obtener la llave o el "password". Esto es que el adversario no le es permitido cualquier búsqueda "test_instance_password" para un intercambio exitoso de llave. Por otro lado, el adversario es capaz o le está permitido una búsqueda de "test_password" para otro intercambio de llaves haciéndose pasar por otro sólo si el adversario adivina en forma correcta el password. Esto comprende un "On-Line Impersonation/password guessing attempt" por parte del adversario. [2]

También modelamos la habilidad para cualquier adversario para dar de alta passwords entre cualquier usuario y sigo mismo usando una búsqueda de "Set_password". Esto puede explicarse mediante la situación en que un adversario crea cuentas en cualquier computadora que desee, siempre y cuando esas cuentas tengan identificadores de usuarios "IDs" diferentes a la de todos los usuarios válidos [2.].Un esquema general se presenta en la figura 4.1

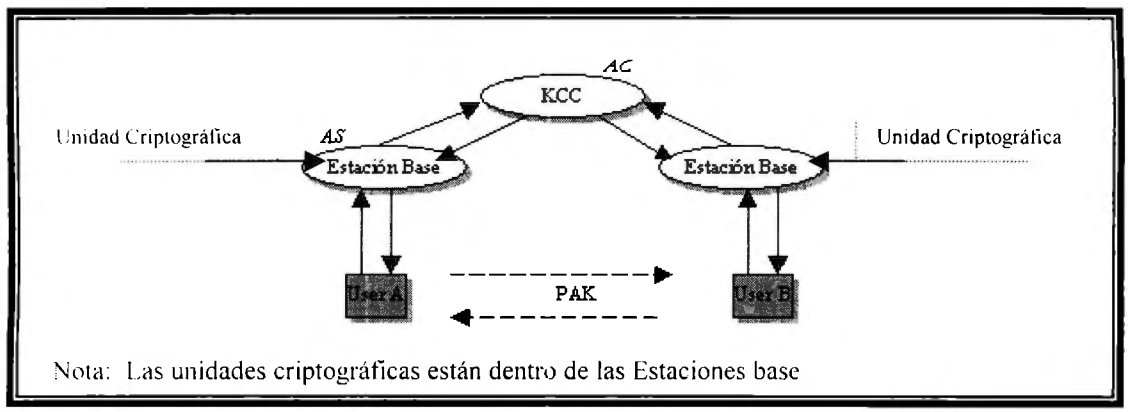


Fig. 4.1 Esquema general

De acuerdo con el Comité Técnico FrameWork1.0 sobre la seguridad en ATM, No es posible modificar los encabezados, por lo que buscamos obtener la seguridad de alguna otra forma. [4]

Debido a lo anterior incorporamos una unidad de encripción encargada de encriptar las celdas de datos, dejando los encabezados sin alterar. Eso permite privacidad en la comunicación con una gran flexibilidad sobre la red ATM sin exponer información sensible en los nodos [5].
El método más sencillo es utilizar una llave única para todo el tráfico de la sesión. Pero es necesario que incorporemos una unidad criptográfica en nuestro esquema. En la siguiente figura se puede ver gráficamente esto.

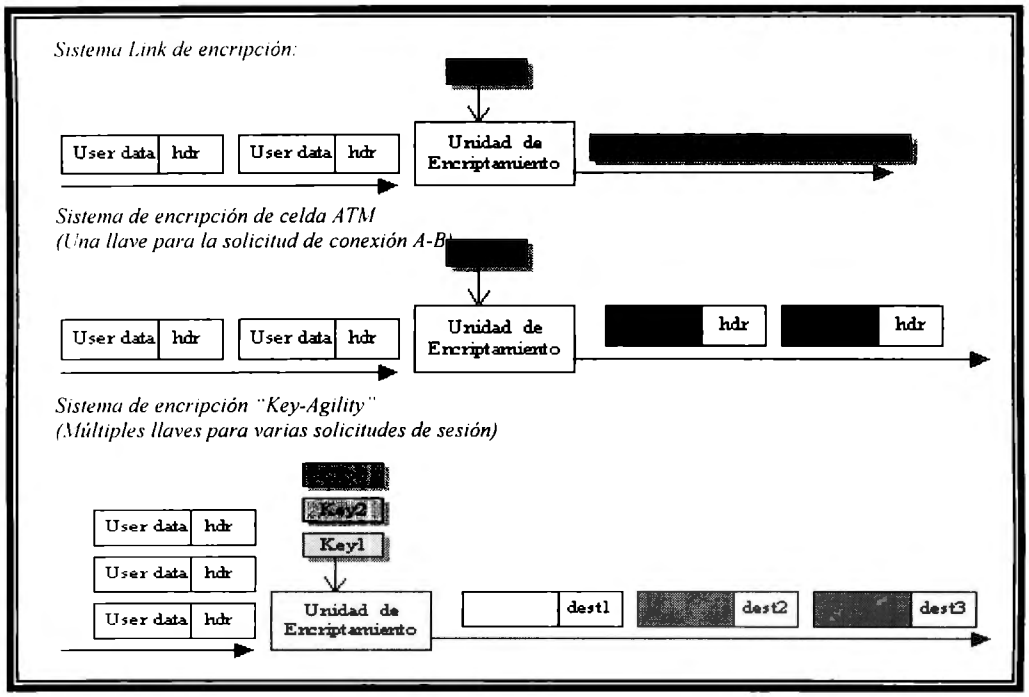


Fig. 4.2 Unidad de encripción con llave única para todo el tráfico de la sesión

Tenemos que cumplir con los requisitos establecidos para seguridad en ATM [4, 5]:

- Toda la información generada por las autoridades debe ser firmada digitalmente usando sus llaves privadas.
- La información utilizada tiene un tiempo de vida, dependiendo de la política de seguridad.
- El proceso básico de manejo de llaves para las conexiones punto a punto involucra intercambio de certificados-de-llave-pública y un número aleatorio generado especialmente para esta conexión (Solicitud de conexión.)

  Esto es seguido de un intercambio de mensajes firmados y encriptados conteniendo una llave de encripción (Encryption key) y un número aleatorio identificador que fue recibido en un mensaje de intercambio anterior.

  El intercambio de la información "nonce" previene ataques en contra del sistema de manejo o administración de llaves (Key management)

- Una vez que toda la información ha sido enviada y verificada; se usa o se tiene un sistema de seguridad de mensajes para la conexión y se manda un mensaje al (solicitante de conexión) origen para avisar que la conexión se llevó a cabo con éxito

Fundamentos del protocolo

De acuerdo con los criterios de diseño, así como las propiedades de los sistemas de comunicación móvil se crea el siguiente fundamento del protocolo, el cual es descrito a continuación:

Las siguientes notaciones son empleadas en este protocolo:

$Cert_A$	Certificado de Autentificación
α	Base del problema de logaritmos discretos D-H
m_S	Modulo de D-H
A_S	Llave secreta de sesión
m_A	Modulo A_S
p_A	Numero primo para generar las llaves A_S
q_A	Numero primo para generar la llave AS
n_p	La llave publica del KCC
m_n	Modulo de la PK del KCC
iddR	Identificación del dominio remoto
idA, idB	Identificación de usuarios A y B en forma aleatoria
$m\alpha$,	Primo grande para firma digital y /o parámetro de curva elíptica
N_R	Nonce generado por el visitante del dominio
AS_R	Servidor de Autentificación del dominio visitado
π	Denota la función de asignación de passwords a Usuario A y B
K	Llave de sesión entre A y B
DSAkcc-priv	Digital Signature Algorithm de llave KCC privada.
$Kpub_A$	Llave publica del usuario A.
$Kpub_B$	Llave publica del usuario B.
TE	TimeStamp (Tiempo de vida.)

Preliminares:

K Parámetros generales de seguridad para funciones hash y llaves secretas (128 o 160 bits)

$l > K$ Parámetros de seguridad para "discrete_log_based public key" (1024 o 2048 bits)

$\{0,1\}^*$ Serie finita de "string" binario.

$\{0,1\}^n$ Serie de strings de longitud n

$\in (n)$ Es una función real evaluada (es NEGLIGIBLE) para cada $c > 0$

Existe $nc > 0$ tal que $\in(n) < {}^1/n^c \quad \forall \quad n > nc$

Dejemos q de tamaño k y p de tamaño l ser primos tal que $p = rq + 1$ para algún valor r co-primo de q.

Dejemos g ser una generación de un subgrupo de Z_p^* de tamaño q, el grupo le llamaremos G_{pq}.

DH(X,Y) denota un valor Diffie_Hellman $g^{(X,Y)}$ de $X = g^{(x)}$ y, $Y = g^{(y)}$. Asumimos la dureza o robustez del problema DDH está en G_{pq}

Nota: Omitiremos generalmente "Mod $_p$" de las expresiones cuando es obvio que estamos trabajando con Z_p^*

Una formulación tal que, dados g, X, Y, Z en G_{pq} donde $X = g^{(x)}$ y $Y = g^{(y)}$. Son seleccionados aleatoriamente, y Z es ya sea DH(X,Y) o un valor aleatorio, cada uno con el 50% de probabilidad de ser seleccionado. Romper DDH "Z = DH (X,Y)" implica construir un "Time-polynomial adversary" que distingue Z = DH (X,Y) de la Z-random con un "non-negiligible advantege" sobre búsqueda aleatoria.

Definimos funciones Hash H_{2a}, H_{2b}, H_3: $\{0,1\}^* \rightarrow \{0,1\}^k$ y;
H_1: $\{0,1\}^* \rightarrow \{0,1\}^n$; Donde $(n \geq l+k)$
Asumimos que H_{2a}, H_{2b}, H_3 son funciones independientes.
Nótese que mientras H_1 esta descrita como la que regresa un string de bits, nosotros operaremos sobre su salida como un número Modulo p.
π es generado por el KCC.
π Denota la función de asignación de passwords a pares de usuarios.
π [A,B] = π[B,A], y por tanto tiene que ser repartida a los usuarios A y B a través de las estaciones básicas involucradas.

Con el protocolo PAK y $\pi = \pi$ [A,B], se obtiene la llave de sesión resultante es K sólo si la prueba regresa "Verdadero". Si la prueba regresa falso, entonces el protocolo aborta.

Procedimiento para KEA – PAK (Password-Authentication Key)

El KEA utilizado en los protocolos propuestos en el capitulo 3 es el mismo que se utiliza en combinación con el PAK, para este caso.

En PAK Se genera una llave aleatoria a través de un tercer ente (KCC) el cual otorga esa llave de inicio de sesión a Alice y Bob con la cual junto con valores numéricos primos seleccionados de un conjunto Z_p^* se generarán nuevos valores que a su vez confirmarán una llave de sesión. El intercambio de llaves es precisamente donde se basa el método para la autentificación con password.

Describiremos el procedimiento general que incluye precálculos, formación de usuario y seguridad (nivel) de usuario. Es importante distinguir que los procedimientos de firma digital serán implementados en ambientes de comunicación usuario-a-usuario. A continuación, presentamos las funciones para generación de llave, generación de firma y verificación de firma empleando DSA y PAK

Interpretación y ejecución del protocolo PAK

1. Primero el protocolo inicializa las variables "x" y "y´" con número primos generados y que están el Z_p^*.
Es importante destacar que "x" no es la llave secreta utilizada en la parte de firma digital. Y y´ no es la llave pública. Aunque podrían utilizarse esos valores para generar las llaves de intercambio para la autentificación a través de PAK. La razón por la cual no se utilizan en este artículo es precisamente para mantener la modularidad del protocolo y que sean independientes. De esta forma, se logra dar redundancia en la autentificación.

Se genera un valor de "m", $m = g^x (H_1(A,B,\pi))^r$

Donde g dada con anterioridad, r es co-primo de q y π fue generada por la entidad KCC al momento de solicitar la instancia de conexión entre A y B. π Es un valor generado al azar. Esta llave simétrica π es transmitida al usuario B en la sesión de inicialización de usuarios móviles con la finalidad de separarla de los valores "g" y "m" que son transferidos en esta parte. Puesto

que estos valores son utilizados para generar la llave de "Password-Authentication," le será más difícil al adversario adivinar el password ya que tendría que obtener el valor de π, y relacionarlo con los valores "m" y "g" correspondientes para poder generar la llave de sesión entre A y B.

2. El usuario B genera la llave $\mu = g^{y'}$ y utiliza los valores m, π, y', r para la generación de σ el cual a su vez es utilizado para generar a "k". Luego envía el valor de μ, k al usuario "A".

3. Genera $\sigma = \mu^x$
Hace una comparación entre k y el valor obtenido por la función hash con los valores m,μ,σ,π para generar una nueva llave k' y transmitirla al usuario B. Así mismo utiliza otra función hash H_3 que será la llave de sesión si y solo si coincide $k' = H_{2b}(A, B, m, \mu, \sigma, \pi)$

4.- Si $k' = H_{2b}(A, B, m, \mu, \sigma, \pi)$ en el lado del usuario B coincide, entonces el protocolo realiza una función Hash H_3 para generar la llave K la cual es la llave de sesión que se utilizará entre A y B.

Ahora veamos los detalles del protocolo:

Precálculos

Inicialización de la red (AS_R) (para ambos) KCC

Estación base $m\alpha, n_p, \alpha, idd_R$ ← KCC escoge $m\alpha, n_p, idd_R, \alpha \in G$

Almacena: $m\alpha, n_p, \alpha, idd_R$

Inicialización Usuarios móviles (para ambos) KCC escoge id_A, id_B para A y B

Usuarios eligen R_A R_B aleatorias

$_{KCC}\{ UserA, m_1 \{h(m_1),KCC\}_{DSA_{kcc\text{-}priv}} \}Kpub_A$

$_{KCC}\{ UserB, m_2 \{h(m_2),KCC\}_{DSA_{kcc\text{-}priv}} \}Kpub_B$ ←

Donde : $m_1 = m\alpha, \alpha, id_A, \pi$ y $m_2 = m\alpha, \alpha, id_B, \pi$

Donde $\pi = \pi[id_A, id_B]$

Calcula $P_A = \alpha^{R_A} \bmod m\alpha$

Calcula $P_B = \alpha^{R_B} \bmod m\alpha$

P_A, P_B → Calcula $Cert_A, Cert_B$

$_{UserA}\{ DSA_{kcc\text{-}priv}\{Cert_A\}\}$

$_{UserB}\{ DSA_{kcc\text{-}priv}\{Cert_B\}\}$ ←

Uso Firma Digital para mandar los certificados.

Almacena: $R_A, P_A, m\alpha, \alpha, id_A, Cert_A$

$R_B, P_B, m\alpha, \alpha, id_B, Cert_B$

π *(llave simétrica para usuarios a conectarse A y B)*

Ejecución (formación de usuario)

Usuario Estación base KCC

Se forma un campo infoA:

$infoA = (id_A, A_S\alpha, *(idd_R), id_B, B_S\alpha, *(idd_R), TE, Cert_AA)$

Formación del mensaje "S"

$S = \{ \{Cert_AA\}, \{Ex, DSA_S \{S, info_x, D\text{-}H\ m_n\}\}\}$

id_A, TE, S →

$id_A, idd_R, TE, N_R, DSA_S (N_R, S, idd_R)$ →

KCC busca archivo de usuario, saca a llave de usuario, valida Si se cumple acepta en caso contrario rechaza; Hace todos los recálculos necesarios y compara con el Correspondiente mensaje que llega. Y emite un certificado

$Cert A'_s(KCC, A_s, AS_R, U)$ ←

Hace los correspondientes Cálculos de autentificación Compara con lo recibido del Usuario y procede a dar las Credenciales temporales

$Cert A_s(AS_R, K_s, U)$ ←

Generar llave

Selecciona h $\in Z^*p$
Y calcula $g = h^{(P_A - 1)/q_A} \bmod q_A$
Repite hasta g = 1

$p_A\ q_A$ ←

selecciona p_A q_A tal que:
q_A primo: $2^{159} < q_A < 2^{160}$
p_A primo: q/p-1
Obs: DSS indica que p_A es primo:
Tal que: $2^{159+64t} < q_A < 2^{160+64t}$
donde : 0 <= t <= 8

selecciona aleatoriamente
un entero "x" dentro del
intervalo $[1, q_A - 1]$
Calcula $y = g^x \bmod p_A$

(p_A, q_A, g, y) llaves públicas
x llave secreta

Guarda:
p_A, q_A, g, y →

almacena:
p_A, q_A, g, y

Generar Firma

Sea m el mensaje:
Usuario

Seleccionar k ε Z k ε[1, p-1], k aleatorio
Calcula: $r = (g^k \bmod p_A) \bmod q_A$

r →

Calcular $k^{-1} \bmod p$

Calcular $s = k^{-1} \{h(m)+X^r\} \bmod p$,
h es la función SHA-1 Si s = 0 ir a 1
Se firma el mensaje (Es el par (r, s))

Para la verificación de firmas (r, s), el Rx:

Obtiene una copia autentica de la llave pública de Tx (p,q, g, y)

Calcular $w = s^{-1} \bmod q$ y h(m)

w, m ←

Calcular $u_1 = h(m)\ w \bmod q$, $u_2 = rw \bmod q$

u_1, u_2 →

Calcular $v = (g^{u_1}\ y^{u_2} \bmod p) \bmod q$

v ←

Aceptar la firma cuando y solamente cuando v = r.

A(User) **B(User)**

$x \in_R Z_q$

Donde x es seleccionado en forma aleatoria. Esta "x" NO es la misma que se utiliza anteriormente. Es decir, No es la llave secreta.

m, g

$m = g^x (H_1(A,B,\pi))^r$ ⟶ Test $m \neq 0 \bmod p$

Donde r es co_primo de q

$y' \in_R Z_q$

y' es otro numero primo, No es la llave publica

$\mu = g^{y'}$

μ,k

$\sigma = (m/(H_1(A,B,\pi))^r)^{y'}$

$k = H_{2a}(A, B, m, \mu, \sigma, \pi)$

$\sigma = \mu^x$ ⟵

Test $k \stackrel{?}{=} H_{2a}(A, B, m, \mu, \sigma, \pi)$

$k' = H_{2b}(A, B, m, \mu, \sigma, \pi)$ k'

$\mathbf{K} = H_3(A, B, m, \mu, \sigma, \pi)$ ⟶

Test $k' \stackrel{?}{=} H_{2b}(A, B, m, \mu, \sigma, \pi)$

$\mathbf{K} = H_3(A, B, m, \mu, \sigma, \pi)$

(Donde **K** es la llave de sesión entre A y B)

Si "Connect" "Connect"

Entonces queda confirmado ⟵

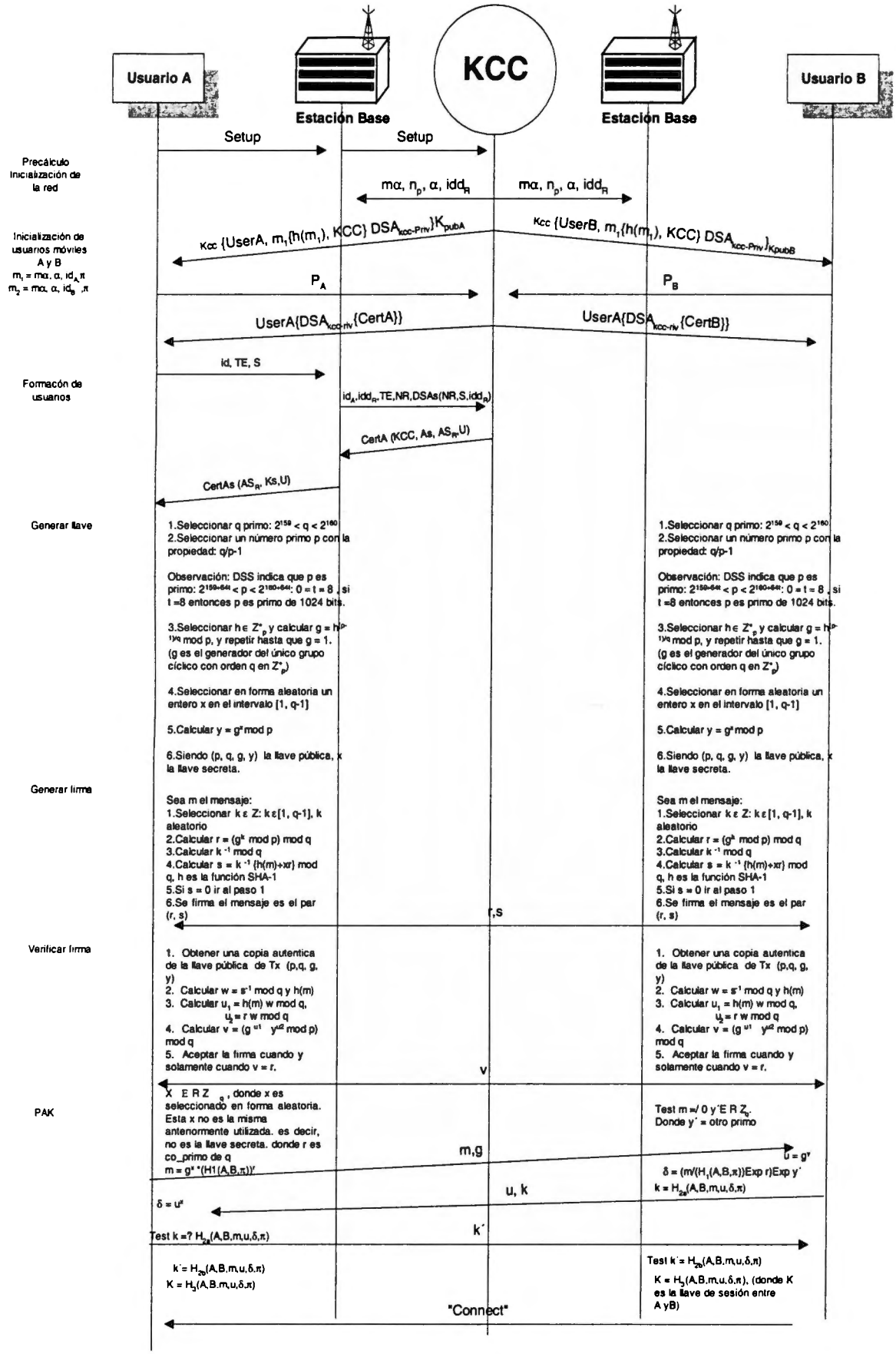
De otra forma, Aborta

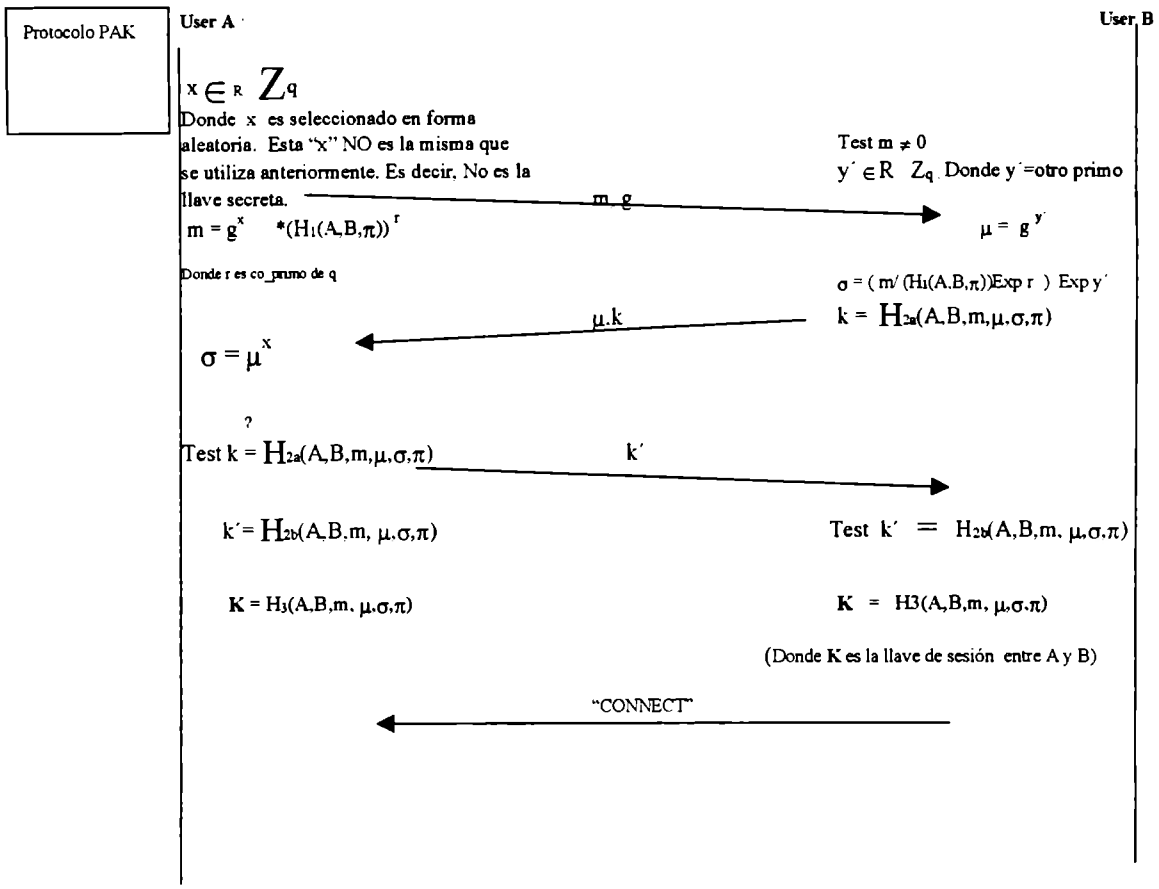
Una vez que toda la información ha sido enviada y verificada se manda un mensaje de confirmación "Connect"

El protocolo PAK con $\pi = \pi$ [A,B]. La llave de sesión resultante es K. Si la prueba regresa falso, entonces el protocolo aborta.

El protocolo PAK es un protocolo seguro de "Autentificación-Password" de intercambio de llaves en el modelo de autentificación explícita.

Se incorpora un protocolo ligero de intercambio de llaves (PAK) entre A Y B, para la autentificación de las partes interesadas (A y B.) (Autentificación por password con intercambio de llaves.)


Usuario A
Estación Base
KCC
Estación Base
Usuario B
Setup
Setup
Precálculo
Inicialización de
la red
mα, n_p, α, idd_R
mα, n_p, α, idd_R
Inicialización de
usuarios móviles
A y B
m1 = mα, α, idA, π
m2 = mα, α, idB, π
Kcc {UserA, m1{h(m1), KCC} DSA_Kcc-Priv}K_pubA
Kcc {UserB, m1{h(m1), KCC} DSA_Kcc-Priv}K_pubB
P_A
P_B
UserA{DSA_Kcc-Priv{CertA}}
UserA{DSA_Kcc-Priv{CertB}}
Formacón de
usuarios
id, TE, S
id_A,idd_R,TE,NR,DSAs(NR,S,idd_R)
CertA (KCC, As, AS_Pr,U)
CertAs (AS_R, Ks,U)
Generar llave
1.Seleccionar q primo: 2^159 < q < 2^160
2.Seleccionar un número primo p con la propiedad: q/p-1
Observación: DSS indica que p es primo: 2^159+64t < p < 2^160+64t: 0 = t = 8 , si t =8 entonces p es primo de 1024 bits.
3.Seleccionar h ε Z*_p y calcular g = h^(p-1)/q mod p, y repetir hasta que g = 1. (g es el generador del único grupo cíclico con orden q en Z*_p)
4.Seleccionar en forma aleatoria un entero x en el intervalo [1, q-1]
5.Calcular y = g^x mod p
6.Siendo (p, q, g, y) la llave pública, x la llave secreta.
Generar firma
Sea m el mensaje:
1.Seleccionar k ε Z: k ε [1, q-1], k aleatorio
2.Calcular r = (g^k mod p) mod q
3.Calcular k^-1 mod q
4.Calcular s = k^-1 {h(m)+xr} mod q, h es la función SHA-1
5.Si s = 0 ir al paso 1
6.Se firma el mensaje es el par (r, s)
r,s
Verificar firma
1. Obtener una copia autentica de la llave pública de Tx (p,q, g, y)
2. Calcular w = s^-1 mod q y h(m)
3. Calcular u1 = h(m) w mod q,
u2 = r w mod q
4. Calcular v = (g^u1 y^u2 mod p) mod q
5. Aceptar la firma cuando y solamente cuando v = r.
v
PAK
X E R Z_q, donde x es seleccionado en forma aleatoria. Esta x no es la misma antenormente utilizada. es decir, no es la llave secreta. donde r es co_primo de q
m = g^x *(H1(A,B,π))^r
Test m ≠ 0 y´E R Z_q.
Donde y´ = otro primo
m,g
u = g^y
δ = (m/(H1(A,B,π))Exp r)Exp y´
k = H2a(A,B,m,u,δ,π)
u, k
δ = u^x
k´
Test k =? H2a(A,B,m,u,δ,π)
k´= H2b(A,B,m,u,δ,π)
K = H1(A,B,m,u,δ,π)
Test k´= H2b(A,B,m,u,δ,π)
K = H1(A,B,m,u,δ,π), (donde K es la llave de sesión entre A yB)
"Connect"



Comentarios

En este protocolo se muestra como en condiciones de movilidad al incorporar el protocolo PAK éste provee una autentificación adicional al generado por la firma digital y entidades certificadoras, logrando una redundancia en el servicio de autentificación y fortaleciendo la seguridad del esquema asegurándose de la identidad de las partes involucradas.

El protocolo PAK recibe la dureza o robustez del problema DDH - Diffie_Hellman y de la aleatoriedad de los valores seleccionados. Romper DDH implica construir un "Time-polynomial adversary" que distingue Z = DH (X,Y) de la Z aleatoria con un "non-negiligible advantege" sobre la búsqueda aleatoria. Así como de las funciones hash implementadas para el intercambio de llaves para la autentificación.

Por otro lado, incorporamos unidades de encripción para las celdas ATM permitiendo privacidad en la comunicación y sin exponer información sensible en los nodos. Además, cumpliendo con los requisitos establecidos para la seguridad en ATM. Así mismo justificando las características por la cual hemos decidido adoptar ATM para el esquema PCN.

Otro de los puntos de oportunidad de aplicación donde pueden ser implementados los protocolos propuestos es en las aplicaciones de comercio-móvil, mismo, que se trata en la siguiente sección.

4.2 DISEÑO DE APLICACIONES DE COMERCIO-MÓVIL

El "E-commerce" esta cambiando el comercio y ha influenciado ampliamente y cambiado las mercancías, es decir, la manera en que se intercambian en términos del espacio y del tiempo. Las redes de área amplia, Internet por ejemplo, han hecho posible la creación de un mercado global en gran parte independiente de la localización física de usuarios y de las mercancías [6, 7, 8]. La movilidad ha agregado una nueva dimensión, el "mercado global". Los dispositivos móviles, los teléfonos portátiles notablemente, y las redes inalámbricas han hecho este mercado virtualmente accesible a millones de consumidores. Esto ha motivado la creación de un conjunto grande de servicios de valor agregado apuntados específicamente a los consumidores móviles. El mercado móvil tiene una potencialidad enorme, según lo publicado en varios informes [9], pero también plantean un gran número de diversos problemas con respecto a los ya tratados tradicionalmente en el mundo de Internet, basado en sistemas PC-alámbricas. En esta sección tratamos los problemas relacionados con el diseño y el desarrollo de las aplicaciones del comercio-móvil con particular énfasis en el uso de teléfonos celulares como los dispositivos móviles [10].

Introducción.

Las consideraciones presentadas en este apartado se basan en experiencias en el diseño y aplicaciones de comercio-móvil desarrolladas por [11,12]. Aunque el enfoque se ha desarrollado en el contexto específico de GSM y de infraestructuras de Internet, y en el uso de teléfonos celulares como dispositivos móviles, creemos que pone en evidencia un conjunto de problemas que se pueden considerar representativos de la clase entera de las aplicaciones comercio-móvil.

Marco de referencia para las aplicaciones del comercio-móvil.

Desde un punto de vista particular, el mundo que emerge de las aplicaciones del comercio-móvil necesita un conjunto de lineamientos para conducir su desarrollo. El punto clave para estas aplicaciones no es la complejidad de la lógica de aplicación en comparación con la calidad del servicio que la aplicación debe exhibir al ejecutarse. Bajo el requisito vago de la calidad del servicio se incluye deliberadamente una amplia gama de características que se extienden desde las capacidades de interacción del usuario hasta la seguridad, a partir del momento de la transacción al ancho de banda de la red. Así el diseño y el desarrollo de estas aplicaciones se deben hacer de una manera disciplinada y cuidadosa que sigue posiblemente una metodología bien definida y que adopta los mecanismos apropiados. Ya que finalmente la meta es proporcionar elementos suficientes a la definición de una metodología y mecanismos útiles, partiendo de la situación actual del contexto móvil, que parece no prestar casi ninguna atención al nivel de la aplicación en comparación con el nivel de la infraestructura.

Una primera observación relevante es que la configuración cliente/servidor común, utilizada extensamente en el "E-commerce" basado en Internet, podría no ser conveniente para un ambiente móvil. En el modelo cliente/servidor la mayoría de la lógica de aplicación es generalmente más exigida al cliente, lo cual en comercio-móvil es apoyado por un pequeño dispositivo limitado, tal como un teléfono celular, y que no puede contar en una conexión de red confiable con la parte de la transferencia de sus cómputos en el servidor. Esto sucede debido a la

diversa naturaleza de las redes alámbricas e inalámbricas. El último es, de hecho, intrínsecamente no fiable, inestable e imprevisible.

Por lo tanto, se crea la necesidad de nuevos modelos arquitectónicos desarrollados explícitamente para el comercio-móvil [13]. La nueva configuración debe ser altamente escalable para utilizar la amplia variedad de dispositivos móviles actuales en el mercado y de los que aparecerán en el futuro. Por otra parte, la configuración tiene que considerar la evolución de las infraestructuras y de las tecnologías de red.

En primer lugar en esta dirección, estudiaremos las aplicaciones actuales del "E-commerce" para contornear los que puedan ser convenientes para el marco del comercio-móvil.

Tipología de las aplicaciones del comercio electrónico.

Centrándonos en las relaciones de Negocios-a-Consumidor (Business-to-Consumer (B2C)), en donde un individuo interactúa a distancia con el sistema a través de una red de telecomunicación, podemos clasificar las aplicaciones referentes a este campo, dependiendo de la cantidad de tiempo que intervenga entre la petición del comprador y de la aceptación del contrato de venta del vendedor y del comprador. El tiempo necesario para elegir lo más conveniente a comprar puede desempeñar un papel en la clasificación siguiente:

- Servicios de Tiempo Cortos (STS). La oferta del servicio es directa, es decir no hay mediación; el objeto de las mercancías de la transacción se estandariza altamente, esto es que ellos son de uso común y exhibe un valor totalmente definido.
  En este escenario el comprador desea reducir al mínimo la interacción con el sistema que es el tiempo de las negociaciones necesarias para comprar mercancías. Un ejemplo es representado por los servicios que ofrecen posibilidad para comprar boletos o para recargar tarjetas de teléfono pre-pagadas. Por otra parte, las maneras sofisticadas de presentar las mercancías (ejemplo, presentaciones multimedia) no son necesarias.

- Servicio de Tiempo Medio (MTS). Es posible asociar esta categoría al modelo de una plaza comercial, donde el comprador tiene una oferta grande de mercancías, puede aprovecharse de ofertas especiales y tiene opciones múltiples para la personalización del servicio tal como fecha de salida, localización de salida, sistema de pago, etc.; la opción de mercancías y aceptación de contrato de venta ocurrido en la misma sesión de conexión

- Servicios de Tiempo Largo (LTS). Pertenecer a esta categoría de aplicación como servicios de subasta, donde la negociación implica varios compradores quienes compiten por un limitado recurso que se asigna después de un largo tiempo (hora, día, mes); el protocolo de comunicación entre las partes debe utilizar las transacciones asíncronas [14].

Las actuales tecnologías, notablemente las redes de acceso y las características de los teléfonos celulares, obligan fuertemente al conjunto de servicios a que se puedan adaptar al ambiente móvil. Refiriéndonos a la anterior clasificación de las tipologías de aplicación del "E-commerce" el STS y el MTS parecen ser las más convenientes de ser utilizadas con la actual tecnología. La base de todas estas aplicaciones es siempre una transacción entre el usuario y un sistema alejado. Así, el segundo paso en nuestro estudio debe ser una clasificación de la naturaleza de las

transacciones en un ambiente móvil, para definir más adelante el dominio de la aplicación del comercio-móvil.

Tipología de transacciones de comercio-móvil.

En el escenario móvil, el proceso de la transacción representa el aspecto más relevante de todas las aplicaciones, así una clasificación de las tipologías de la transacción del comercio-móvil puede ayudarnos en la definición de un marco arquitectónico. La clasificación se basa en modalidades del acceso al servicio y en la clase de dispositivos que puedan estar implicados. Desde el punto de vista de las modalidades de acceso, los servicios del comercio-móvil se pueden caracterizar como suscritos o no-suscritos.

Los servicios suscritos, la mayoría de éstos hoy en día son ofrecidos por los mismos proveedores de servicio, son a menudo personalizados para el usuario específico y tienen un nivel de seguridad más fuerte [15][16]. La desventaja más grande de este acercamiento es que la activación de la suscripción puede ser un proceso complejo, que no se puede lograr a menudo a través del mismo dispositivo usado para la transacción, es decir puede requerir una conexión alámbrica, o una interacción directa con el proveedor del servicio. Por otra parte, los usuarios tienen aversión a esta clase de servicios debido a la naturaleza de la suscripción, que implica una cierta clase de contrato permanente con el proveedor, la cual también recoge una cantidad valiosa de datos sensibles sobre ellos.

Los servicios no-suscritos, debido a su naturaleza de tiempo-limitado, necesitan siempre *una* interacción más compleja entre el usuario y el sistema, lo cual implica un tiempo más largo de tener acceso al servicio, y la hacen menos confiable y expuesta a los problemas de la red. Por otra parte, muchos de los servicios más importantes del comercio-móvil requieren de un alto nivel de seguridad de autentificación, que no es satisfecha por este esquema de transacción. Y es aquí donde nuestras propuestas de protocolos encuentran una oportunidad de aplicación. Sin embargo, la experiencia muestra que esta clase de servicio es a menudo aceptado por la mayoría de los usuarios.

Los dispositivos involucrados en transacciones de comercio-móvil se pueden clasificar sobre la base de sus capacidades de poder computacional y de la capacidad de entrada (desde los teclados numéricos de teléfonos portátiles hasta pantallas al tacto y reconocimiento de la escritura ofrecido por PDAs modernas.) El acceso y la compra de un servicio implican a menudo una transacción ligera, así que estos pueden ser logrados fácilmente incluso a través de un dispositivo limitado como un teléfono celular. La verificación, es decir el control del título de la compra (y posiblemente de su transformación en un título más común, por ejemplo, un boleto de papel) requiere siempre una conexión alámbrica. Si la conexión es demasiado difícil de obtener, una cantidad de espacio grande para salvar la información y para realizar una verificación retrasada es necesaria. Esto sugiere que estas dos fases de un servicio inalámbricas se pueden lograr utilizando dispositivos distintos:

- El cliente compra el servicio utilizando un equipo de funciones mínimas, que a menudo puede ser el teléfono celular que ya se posee. El usuario recibe del sistema un pequeño "token", que se puede salvar generalmente en la memoria del dispositivo, como un equivalente electrónico del título de la compra.

La verificación del servicio se puede lograr directamente por el vendedor del servicio, o por terceros. En ambos casos, la validación requiere una verificación del título de compra. Esto implica una transacción entre el dispositivo del usuario y el sistema del vendedor que libera el título. Si una conexión está disponible, el proveedor puede interconectar directamente con el dispositivo del usuario, el cual salva el título, y lo verifica. Si no, el dispositivo del proveedor debe salvar el token y la identidad del cliente, y se retrasa el control. Ambas de estas opciones requieren un hardware dedicado, que sin embargo se carga solamente en el vendedor.

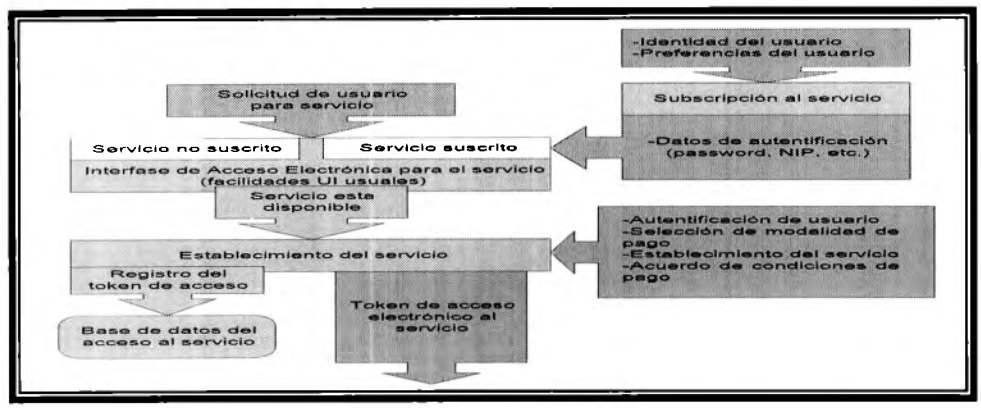


Fig 4.3. Configuración lógica de la suscripción del servicio y de la compra.

Una configuración lógica de referencia.

En los cuadros siguientes se bosqueja una configuración lógica de referencia que se puede proponer para utilizar las aplicaciones del STS y posiblemente de LTS basadas en el análisis anterior. Los diagramas están partidos en dos por razones de espacio. En la figura 4.3 vemos los componentes principales de la configuración con respecto al registro y a la compra del servicio. La figura 4.4 ilustra los componentes que hacen frente a la conversión entre el token de servicio virtual y el token de acceso al servicio actual (por ejemplo, un boleto real) y la verificación de su validez. Así podemos seleccionar 4 subsistemas principales.

1) La interfase de acceso al servicio es una aplicación amigable, típicamente un conjunto de Web o de las paginaciones de WAP, donde el usuario puede operar recíprocamente con el proveedor de servicio. El usuario puede primero examinar el servicio(s) ofrecido, solicitando posiblemente una cantidad limitada de información adicional al vendedor, entonces define los detalles del servicio y controla la disponibilidad del producto deseado. Estos pasos se pueden realizar sin ninguna identificación del usuario, puesto que son solamente "informativos". Sin embargo, algunos servicios pueden requerir una suscripción previamente establecida que permita que el usuario tenga acceso a la interfaz del servicio. Este registro es útil para presentar un perfil del usuario preliminarmente (de modo que la interacción durante el acceso al servicio sea limitada) y/o para limitar el acceso de los usuarios que no están realmente interesados en el servicio (es decir, que desean solamente "dar una mirada", pero al hacer esto consumen recursos del sistema).

2) Una vez que el usuario sea satisfecho por el servicio propuesto, y este último está disponible con las características solicitadas, el establecimiento del servicio es responsable del establecimiento del contrato de servicio desde la autentificación al pago. El usuario debe ahora identificarse por si mismo y el proveedor del servicio debe

controlar la identidad dada. Por otra parte, el proveedor de servicio debe también asegurar su identidad y confiabilidad. Esto se puede lograr de dos maneras: los proveedores bien conocidos podrían garantizar por sí mismos, mientras que la mayoría de los otros deben utilizar terceros legalmente confiables para realizar la identificación de todos los temas implicados en la transacción. Entonces el usuario es presentado con un contrato en línea que se firmará, y eventualmente otros pasos requeridos por la modalidad de pago particular son tomados. La salida de este subsistema es un "token de acceso electrónico" (por ejemplo un NIP) que se puede salvar generalmente en el dispositivo utilizado para tener acceso al servicio (por ejemplo, la memoria de la tarjeta SIM de un teléfono celular GSM.) Este token será utilizado más adelante para tener acceso al servicio, y también se coloca en la base de datos del servicio para el paso de validación.

3) El acceso al servicio puede proporcionar varias modalidades de verificación y acceso. Cuando el usuario acceda al servicio, el o ella pueden ser requeridos para cambiar el token de acceso electrónico en un título de acceso estándar (por ejemplo, un boleto), o utilice directamente el título electrónico.

4) Finalmente, los controles de la verificación del título de acceso si un usuario ha adquirido el título correcto para utilizar el servicio. La verificación se puede realizar directamente durante el acceso al servicio, o más adelante mientras que el servicio está en uso. En ambos casos, puede ser inmediata o retrasada. Cuando se retrasa la verificación, la identidad del usuario y el título de acceso se pueden salvar en un dispositivo de almacenamiento temporal y se revisan más adelante en la base de datos del servicio. Si utiliza un título de acceso electrónico para realizar la validación, el usuario puede interconectar directamente su dispositivo portátil con el dispositivo del inspector (por ejemplo, un teléfono celular puede interconectarse con el dispositivo de control vía un acceso infrarrojo o su señal de radio estándar).
Esta configuración parece absolutamente general y permite la caracterización de una clase de servicios grande. Un problema muy interesante es cómo podemos asociar esta configuración en la infraestructura real de la red, esto es, qué componentes lógicos (subcomponentes) residen en qué componentes de la infraestructura. Esta asociación no es directa y permite diversos grados de lo que percibirá el usuario final como calidad del servicio.

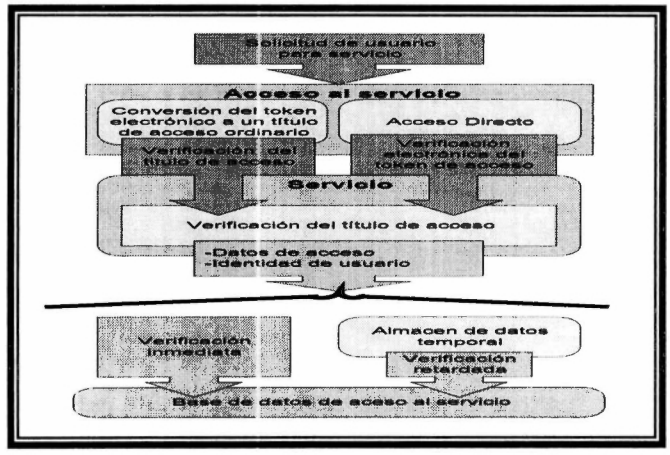


Fig. 4.4 Configuración lógica de la verificación y acceso al servicio.

Comentarios

En esta sección damos una clasificación de las transacciones del comercio-móvil basadas en modalidades de acceso y de los dispositivos implicados, y un bosquejo qué consideramos como los componentes arquitectónicos básicos de las aplicaciones del comercio-móvil de STS. Tales componentes son todos requeridos para estar presentes en una aplicación final y se deben diseñar cuidadosamente para asegurar una calidad aceptable del servicio, que es obligado fuertemente por las limitaciones de la tecnología.

Una metodología de diseño apropiada por lo tanto es necesaria, así como un conjunto de herramientas refinadas y de prueba. De hecho, está claro que esta clase de aplicaciones deben ser altamente modulares, y por lo tanto su implementación se parte naturalmente en varios pasos, cuyas salidas se pueden probar por separado.

Por otra parte, muchos componentes de esta configuración deben ser reutilizables en aplicaciones similares. En un mundo rápidamente creciente (y cambiante) como el comercio-móvil, la reutilización de componentes pre-construidos permitirá un desarrollo rápido de aplicaciones, acortará su tiempo de poner a punto, y asegurará niveles más altos de seguridad y de calidad.

4.3 WAP

WAP es el Protocolo de Aplicaciones Inalámbricas, mismo que surge debido a la combinación de dos tecnologías de amplio crecimiento y difusión durante la última década: Las Comunicaciones Inalámbricas e Internet.

WAP fue realizado por 4 compañías (Nokia, Motorola, Ericsson y Unwired Planet), la Fig.4.5 nos muestra el modelo de funcionamiento del sistema WAP, el cual se encuentra definido en [17]. Dicho protocolo proporciona todos los servicios (Navegación, Correo Electrónico, Comercio Electrónico, etc.) que tiene disponible el usuario con Internet, entre otros propios de los servicios de voz.

Para ello, se parte de una arquitectura basada en la arquitectura definida para el World Wide Web (WWW), pero adaptada a los nuevos requisitos del sistema.

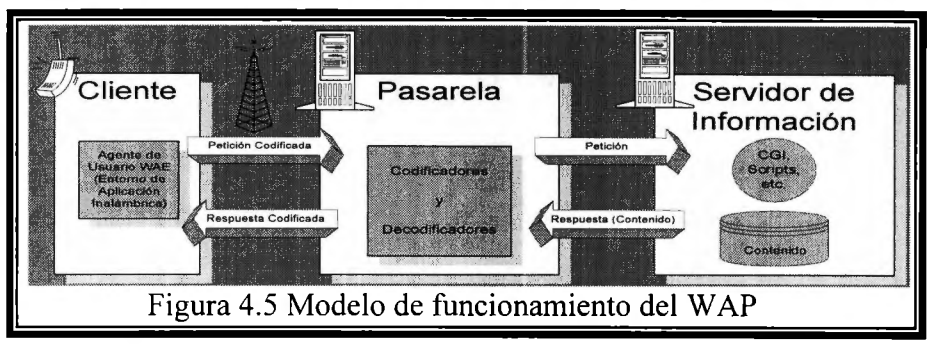


Figura 4.5 Modelo de funcionamiento del WAP

De esta forma, en el terminal inalámbrico existe un *"micro navegador*[29]*"* encargado de la coordinación con la pasarela, a la cual le realiza peticiones de información que son adecuadamente tratadas y redirigidas al servidor de información adecuado. Una vez procesada la petición de información en el servidor, se envía esta información a la pasarela que de nuevo procesa adecuadamente para enviarlo al terminal inalámbrico.

Para conseguir consistencia en la comunicación entre el terminal móvil y los servidores de red que proporcionan la información, WAP define un conjunto de componentes estándar:

- Un modelo de nombres estándar. Se utilizan las URIs (*Universal/Uniform Resource Identifier* ó Identificador Uniforme/Universal de Recurso) definidas en WWW para identificar los recursos locales del dispositivo (tales como funciones de control de llamada) y las URLs (también definidas en el WWW) para identificar el contenido WAP en los servidores de información.
- Un formato de contenido estándar, basado en la tecnología WWW.
- Unos protocolos de comunicación estándares, que permitan la comunicación del *micro navegador* del terminal móvil con el servidor Web en red.

Veamos ahora un modelo global de funcionamiento de este sistema en la Figura 4.6.

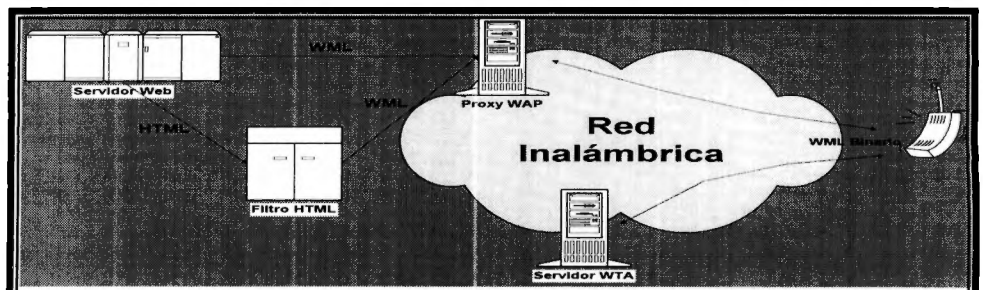


Figura 4.6. Ejemplo de una red WAP

En el ejemplo de la figura, el terminal móvil tiene dos posibilidades de conexión: a un proxy WAP, o a un servidor WTA. El primero de ellos, el proxy WAP traduce las peticiones WAP a peticiones Web, de forma que el cliente WAP (el terminal inalámbrico) pueda realizar peticiones de información al servidor Web. Adicionalmente, este proxy codifica las respuestas del servidor Web en un formato binario compacto, que es interpretable por el cliente. Por otra parte, el segundo de ellos, el Servidor WTA[30] está pensado para proporcionar acceso WAP a las facilidades proporcionadas por la infraestructura de telecomunicaciones del proveedor de conexiones de red.

4.3.1 COMPONENTES DE LA ARQUITECTURA WAP

El modelo WAP esta basado en la arquitectura definida en el World Wide Web (WWW) adaptándolo a los nuevos requisitos del sistema haciendo uso de una pila de protocolos similares a los empleados en Internet, basándose en un modelo de capas al igual que el sistema OSI, cada

[29] Se pretende que este *micro navegador* actúe de interfaz con el usuario de la misma forma que lo hacen los navegadores estándar.

[30] *Wireless Telephony Application* ó Aplicación de Telefonía Inalámbrica

una de estas capas del modelo de referencia emplea uno o varios protocolos los cuales tienen la función de interpretar la información que recibe de la capa inmediata inferior y adaptarla para que la capa inmediata superior pueda repetir la misma operación y llegar a la capa de aplicación. En la Fig. 4.7 se muestra la forma en que se encuentra distribuida la arquitectura antes mencionada [17], por lo que el terminal móvil hará uso de un "pequeño navegador" similar a Netscape Navigator o Internet Explorer encargado de la coordinación con la pasarela, a la que realiza peticiones de información; peticiones que son tratadas y encaminadas al servidor de información adecuado. Una vez procesada en el servidor, la información se envía a la pasarela, que la procesa y la envía al teléfono móvil.

Una vez introducido el sistema, vamos a ver la arquitectura que le da consistencia. La arquitectura WAP está pensada para proporcionar un "entorno escalable y extensible para el desarrollo de aplicaciones para dispositivos de comunicación móvil". Para ello, se define una estructura en capas, en la cual cada capa es accesible por la capa superior así como por otros servicios y aplicaciones a través de un conjunto de interfaces muy bien definidos y especificados. Este esquema de capas de la arquitectura WAP la podemos ver en la Figura 4.7.

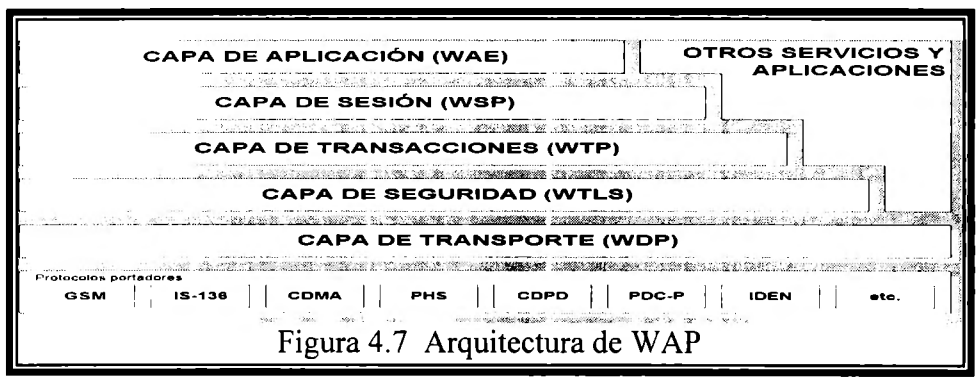


Figura 4.7 Arquitectura de WAP

- La capa de transporte viene definida por el protocolo WDP (Wireless Datagram Protocol), este permite hacer uso de las mismas aplicaciones en diferentes tipos de portadoras (distintas frecuencias o distintos protocolos de acceso al medio) o señales de información.
- En la capa de seguridad se emplea el protocolo WTLS (Wireless Transport Layer Security) el cual es derivado del SSL 3.1, es basado en el sistema abierto TLS 1.0 proporcionando los elementos de seguridad de confidencialidad, integridad y autentificación; la verificación de la autentificación, no-repudio son dadas por una PKI.
- La capa de transacción esta basada en WTP (Wireless Transaction Protocol) derivado del TCP. la función principal de esta capa es eliminar los datagramas no utilizados y preparar la información para la capa superior.
- WSP (Wireless Session Protocol), es el protocolo que se empleará en la capa de sesión y está preparado para agrupar varias operaciones WTP siendo encargado también del restablecimiento de las conexiones que excedan el tiempo de vida asignado al iniciar la conexión.
- La última capa, la de aplicación define la interfaz de usuario en el teléfono, hace uso de: Wireless Markup Language (WML), WMLScript y Wireless Telephony Application (WTA).

Al igual que UMTS los mecanismos de seguridad de WAP se encuentran en una etapa de desarrollo aunque ya existen algunas herramientas que se apoyan en dicho estándar para ofrecer los elementos de confidencialidad, integridad, autenticidad y no-repudio. Así, se

cuenta con W/Secure [18] y Baltimore Telepathy [19] los cuales contienen una implementación de WTLS, existen diferentes formas de implementar dichos mecanismos de seguridad, entre los cuales tenemos:

- Autentificación mutua sobre la interfaz aire, la cual serviría para establecer parámetros importantes de seguridad.
- Cifrado interfaz aire, para emplear este tipo de cifrado es necesario hacer uso de diferentes claves de control junto con la información de señalización.
- Cifrado punto a punto, por medio de este tipo de cifrado la aplicación puede verificar las claves de administración sin ningún problema y de esta manera los datos que hace uso la aplicación nunca serán expuesta fuera de ella.

4.3.2 MECANISMOS DE SEGURIDAD

WAP ofrece una arquitectura flexible de seguridad, centrándose en proporcionar seguridad entre la conexión que posee un usuario y un servidor WAP, es decir, en general no ofrece mecanismos de seguridad extremo a extremo entre el usuario del terminal móvil y el servidor Web de Internet. Sin embargo, muchas aplicaciones requieren servicios de seguridad extremo a extremo (en particular es especialmente crítica la autentificación entre clientes y servidores Web). A continuación se evalúan diferentes opciones para ofrecer estos servicios extremo a extremo, estudiando el compromiso entre el nivel de seguridad requerido y la complejidad y costo de la solución adoptada. Se consideran las siguientes opciones:

- Confiar en el servidor WAP y utilizar el mecanismo de autenticación de la red móvil: En este caso se cede toda la autenticación del cliente a la propia red móvil, y el servidor WAP establece una conexión SSL con el servidor Web. Esta solución requiere confianza total en el servidor WAP, pero es fácilmente implantable y en la red móvil no es necesario utilizar el protocolo WTLS.
- Confiar en el servidor WAP y utilizar WTLS entre cliente y servidor WAP: Se aumenta la seguridad en la red móvil, pero nuevamente es necesaria una confianza total en el servidor. Requiere que los terminales móviles y el servidor WAP implementen WTLS.
- Utilizar una conexión WTLS con el servidor Web remoto: Esta solución no requiere confianza en el servidor WAP (las medidas de seguridad se implementan extremo a extremo). A cambio, requiere que el servidor de Internet ofrezca un servidor WTLS.
- Proteger la comunicación al nivel de aplicación: Ciertas aplicaciones críticas requerirán servicios especiales de seguridad (como no repudio) que forzosamente se deben ofrecer al nivel de aplicación.

4.4 LAN INALÁMBRICA (Wireless Local Area Netwok WLAN)

Redes de área local inalámbricas son generalmente operadas por iniciativas privadas ver figura 4.8, proveen generalmente de una tasa alta de comunicación de datos sobre una área pequeña. Con la excepción de algunos sistemas (ejemplo, Altair [20].) Las WLANs típicamente operan en Bandas ISM (excepto para sistemas infrarrojos.). Estas pueden ser preferidas con respecto a sus contrapartes alámbricas debido a que las redes alambradas son impracticas en el grado de necesidad de movilidad

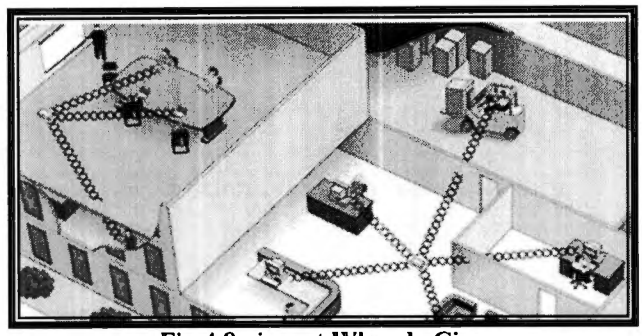


Fig.4.8 aironet Wlan de Cisco

Las tecnologías existentes para WLANs son sistemas de celular con licencia, operando en los 18-19 MHz. [21] de frecuencias de radio, Sistemas de espectro-expandido no licenciados operando en las bandas de ISM [22] tal como FreePort y Wave LAN, sistemas de infrarrojo difundido (DF/IR), y sistemas de infrarrojo "directed-beam" (DB-IR) [23]. La siguiente tabla muestra algunos sistemas disponibles en el mercado.

Tabla 4.1

Técnica	Óptica		RF		
	DF/IR	DB/IR	RF	DSSS	PHSS
Data rate (Mb/s)	1-4	10	5-10	2-20	1-3
Movilidad	Estacionario / móvil	Estacionario con LOS	Estacionario / móvil		Móvil
Alcance (ft)	50-200	80	40-130	100-800	100-300
Detectabilidad	Negligible		Algunos	Poco	Poco
Longitud de Onda /Frecuencia	λ = 800 900nm		18 GHz o ISM	15M bands	
Técnica de modulación	OOK		FS/QPSK	QPSK	GFSK
Poder de radiación			25 mW	<1W	
Método de acceso	CSMA	Token Ring, CSMA	Reservación ALOHA CSMA	CSMA	

4.4.1 OPCIONES TECNOLÓGICAS

4.4.1.1 Cuando se diseña una WLAN se podrían tener varias opciones tecnológicas a escoger, cada una de estas ofrecen ventajas y limitaciones. A continuación se da una breve descripción de cada una de ellas:

- Tecnología de Banda Angosta Un sistema de radio de banda angosta transmite y recibe información de los usuarios en una frecuencia de radio especifica. La señal de radio de banda angosta trabaja bajo una señal de frecuencia cuyo ancho esta limitado por la capacidad que debe tener para dejar pasar la información. Una línea telefónica privada es mucho más segura que un sistema de radio frecuencia, cuando se realiza una llamada en una casa, esta no puede ser escuchada o interceptada tan fácilmente como en los sistemas de radio. Hoy en día se busca crear sistemas mucho más seguros que garanticen el empleo del sistema de radio.

- Espectro expandido (Spread Spectrum) En la actualidad muchos sistemas WLANs están usando la tecnología del Spread Spectrum, la cual es una técnica desarrollada por los militares que trabaja en la frecuencia de radio de banda ancha, para ser utilizada en aquellos sistemas de comunicaciones considerados críticos. Este sistema es empleado para utilizar más eficientemente y con mayor seguridad el ancho de banda. Existen dos tipos de radio spread spectrum: frecuencia Hopping (HFSS) y secuencia directa.

DSSS - Direct Sequence Spread Spectrum & FHSS - Frequency Hopping Spread Spectrum

FHSS y DSSS son técnicas de modulación desarrolladas en los años 40's. Dispersan la señal de transmisión sobre una banda amplia de radio frecuencias. Está técnica es ideal para comunicaciones de datos ya que es menos susceptible al ruido y pequeñas interferencias.

FHSS está limitado a 2Mbps y se recomienda en aplicaciones muy específicas. Para las demás redes, DSSS es una mejor opción ya que soporta velocidades de 11Mbps.

La HFSS emplea una portadora de banda angosta que cambia la frecuencia en un patrón conocido para la transmisión y la recepción. Sincronizado correctamente, el efecto final es el de mantener un canal lógico individual. En el HFSS pueden aparecer impulsos de ruido de corta duración. (ver figura 4.9)

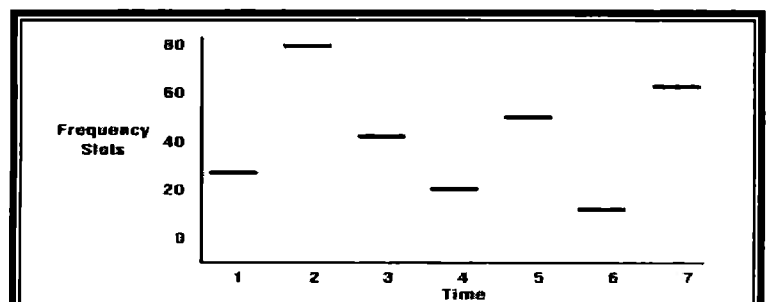


Fig. 4.9 HFSS

La DSSS genera un bit de redundancia por cada bit a ser transmitido. Este patrón es llamado Chip. El chip más largo aumenta la probabilidad de que los datos originales puedan ser recuperados. Si uno o más chip se dañan durante la transmisión se pueden recuperar los datos originales mediante técnicas estadísticas implantadas en la radio sin que sea necesaria la retransmisión de la señal, ver figura 4.10.

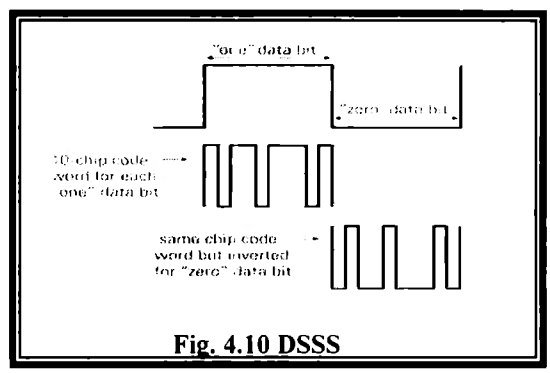


Fig. 4.10 DSSS

4.4.1.2. Técnica Infrarroja (IR)

El sistema Infrarrojo (IR) trabaja a frecuencias muy altas justo por debajo de la luz visible del espectro electromagnético, para portadora de datos. Con este sistema no se pueden penetrar objetos opacos, pudiendo ser esta una tecnología dirigida o difusa.

Los sistemas dirigidos económicos proveen un rango muy limitado y son utilizados en aplicaciones especificas de WLANs. Un alto desempeño del IR dirigido no es práctico para usuarios móviles y es usado solamente para implementar subredes fijas. El sistema difuso (o reflexivo) IR WLAN no requiere línea de vista, pero las celdas son limitadas a sectores individuales.

4.4.2 APLICACIONES DE LAS WLANS

- Para acceder a la red de forma más rápida permitiendo aumentar la productividad de consultores, auditores, pequeños grupos de trabajo, sitios de entrenamiento, así como permite aumentar el nivel de usuarios simultáneos en universidades por ejemplo.
- Para minimizar el overhead de los cambios en los movimientos en las WLANs.
- Para disminuir los costos asociados a la instalación de una infraestructura en edificios viejos.
- Para simplificar las frecuentes reconfiguraciones de la red.
- Para poder intercambiar información en tiempo real con la base de datos central.
- Para proveer un backup en aplicaciones criticas que corren sobre redes alámbricas.

4.4.2.1. Beneficios de las WLANs

Las WLANS ofrecen como se mostrará a continuación productividad, conveniencia y ventajas en cuanto a costos sobre las tradicionales redes alámbricas:

- Los sistemas móviles WLANs pueden proveer a los usuarios de la LAN acceso a la información en tiempo real en cualquier momento. Esta movilidad permite dar servicios que no son posibles con redes alámbricas.
- La velocidad y simplicidad en la instalación puede eliminar la necesidad de colocar cable a través de paredes y techos.
- La flexibilidad en la instalación permite a la red llegar donde el sistema de cableado no puede.
- Reduce los costos de infraestructura, la inversión inicial requerida para el hardware de la WLAN puede ser alta pero los costos totales de instalación así como los de ciclo de vida pueden ser significativamente más bajos que los generados por un sistema de cableado.
- El sistema WLAN ofrece escalabilidad y puede ser configurado en una gran variedad de topologías. Las configuraciones pueden ser cambiadas de forma sencilla.

La tecnología (WLAN) complementa tecnologías de acceso para la Tercera Generación de las redes celulares. Los estándares WLAN, HIPERLAN/2 e IEEE 802.11a, permiten incluso rangos de datos más altos que UMTS (por arriba de los 54 Mbps) para cubrir en áreas conflictivas y en la ciudad.

4.4.3 REDES INALÁMBRICAS DE BANDA ANCHA (WANS INALÁMBRICAS.)

El espectro de banda ancha inalámbrico esta comenzando a trabajar bajo licencia en las bandas de frecuencias desde 2 GHz a 42 GHz. En Canadá la tecnología de redes inalámbricas es conocida como Sistema de Comunicaciones Multipunto Local (LMCS). En Estados Unidos y el resto del mundo es referida al Servicio de Distribución Multipunto Local (LMDS).

El acceso inalámbrico en banda ancha es una opción atractiva para establecer servicios porque permite extender su área de cobertura en la forma más económica y provee competitividad añadiendo valor a los servicios ofrecidos sobre el aire.

A la nueva generación les permite eliminar la necesidad de instalar costosas redes de fibra.

Los operadores de Redes de Área Amplia Inalámbrica (WWANs) utilizan espectro permitido y proporcionan servicios de datos de baja-tasa para sus clientes.

La siguiente tabla muestra algunos de los servicios de WAN

Tabla 4.2

País	MOBITEX	RD-LAP	CDPD
Estados Unidos de N.A	RAM Mobile Data	ARDIS	Operadores de Celular
Canada	Rogers Cantel	Bell-ARDIS	
U.K.	RAM Mobile Data	Licencia sin uso	
Francia	France Telecom. Y TDR		
Alemania	GfD	German Bundespost	
Suecia	Telia Mobitel		
Finlandia	Telecom. Finland		
Noruega	Tele-Mobile		
Bélgica	BellSouth Mobile Data		
Netherlands	RAM Mobile Data		
Suiza		Licencia en espera	
Australia	BellSouth		
Chile	CTC Cellular		
Tailandia		Operador no conocido	
Malasia		Operador no conocido	
México	En proceso de regulación		
Singapur	ST Mobile Data	Singapur Telecom	
Hong Kong		Hutchinson	
Total	14 operadores en 13 países		

La arquitectura Mobitex fue originalmente desarrollada por Telia, el operador nacional Sueco. Para animar el desarrollo de fuentes de equipamiento múltiple, el software y hardware de Mobitex esta disponible con ninguna licencia u honorario. Las redes Mobitex están operando en 10 países además de Estados Unidos [26]. Existe otra rama de WANs que no tiene sus propias redes especializadas, pero prefieren utilizar marcha lenta en los existentes canales analógicos AMPS para transmitir paquetes de datos a una tasa de 19.2 Kb/s. Este sistema es el CDPD

El estándar celular digital IS-54 e IS-95 soportan eventualmente un arreglo de servicios de datos incluyendo tanto modo de circuitos como modo de paquetes. En Europa, el ETSI comenzó a desarrollar un estándar publico para radio "trunked" y sistemas móviles. Este estándar es conocido como "Trans-European Trunked Radio standard" (TETRA).

4.4.3.1 Elementos esenciales para el establecimiento de la tecnología inalámbrica en banda ancha

La distribución de las señales a través de la tecnología inalámbrica a subscriptores remotos requiere de tres elementos esenciales:

- Una estación base. La estación base, es el sitio central que reúne todo el tráfico proveniente de los subscriptores de una celda dada. La estación base incluye equipamiento en la parte interna y al aire libre.

  El equipo de la parte interna provee una interfaz al "backbone" inalámbrico o cableado. El equipo al aire libre consiste de un transmisor y receptor usualmente localizados sobre una torre o en lo alto de una montaña. Estos entregan y reúnen todo el tráfico proveniente de los subscriptores dentro de un sector o celda.

- Un equipo local para el cliente. El equipo local del cliente refleja funcionalmente el equipo de la estación base, en el sitio local del cliente el transmisor, el receptor y la antena se encuentran generalmente protegidos en una unidad compacta que es altamente direccional. La modulación, demodulación y funciones de interfaz en edificios cableados ocupan lugar en una unidad de interfaz de red (NIU.)
  Las NIUs están diseñadas para direccionar un rango de subscriptores mediante tarjetas y cuyos requerimientos de conectividad pueden incluir T1/E1, POST y vídeo digital.

- Un sistema de administración de red. El sistema de administración de red (NMS) controla los componentes cableados e inalámbricos, así como los servicios que son entregados. Idealmente el NMS provee funcionalidad punto a punto a través de la red, incluyendo el backbone y a los clientes locales.

4.4.4 FUNCIONALIDAD EN ATM

La combinación de la tecnología inalámbrica en banda ancha con redes de multiservicios integrada permite proveer mayores servicios. Sobre ATM puede llegar a consolidar múltiples servicios, capacitando un ancho de banda dinámico y garantizando una alta calidad de servicio (QoS) a los usuarios finales. Otras aplicaciones incluyen redes privadas virtuales, encriptación, teleconferencias, voz sobre IP y aprendizaje a distancia, así como una calidad de servicio en la facturación y un soporte en el mantenimiento.

4.5 HIPERLAN Y HIPERLAN/2

HIPERLAN es el estándar de la Comunidad Europea para LANs inalámbricas. Fue diseñado por un comité establecido por el Instituto de eStándares de Telecomunicación Europeo (ETSI). El estándar diseñado estuvo dirigido para estar tan cercano como fuera posible en funcionamiento al LAN-alámbrico, tal como Ethernet. Un bit rate de 23,529 Mbps fue proporcionado por canal; y los 5 canales existentes ocupan 150 Mhz del ancho de banda en el rango de 5.15-5.30 GHz o el rango 17.1-17.2 de GHz.

El estándar no distingue entre el modo de infraestructura y el modo "ad-hoc" de operación. En su lugar asume que algunos de los nodos pueden convertirse en nodos "expedición" y actuar como nodos intermedios en las rutas del paquete; esto significa que las terminales tienen que guardar y poner al día las tablas de encaminamiento para una red que constantemente cambia. El estándar especifica algunos métodos para tratar esta emisión.

En el nivel de la capa física, se utiliza GMSK (Gaussian Minimum Shift Keying ó Intercambio Mínimo Gaussiano.) Además, la codificación BCH cae en la tasa de error de bit a 10^{-3}. Los bloques cifrados son de 496 bits en longitud de la cual 416 dígitos binarios son datos. El esquema de codificación proporciona protección contra por lo menos 2 errores de dígito binario al azar y en la mayoría 32 errores de dígito binario consecutivos. Los paquetes se forman de un máximo de 47 bloques cifrados; el entrenamiento del ecualizador y las secuencias de sincronización también se insertan. Para los detalles exactos del formato del "frame", consultar [24].

HIPERLAN/2 es un estándar global para anchos de banda inalámbricos LAN que operan con un rango de datos de 54 Mbps en la frecuencia de banda de 5 GHz. HIPERLAN/2 es una solución estándar para un rango de comunicación corto que permite una alta transferencia de datos y Calidad de Servicio del tráfico entre estaciones base WLAN y terminales de usuarios. La seguridad esta provista por técnicas de encripción y protocolos de autenticación.

HIPERLAN/2 permite, acceder a la Intranet de la compañía donde uno trabaja y correr aplicaciones en tiempo real, consiguiendo un ambiente de trabajo verdaderamente flexible y móvil.

La ETSI desarrollo esta nueva tecnología WLAN, a la que nombro HiperLAN tipo2. HiperLAN2 proporciona:

- Capacidad alta y escalable así como el número de usuarios aumente en el sistema
- Manejo de ancho de banda con funcionamiento fiable para cada usuario y aplicación
- Un nivel alto de seguridad
- Capacidades de QoS para soportar virtualmente cualquier tipo de servicio o aplicación
- De fácil uso a través de un grupo de herramientas de auto configuración.

Además, nos ayuda a satisfacer las necesidades de banda amplia para móvil de banda ancha

- QoS para comunicaciones multimedia en tiempo real
- Control eficiente de la energía para la integración en los dispositivos portátiles.
- Capa de Control de Acceso al Medio (MAC) desarrollada y optimizada para comunicaciones de radio para entregar un rendimiento de procesamiento posible más alto sobre la interfase de aire, inclusive cuando el número de usuarios sea potencialmente elevado dentro de una célula.
- Selección de Frecuencia Dinámica (DFS)
- Capa de convergencia, ofreciendo independencia de la red de backbone teniendo en cuenta interfases con:
    - Ethernet
    - IEEE1394
    - ATM
    - 3G para sistemas móviles.

- Características de seguridad fuerte con soporte para autentificación individual y llaves por sesión, incluyendo soporte para utilizar cualquier llave pre-compartida o PKI.
- Estado avanzado de la estandarización (continuación de las extensiones 802.11 "g" y "h")
- Superior rendimiento del procesamiento, soportado por el departamento de investigación de la universidad de Bristol.
- Tiene un mayor costo efectivo.

IEEE 802.11a es un estándar para LANs inalámbricas que operan con un rango de datos de 54 Mbps en la frecuencia de banda de 5 GHz. IEEE 802.11a es una comunicación de paquetes de datos de rango corto entre estaciones base WLAN y terminales del usuario final. La comunicación directa entre terminales es posible. IEEE 802.11a también permite al usuario acceder a la Intranet de la empresa en la que trabaja.

4.5.1 Comparación entre IEEE802.11 e HIPERLAN

IEEE802.11 utiliza un mecanismo de contención para permitir a las estaciones compartir un canal inalámbrico, esta basado en acceso múltiple por censado de portadora (CSMA), como 802.3. 802.11 no puede utilizar todo lo de 802.3 por que no es posible en un ambiente inalámbrico. El MAC de 802.11 utiliza un mecanismo de evitación de la colisión para reducir la probabilidad de las colisiones. El MAC de 802.11 es designado para operar sobre capas físicas múltiples y no especifica varios parámetros dependientes del medio. Para una descripción a mayor detalle de IEEE 802.11 se puede ver [25]

IEEE802.11 siendo un sistema "médium-bit-rate" para los ambientes de interior evita ISI incluso con la antena omni-direccional, así, no necesita un alto costo del ecualizador; sin embargo, su bit rate bajo limita su uso especialmente con el advenimiento previsto de las aplicaciones multimedia. El alto bit rate de HIPERLAN's induce una necesidad de la igualación; el costo más alto de este sistema y su funcionamiento degradado cuando se presentan los problemas ocultos del nodo hacen menos atractivo para los desarrolladores del producto.

OTRAS TECNOLOGÍAS DE APOYO

Existen varias tecnologías de apoyo, algunas de estas se pueden consultar en el anexo C, ya que estas son complementarias al objetivo de esta tesis.

REFERENCIAS

[1] Gustavo A. Santana. T, Arturo Torres D., David Higuera R., " Protocolo de Autentificación de Usuario Móvil para ATM Inalámbrico" X Congreso Internacional de Computo CIC 2001, 12-16 Noviembre de 2001, Ciudad de México, D.F.

[2] V. Boyko and P. MacKenzie, Provably Secure Password – Authenticated Key Exchange Using Diffie-Hellman, Eurocrypt 2000 , September 8 2000.

[3] G. Santana and D. Higuera , "A Mobile User Authentication Protocol for Personal Communication Networks", Submitted to Wireless and Optical Communications (WOC2001) , June 27-29,2001 Banff, Canada. IASTED International Conference

[4] Technical Committee, ATM Security Framework 1.0, AF-SEC-0096.000, February 1998.

[5] Communication of the ACM, February 1995/ vol38. No2.

[6] Forrester Research Center, http://www.forrester.com

[7] Directive 2000/31/EC, http://europa.eu.int/eur-lex/en/lif/dat/2000/en_300L0031.html

[8] Dhiman Chatterjee, VC Ramesh, *Real Options for risk management in information technology projects*. ECE Department, Illinois Institute of Technology, www.Oberthurcs.com/fichier/solutions/Mobile/us/

[9] Upkar Varshney, Ron Vetter (2000), *Emerging mobile and wireless networks*. Communications of the ACM, June 2000, vol.43, no. 6

[10] Luca Dionisio, Giuseppe Della Penna, Benedetto Intrigila, Paola Inverardi Area Informatica, Universita' dell'Aquila {dionisio,gdellape,intrigila,inverardi}@univaq.it

[11] L. Dionisio, B. Intrigila, P. Inverardi, *On Mobile Commerce and Cellular Phones*. Technical Report, University of L'Aquila, submitted for publication.

[12] C. Colafigli, P. Inverardi, R. Matricciani, *InfoParco: an experience in designing an information system accessible through WEB and WAP interfaces*. Proc. HICCS-34, Maui January 2001. Extended version submitted for publication.

[13] S. Mann (2000), *Programming Applications with the Wireless Application Protocol*. Wiley

[14] Mostafa Hashem Sherif (2000), *Protocols for secure electronic commerce*. CRC Press.

[15] WAP Forum, *Wireless Transport Layer Security Specification*, http://www.wapforum.org/

[16] W. Ford, M.S. Baum (2001), *Secure electronic commerce: building the infrastructure for digital signatures and encryption*. Prentice Hall

[17] WAP FORUM. *Wireless Application Protocol Architecture Specification*. URL: http://www.wapforum.com

[18] *W/Secure SDK Whitepaper*, URL: http://www.baltimore.com/library/whitepapers/wsecure.html

[19] BALTIMORE TECHNOLOGIES LAUNCHES TELEPATHY, *Wireless Security For Mobile Commerce*. URL: http://www.baltimore.com/news/press/pr20000111b.html

[20] J. E. Padgett, C. G. Gunther, & T. Hattori, "Overview of wireless personal communications", IEEE commun. Magazine, pp. 28-42, Jan 1995.

[21] T. Freeberg, "A new technology for high speed wireless local area networks", Proc. IEEE Workshop on Wireless LANs, pp. 127-139, May 1991.

[22] B. Tuch, "An ISM band spread spectrum local area network: WaveLAN", Proc. IEEE Workshops on Wireless LANs, pp. 103-111, May 1991.

[23] A. Santamarina & F.J. López-Hernandez, Eds., "wireless LAN systems", Boston, MA: Artech House, 1994.

[24] A. Wittneben, W. Liu, "The European Wireless LAN standard HIPERLAN: key concepts and testbed results," IEEE 47th Vehicular Technology Conference, Technology in Motion (Cat. No.97CH36003), vol. 3, Phoenix, AZ, USA, 4-7 May 1997, pp. 1317-1321.

[25] Antonio DeSimone and Sanjiv Nanda, Wireless Data: Sistemas, Standards, Services, Baltzer Journals, 1995.

[26] J.E. Padgett, C. G. gunther, and T. Hattori, "overview of wireless personal communications", IEEE commun. Magazine, pp. 28-42, Jan 1995.

COMENTARIOS, CONCLUSIONES Y TRABAJO A FUTURO

COMENTARIOS Y CONCLUSIONES

Cabe, mencionar que ha lo largo del documento se hicieron comentarios parciales en cada capítulo, donde se creía conveniente hacer dichos comentarios. Sin embargo, se considera conveniente dar una serie más de comentarios y conclusiones finales a fin de cerrar los temas tratados en esta tesis.

Se realizo un estudio exhaustivo, no necesariamente total sobre los protocolos de autentificación tanto para sistemas distribuidos como para autentificación móvil.

Partiendo de un enfoque basado en métodos formales se desarrollaron protocolos de autentificación de usuario en ambientes móviles tipo PCS/PCN. Los cuales tienen las siguientes características:

- En primer lugar se obtiene un protocolo de autentificación de usuario móvil en un ambiente que considera "inter-dominios" con algunas ventajas para la autentificación de los usuarios empleando una modificación del esquema de firma digital con llave pública, que emplea certificados y curvas elípticas, como una alternativa conveniente para evaluar la seguridad

- En una segunda propuesta se presenta una variación del protocolo mencionado anteriormente, en la cual se considera la situación de autoridades de certificación distribuidas.

- Posteriormente en una tercer propuesta, se presenta un protocolo de usuario móvil el cual es pensado para los usuarios con alto grado de movilidad; donde el procedimiento de la autentificación, no emplearía los certificados debido a la movilidad experimentada por el usuario. Dicho protocolo, emplea autentificación mutua y criptografía de llave pública distribuida.

- Además, se propone incorporar el primer protocolo en una propuesta de aplicación para un protocolo de autentificación de usuario móvil para ATM Inalámbrico.

- Y se dejan varias líneas de trabajo futuro, mencionando entre estas por ejemplo, el desarrollo de otras ideas de protocolos para sugerencias metodológicas de diseño de sistemas multi-agente con autentificación y autorización en ambientes móviles.

Por otro lado, resulta importante comentar en que basamos la hipótesis, de que los protocolos propuestos resultan fuertes en términos de seguridad. Aunque ciertamente, no se hizo una demostración formal en este sentido, podemos argumentar sobre la base de [1,2,3] lo siguiente:

Debido a que desde el primer protocolo se plantea la idea de utilizar curvas elípticas (CCE), es importante que el primer paso en el proceso de generación de los parámetros de un esquema de CCE se determine el nivel de seguridad que se intenta alcanzar. Varios factores podrían influenciar esta determinación, tales como: el valor de la información, el tiempo que debe ser protegida, el tamaño de los parámetros que serán utilizados, el nivel de seguridad provisto por el esquema, el impacto de los tipos de ataques conocidos en la seguridad y eficiencia de los métodos basados en problemas de factorización entera, logaritmo discreto y logaritmo discreto de curva elíptica[1].

Como es bien sabido, varios estándares evitan el uso de algoritmos de llave pública debido al alto número de operaciones, requerimiento de llaves grandes y consumo de ancho de banda. Pero si por ejemplo, efectuamos las comparaciones de eficiencia de un criptosistema de curva elíptica, de RSA y DSA. En términos del número de operaciones en RSA, un exponente público corto puede ser empleado (aunque esto presenta algunos riesgos de seguridad) para acelerar la encriptación y la verificación de firma. En DSA y CCE, una gran proporción de la generación de una firma y transformaciones de encriptación pueden ser precalculada.

Respecto al tamaño de llaves, se puede tomar como ejemplo que los tamaños de llave pública para RSA, DSA y CCE son de 1088, 1024, y 161 bits y los tamaños de llave privada son de 2048, 160, y 160 respectivamente, por lo que resulta claro que los CCE son mas cortos que los de RSA y DSA.

Y por ultimo con respecto al ancho de banda se puede considerar el caso cuando mensajes cortos están siendo transformados, y para hacer una comparación concreta, supongamos que cada sistema esta siendo utilizado para firmar un mensaje de 2000 bits, o para encriptar un mensaje de 384 bits. Al comparar las longitudes de firmas y mensajes encriptados respectivamente, tenemos que para RSA son de 1024 y 1024, para DSA de 320 y 2048, y para CCE de 320 y 320, de acuerdo con [1,2,3]. Entonces, se puede concluir que CCE ofrece un ahorro de ancho de banda considerable sobre los otros tipos de sistemas de clave pública. Por lo tanto, todos estos ahorros redundan en velocidades más altas, menor consumo de energía y reducciones en el tamaño del código. Por lo que una implementación de DSA o CCE resulta beneficiosa particularmente en aplicaciones donde el ancho de banda, capacidad de procesamiento, disponibilidad de energía o almacenamiento están restringidos. Tales aplicaciones incluyen transacciones sobre dispositivos inalámbricos, computación en dispositivos portátiles tales como: PDAs, teléfonos celulares, computadoras portátiles, y tarjetas inteligentes por ejemplo.

---

[1] Dentro de los algoritmos de ataques mas conocidos se encuentran el Quadartic Sieve (QS), Number Field Sieve (NFS), Pollard-ρ, Pollar- λ, Pohlig-Hellman, index-calculus principalmente.

Para cerrar nos resta comentar, como con todos los criptosistemas, especialmente con criptosistemas de llave pública, toma años de evaluación pública antes de que un razonable nivel de confianza en un nuevo sistema sea establecido. Las CCEs parecen estar en el camino de alcanzar ese nivel, nos llamo la atención voltear la vista a ellas e incorporarlas en nuestras propuestas, ya que en los últimos años aparecen implementaciones de estas en la seguridad de e-mail, tarjetas inteligentes entre otras.

Sin embargo, un factor que aún queda ampliamente incierto es su seguridad, tal como sucede con todos los criptosistemas de llave pública utilizados en la práctica, su seguridad no ha sido probada, sino que esta basada en la incapacidad de encontrar ataques. Si no existen ataques sobre alguno de esos sistemas, podrían ser descubiertos tarde o temprano. En tal caso se debería cambiar a alguna otra alternativa como los ya tradicionales criptosistemas utilizados en la práctica basados en el problema de factorización entera y logaritmo discreto.

No obstante, los CCE pueden ser implementados eficientemente, y tienen un número de ventajas que pueden hacerlo la mejor elección en un rango de aplicaciones, como aquellas en las que los recursos disponibles (memoria, procesamiento, energía, ancho de banda, etc.) son reducidos. Más aún, con un número de estándares estando en preparación, la interoperabilidad entre los diferentes productos será mucho más fácil de obtener.

Y bien, si consideramos que hemos incorporado herramientas matemáticas y algoritmos para criptografía fuertes (avalados por organismos estandarizadores), a los protocolos propuestos en esta tesis, que fueron diseñados a partir de una extensión de la familia de los protocolos de Beller-Chang-Yacobi, además de la influencia de muchos otros, Aziz-Diffie, Needham-Schroeder, etc. Con amplio camino recorrido y aceptación y adaptación de muchos de estos para su implementación en sistemas reales. Consideramos que vamos por buen camino, claro esta que resta camino por recorrer para probar de manera completa y formal su validez y seguridad y posteriormente pensar en la implementación. Sin embargo, esto abre nuevas líneas de trabajo futuro.

TRABAJOS FUTUROS

Son varias las líneas de trabajo que se desprenden de esta tesis. De las cuales podemos mencionar las siguientes:

- Utilizar cualquiera de los 4 puntos tratados en el articulo [4] en la parte de la verificación
  - Modelando y verificando protocolos utilizando lenguajes de especificación y herramientas de verificación no desarrolladas específicamente para el análisis de protocolos criptográficos.
  - Desarrollando sistemas expertos que un diseñador de protocolo puede utilizar para investigar escenarios diferentes.
  - Modelando y verificando el protocolo utilizando lógicas modales desarrolladas para el análisis de conocimiento y creencia.
  - Desarrollando un modelo formal basado en las propiedades de "reescritura-de-termino" algebraico de sistemas criptográficos.

Particularmente, aunque de manera parcial, en este trabajo de tesis ya se comenzó la verificación con lógica modal (BAN y otra lógica simplificada desprendida de la misma lógica BAN, ver anexo A) quedando para trabajo futuro modelar y especificar de manera completa los protocolos aquí presentados[2].

- Dirigir los protocolos propuestos al desarrollo de otras ideas de protocolos, por ejemplo el trabajo de sugerencias metodológicas de diseño de sistemas multi-agente con autentificación y autorización en ambientes móviles[3].

- Pensar en llevar los protocolos trabajados de esta tesis, a un esquema de "zero-knowledge"

- Aplicar estos protocolos a esquemas que trabajen sobre la base de inter-PKIs, entre otras más líneas futuras de trabajo.

REFERENCIAS

[1] "Current public-key cryptografic systems", Whitepaper 2, Certicom Corp., abril 1997. http://www.certicom.com

[2] NIST, Digital signature standard, FIPS Pub 186-2 NIST, Enero 2000.

[3] Certicom research, Sec 1: Elliptic curve cryptography, working Draft o.5, Standards for Efficient Cryptography, septiembre 1999. http://www.secg.org

[4] C. Meadows. Formal verification of cryptographic protocols: "A survey. In Advances in Cryptology – ASIACRYPT'94, pp. 135-150. Springer-Verlang, 1995.

[2] Cabe aclarar que por las características propias del tercer protocolo propuesto en el capitulo tres, no sería completamente apropiado y justo solo verificar este, con BAN, se considera conveniente realizar otra forma de verificación de las aquí mencionadas.

[3] Este trabajo ha sido enviado para su evaluación en una conferencia Internacional, bajo el nombre "Methodological Issues for Designing Multi-Agent Systems with Authentication and Authorisation in Mobile Environment", por ser tal su status y no contar con una evaluación propiamente hecha, no se agrega en la bibliografía.

ANEXO A

A.1 PANORAMA GENERAL DE LAS LÓGICAS DE AUTENTIFICACIÓN

Se conoce que los protocolos criptográficos son empleados para proporcionar servicios de seguridad en sistemas de comunicación, siendo los sistemas distribuidos los de mayor interés. Sin embargo, si el protocolo no es diseñado con suficiente cuidado, es posible que contenga defectos, los cuales pueden ser el punto ideal de inicio para varios ataques [1, 2, 3, 4]. Dichos defectos pueden ser sutiles y difíciles de encontrar, motivo por el cual el diseño y verificación informal de los protocolos no es suficiente. Métodos formales parecen ser los aptos para resolver estos problemas [5, 6].

Uno de los métodos formales es utilizar lógicas para razonamiento respecto de las propiedades del protocolo, o bien, métodos algebraicos, máquinas de estado. En este sentido muchas publicaciones al respecto de dichas lógicas se pueden encontrar en la literatura académica (como ejemplo tenemos: [3, 7, 8, 9, 10, 11, 12]) Sin embargo la mayoría de estas lógicas así como otras herramientas formales [13], solamente están concentradas en la verificación o en la planeación. Por lo que resulta importante resaltar la característica de esta lógica, la cuál además, nos ayuda en el proceso de diseño de protocolos criptográficos de autentificación.

La importancia de las herramientas de diseño se enfatiza en [5], donde se hace una propuesta de capas, conocido como el principio de diseño general. La propuesta por capas esta basado en un "stack" de modelos con diferentes niveles de abstracción. El diseñador del protocolo primero utiliza un modelo relativamente abstracto para construir y verificar el protocolo (capa superior). Si el protocolo es correcto en ese nivel, entonces el diseñador puede pasarse a una abstracción menor, un modelo más detallado, en el cual "implementa" (refina) el abstracto. Finalmente, a través de la ejecución repetida de este proceso, una especificación detallada o incluso el código del protocolo actual es producida. Para nuestro caso implementación significa refinamiento, pero el resultado de la implementación es una descripción del protocolo más detallada (no necesariamente la final). Posteriormente esta descripción de protocolo podría ser analizada con otras herramientas, basadas en modelos mas detallados.

A.2 PRUEBAS DE PROTOCOLOS

Esta sección contiene algunas pruebas formales de seguridad de algunos protocolos descritos anteriormente en este documento. Una introducción al formalismo y a la notación de la lógica de autentificación de Burrows, Abadi, y Needham (lógica de BAN) es proporcionada. Esto es seguido por las pruebas con la lógica de BAN a algunos de protocolos estudiados, incluyendo una explicación de cómo cada línea en el análisis se obtiene y un resumen de lo que alcanza cada mensaje. Posteriormente se analiza el primer protocolo propuesto en esta tesis[1], ya que de este se derivan los demás protocolos propuestos, más aún, no solo se utiliza esta lógica sino también se hace uso de otra lógica simplificada que nos ayudo en la parte del diseño.

[1] Se deja para trabajo futuro el estudio formal de los otros protocolos propuestos en esta tesis.

A.3 LÓGICA DE LA AUTENTIFICACIÓN

La lógica de autentificación o lógica de BAN [14, 15] proporciona un formalismo, bajo el cual un análisis más riguroso de protocolos criptográficos sea posible que por métodos informales. Esto nos permite razonar sobre la creencia de los principales implicados en los protocolos. Para realizar este análisis, un protocolo se debe primero traducir de las muchas y variadas notaciones simbólicas usadas en descripciones del protocolo en una forma "idealizada" que describa a los mensajes y a participantes. El protocolo idealizado substituye los mensajes originales por una secuencia de fórmulas lógicas. Después, cualquier suposición requerida por el protocolo necesita ser identificada y ser expresada formalmente. Estas suposiciones formales deben apoyar a la verdad para lograr las metas del protocolo. Finalmente, las metas del protocolo se identifican y se expresan formalmente. Esto consiste en típicamente intercambiar una llave compartida y en identificar la identidad verdadera de una o de ambas partes implicadas. Combinando el protocolo idealizado y las suposiciones idealizadas con los postulados de la lógica de BAN una "prueba" se puede obtener describiendo la creencia y las metas logradas por los participantes.

Ha habido observaciones donde se dice que la lógica de BAN no puede distinguir versiones dañadas y seguras de algunos protocolos. En detalle, Boyd y Mao [16] proporcionan dos ejemplos en los cuales aparece esa lógica de BAN aprobando protocolos que son en la práctica falsos. Paul van Oorschot contradice estas afirmaciones en [17], y demuestra que de hecho los resultados erróneos obtenidos en los dos ejemplos eran el resultado de permitir una suposición falsa en un caso e incorrectamente de "idealizar" el protocolo en el segundo caso. Los creadores de la lógica de BAN procuran identificar el alcance de la lógica en [14]. Ellos Describen el propósito del protocolo como sigue:

La meta de la lógica es describir la creencia de las partes dignas de confianza implicadas en la autentificación, y la evolución de estas creencias mientras que los principales se comunican.

En el caso donde la lógica no captura un protocolo, se emplean otras técnicas formales o informales. Éste es el caso del protocolo de seguridad usado en CDPD.

A.3.1 NOTACIÓN

La lógica contiene tres tipos de objetos: principales, llaves de encriptado, y fórmulas (o declaraciones). Los mensajes se identifican con declaraciones en la lógica. Los símbolos tales como A, B, y S denotan típicamente los principales específicos; K_{ab}, K_{as}, y K_{bs} denotan llaves compartidas específicas; K_a, K_b, y K_s denotan llaves públicas específicas, y k_a^{-1}, K_b^{-1}, y K_s^{-1} denotan las llaves secretas correspondientes; N_a, N_b, y el N_c denotan declaraciones específicas. Los símbolos P, Q y R en un rango por encima de los principales. X y Y en el rango por encima de las declaraciones; K en un rango por encima de las llaves de encripción.

La conjunción es denotada por una coma. Las conjunciones se tratan como conjuntos. Además de la conjunción, la lógica de BAN proporciona las construcciones siguientes:

- $P \mid\equiv X$: P cree X, o P tiene la convicción de que X es verdadero. P puede actuar como si X fuera verdad. Esta construcción es central en la lógica.
- $P \lhd X$: P ve a ó alguna vez recibió X. El mensaje X se ha enviado a P, quien puede leer y repetir (posiblemente después de un cierto desciframiento).

- P ~ X: P alguna vez dijo ó envió X. En un cierto momento P envió un mensaje incluyendo la declaración X. No la conclusión inmediata puede ser hecha si el mensaje fue enviado hace tiempo o durante el funcionamiento actual del protocolo. Se sabe que P creyó X cuando el mensaje fue enviado.
- P $\Rightarrow$ X: P tiene jurisdicción (control) sobre X. El principal P es una autoridad en X y se debe confiar en esta cuestión. Se utiliza esta construcción cuando un principal es responsable de una cierta declaración. Por ejemplo, las llaves de encripción necesitan ser generadas con cuidado, y en algunos protocolos ciertos servidores son confiables en hacer esto correctamente. Esto es reflejado por la suposición de que los principales creen que el servidor tiene jurisdicción sobre declaraciones con respecto a la calidad de llaves.
- # (X): La fórmula X es fresca (fue recientemente enviada), es decir, X no se ha enviado en un mensaje en ningún momento antes de la corrida del actual protocolo. Esto es verdadadero para nonces y los timestamps.
- P $\leftrightarrow^{K}$ Q: P y Q pueden utilizar la llave compartida K para comunicarse. La llave K es buena en el sentido de que nunca sea descubierta por cualquier principal excepto P o Q, o un principal confiable para P o Q.
- $\mapsto^{K}$ P: P tiene a K como llave pública. La llave secreta correspondiente K^{-1} es conocida solamente para P o un principal confiable para P.
- P $\leftrightarrows^{X}$ Q: La fórmula X es un secreto conocido solamente por P y Q, y posiblemente principales confiables para ellos. X se puede utilizar solamente por P o Q para probar sus identidades una a la otra. X es a menudo fresco así como secreto. Un buen ejemplo de un secreto compartido es una contraseña o password.
- {X}K: Esto representa la fórmula X encriptada bajo la llave K. El esquema de encripción que es utilizado, generalmente es evidente de la llave que es utilizada.
- <X>Y: Esto representa la combinación de X con la fórmula Y. Se proyecta que Y sea secreta, y que su presencia pruebe la identidad de quienquiera que pronuncie <X>Y.

A.3.2 POSTULADOS DE LA LÓGICA

En cualquier estudio de protocolos de seguridad, es importante distinguir la sincronización de declaraciones o de acontecimientos. Si no, el reenvío de mensajes anteriores pudiese ocurrir desapercibido. La lógica de BAN divide el tiempo en dos épocas: pasado y presente. El presente consiste del tiempo durante la corrida actual del protocolo bajo consideración. Cualquier mensaje enviado antes de éste se considera que esta en el pasado, y se debe rechazar por el protocolo como no fiable. Cualquier creencia sostenida en el presente es estable para la totalidad del funcionamiento del protocolo. Esto permite que los principales deduzcan la creencia de otro principal en las declaraciones que pueden hacer. Sin embargo, la creencia llevada a cabo en el pasado no se lleva necesariamente adelante en el presente. Esta división simple del tiempo es suficiente para analizar protocolos de autentificación.

Un factor importante en cualquier protocolo de seguridad es la función del encriptado. Aunque La lógica de BAN no hace ninguna tentativa de evaluar la fuerza del criptosistema que es utilizado, hace ciertas suposiciones con respecto al proceso de encriptado. Afortunadamente, estas suposiciones se sostienen para la mayoría de los criptosistemas o se pueden dirigir durante la derivación de la forma idealizada del protocolo. Primero, la encripción garantiza que cada sección cifrada no se puede alterar o ensamblar de secciones más pequeñas de encripción. Si un mensaje contiene dos secciones encriptadas separadas, se tratan como si llegaran en mensajes separados. En segundo lugar, un mensaje no puede ser entendido por un principal que no sepa la

llave (o la llave inversa en el caso de criptografía de llave pública.). Finalmente, los mensajes necesitan contener la suficiente información para que un principal detecte (y no hacer caso) sus propios mensajes.

A continuación se encuentra una lista de los postulados lógicos usados en la lógica de BAN. Estos postulados se utilizan en la derivación de las pruebas de los protocolos de seguridad.

- Reglas de Significado de-Mensaje (Message-meaning): Estas reglas se refieren a la interpretación de mensajes. Explican cómo derivar la creencia sobre el origen de mensajes. Hay tres reglas: dos para los mensajes encriptados con llaves compartidas o públicas, y uno para los mensajes con secretos.

  Para las llaves compartidas, se postulan:

$$\frac{P \mid\equiv Q \leftrightarrow^{K} P,\ P \lhd \{X\}K}{P \mid\equiv Q \sim X}$$

  Es decir, si Q y P comparten una llave secreta K y P ve un mensaje encriptado cifrado utilizando K, entonces P cree Q alguna vez dijo X. Similarmente, para las llaves públicas, se postulan:

$$\frac{P \mid\equiv \mapsto^{K} Q,\ P \lhd \{X\}K^{-1}}{P \mid\equiv Q \sim X}$$

  Para secretos compartidos, se postula:

$$\frac{P \mid\equiv Q \leftrightarrows^{Y} P,\ P \lhd \langle X\rangle Y}{P \mid\equiv Q \sim X}$$

  Es decir si P y Q comparten el secreto Y y P ve X con el secreto compartido Y, entonces P cree que Q alguna vez dijo X.

- Regla de verificación de Nonce (Nonce-verification): Comprueba que un mensaje sea reciente (es decir, en el presente) y por lo tanto que el remitente todavía cree en él.

$$\frac{P \mid\equiv \#(X),\ P \mid\equiv Q \sim X}{P\ Q\ X}$$

- Regla de Jurisdicción (Jurisdiction rule): Condiciona que si P cree que Q tiene jurisdicción o control sobre X entonces P confía en Q en la certeza de X.

$$\frac{P \mid\equiv Q \Rightarrow X,\ P \mid\equiv Q \mid\equiv X}{P \mid\equiv X}$$

- Una propiedad necesaria del operador de la creencia es P cree un conjunto de declaraciones si y solamente si P cree cada declaración individual por separado. Esto justifica las siguientes reglas:

$$\frac{P \mid\equiv X,\ P \mid\equiv Y}{P \mid\equiv (X, Y)} \qquad \frac{P \mid\equiv (X, Y)}{P \mid\equiv X} \qquad \frac{P \mid\equiv Q \mid\equiv (X, Y)}{P \mid\equiv Q \mid\equiv X}$$

Reglas similares pueden ser introducidas como requisito.

- Reglas similares como las de arriba aplican para el operador ~:

$$\frac{P \mid\equiv Q \sim (X, Y)}{P \mid\equiv Q \sim X}$$

Hay que notar que si $P \mid\equiv Q \sim X$ y $P \mid\equiv Q \sim Y$ esto no significa que $P \mid\equiv Q \sim (X, Y)$, ya que esto implicaría que las dos partes X y Y fueron dichas al mismo momento.

- Si un principal ve una formula entonces ve sus componentes, con tal que se sepan las llaves necesarias:

$$\frac{P \lhd (X, Y)}{P \lhd X} \qquad \frac{P \lhd \langle X \rangle Y}{P \lhd X} \qquad \frac{P \mid\equiv Q \leftrightarrow^{K} P,\ P \lhd \{X\}K}{P \lhd X}$$

$$\frac{P \mid\equiv \mapsto^{K} P,\ P \lhd \{X\}K}{P\ X} \qquad \frac{P \mid\equiv \mapsto^{K} Q,\ P \lhd \{X\}K^{-1}}{P\ X}$$

- Para permitir el uso de llaves no certificadas, por ejemplo con certificados, el siguiente suplemento para las reglas arriba mencionadas es requerido:

$$\frac{P \mid\equiv R \sim Q \leftrightarrow^{K} P,\ P \lhd \{X\}K}{P \lhd X}$$

- Si una parte de una formula se conoce que es fresca, entonces la formula entera debe también ser fresca:

$$\frac{P \mid\equiv \#(X)}{P \mid\equiv \#(X, Y)}$$

Utilizando estos postulados, las pruebas pueden ser construidas en la lógica de BAN. Una formula X es demostrable en la lógica desde una formula Y si hay una secuencia de formulas Zo, Zn donde Zo = Y, Zn = X, y cada Z_{i+1} puede ser obtenida desde las anteriores por la aplicación de un regla.

A.4 ANALISIS

A 4.1 PROTOCOLO DE AZIZ/DIFFIE

Protocolo Idealizado

Mensaje 1: $A \rightarrow B \{ \mapsto^{Ka} A\}K^{ca\text{-}1}$
Mensaje 2: $B \rightarrow A \{\mapsto^{Kb} B\}K^{ca\text{-}1}, \{(A \leftrightarrow^{RN1} B), Na\}K_b^{-1}$
Mensaje 3: $A \rightarrow B \{(A \leftrightarrow^{RN2} B), \{RN1\}K_a\}K_a^{-1}$

Suposiciones de prueba

a) $A \mid\equiv \mapsto^{Ka} A$
b) $A \mid\equiv \mapsto^{Kca} CA$
c) $A \mid\equiv (CA \Rightarrow \mapsto^{K} B)$
d) $A \mid\equiv \#(Na)$
e) $A \mid\equiv A \leftrightarrow^{RN2} B$
f) $B \mid\equiv \mapsto^{Kb} B$
g) $B \mid\equiv \mapsto^{Kca} CA$
h) $B \mid\equiv (CA \Rightarrow \mapsto^{K} A)$
i) $B \mid\equiv \#(RN1)$
j) $B \mid\equiv A \leftrightarrow^{RN1} B$
k) $CA \mid\equiv \mapsto^{Ka} A$
l) $CA \mid\equiv \mapsto^{Kca} CA$
m) $CA \mid\equiv \mapsto^{Kb} B$
n) $A \mid\equiv (B \Rightarrow A \leftrightarrow^{RN1} B)$
o) $B \mid\equiv (A \Rightarrow A \leftrightarrow^{RN2} B)$

Nota: Los certificados, se asume que sean frescos.

Prueba

Mensaje 1:

$B \lhd \{\mapsto^{Ka} A\} K_{ca}^{-1}$
$B \mid\equiv CA \sim \mapsto^{Ka} A$ (g, message-meaning)
$B \mid\equiv \mapsto^{Ka} A$ (certificado fresco, nonce verification, h, y jurisdiction)

Resultado: B obtiene llave publica de A

Mensaje 2:
$A \lhd \{\mapsto^{Kb} b\} K_{ca}^{-1}, \{(A \leftrightarrow^{RN1} B), Na\} K_b^{-1}$
$A \mid\equiv \mapsto^{Kb} B$
$A \mid\equiv B \sim \{(A \leftrightarrow^{RN1} B), Na\}$ (message meaning)
$A \mid\equiv B \mid\equiv (A \leftrightarrow^{RN1} B)$ (d, nonce verification)

$A \mid\equiv (A \leftrightarrow^{RN1} B)$ (n, jurisdiction)

Resultado: A autentifica a B. A obtiene la parte de B de la llave de sesión.

Mensaje 3:

$B \lhd \{(A \leftrightarrow^{RN2} B), \{RN1\} K_a\} K_a^{-1}$
$B \mid\equiv A \sim \{(A \leftrightarrow^{RN2} B), \{RN1\} K_a\}$ (message meaning)
$B \mid\equiv A \mid\equiv (A \leftrightarrow^{RN2} B)$ (i, nonce verification)
$B \mid\equiv (A \leftrightarrow^{RN2} B)$ (o, jurisdiction)

Resultado: B autentifica A. B obtiene la parte de A de la llave de sesión.

Conclusiones

Dado,
$K_{ab} = RN1 \oplus RN2$

Por lo tanto,

$A \mid\equiv A \leftrightarrow^{Kab} B$	(e) Llave de sesión compartida
$B \mid\equiv A \leftrightarrow^{Kab} B$	(j) Llave de sesión compartida
$A \mid\equiv B \mid\equiv (A \leftrightarrow^{RN1} B)$	A autentifica a B
$B \mid\equiv A \mid\equiv (A \leftrightarrow^{RN2} B)$	B autentifica a A

A. 4.2 GSM

Protocolo Idealizado

Mensaje 3: $C \rightarrow B\ N_c, A \leftrightarrow^{Kab} B, A \leftrightarrows^{\{Nc\}Kac} C, \#\{N_c\}K_{ac}$
Mensaje 4: $B \rightarrow A\ N_c$
Mensaje 5: $A \rightarrow B <\{N_c\}K_{ac} > \{N_c\}K_{ac}$

Suposiciones de prueba

a) $A \mid\equiv A \leftrightarrow^{Kac} C$
b) $C \mid\equiv A \leftrightarrow^{Kac} C$
c) $A \mid\equiv A \leftrightarrows^{\{Nc\}Kac} C$
d) $C \mid\equiv A \leftrightarrows^{\{Nc\}Kac} C$
e) $B \mid\equiv C \mid\equiv (N_c, A \leftrightarrow^{Kab} B, A \leftrightarrows^{\{Nc\}Kac} C, \#\{N_c\}K_{ac})$
f) $A \mid\equiv C \Rightarrow A \leftrightarrow^{Kab} B$
g) $B \mid\equiv C \Rightarrow A \leftrightarrow^{Kab} B$
h) $A \mid\equiv C \Rightarrow A \leftrightarrows^{\{Nc\}Kac} C$
i) $B \mid\equiv C \Rightarrow A \leftrightarrows^{\{Nc\}Kac} C$
j) $B \mid\equiv C \Rightarrow \#\{N_c\}K_{ac}$
k) $A \mid\equiv C \sim N_c$

Nota: La suposición e) esta basada en la creencia de que la conexión alámbrica es segura.

Prueba

Mensaje 3:

$B \lhd N_c, A \leftrightarrow^{Kab} B, A \leftrightarrows^{\{Nc\}Kac} C, \#\{Nc\}K_{ac}$

$B \mid\equiv C\ (N_c, A \leftrightarrow^{Kab} B, A \leftrightarrows^{\{Nc\}Kac} C, \#\{N_c\}K_{ac})$ (Enlace alámbrico seguro)
$B \mid\equiv A \leftrightarrow^{Kab} B$ (g, jurisdiction)
$B \mid\equiv A \leftrightarrows^{\{Nc\}Kac} C$ (i, jurisdiction)
$B \mid\equiv \#\{N_c\}K_{ac}$ (j, jurisdiction)

Resultado: B obtiene la llave de sesión compartida desde C.

Mensaje 4:

$A \lhd N_c$
$A \mid\equiv C \sim N_c$
$A \mid\equiv C \sim A \leftrightarrow^{Kab} B$ ($K_{ab} = A8\ (K_{ac}, N_c)$)

Resultado: A cree que C en algún momento creó una llave de sesión compartida.

Mensaje 5:

$B \lhd \{N_c\}K_{ac} > \{N_c\}K_{ac}$
$B \mid\equiv A \sim \{N_c\}K_{ac}$ (jurisdiction)
$B \mid\equiv A \mid\equiv \{N_c\}K_{ac}$ (nonce verification)

Resultado: B autentifica a A.

Conclusiones

Por lo tanto,
$B \mid\equiv A \leftrightarrow^{Kab} B$ Llave de sesión compartida
$A \mid\equiv C \sim A \leftrightarrow^{Kab} B$ A no esta convencida que la llave de sesión compartida es actual.
$B \mid\equiv A \mid\equiv \{N_c\}K_{ac}$ B autentifica a A

A.4.3 CDPD

Protocolo idealizado

Mensaje 1: $A \rightarrow B\ A \leftrightarrow^{Na} B$
Mensaje 2: $B \rightarrow A\ A \leftrightarrow^{Nb} B$
Mensaje 3: $A \rightarrow B\ \{<N_c>K\}K_{ab}$
Mensaje 4: $B \rightarrow C\ <N_c>K$
Mensaje 5: $C \rightarrow B\ <N_c>K, \#<N_c>K$
Mensaje 6: $B \rightarrow A\ \{<N_c>K\}K_{ab}$

Suposiciones de prueba

a) $A \mid\equiv B \mid\equiv A \leftrightarrow^{Nb} B$
b) $B \mid\equiv A \mid\equiv A \leftrightarrow^{Na} B$

c) $A \mid\equiv A \leftrightarrows^{K} C$

d) $C \mid\equiv A \leftrightarrows^{K} C$

e) $A \mid\equiv B \Rightarrow A \leftrightarrow^{Nb} B$

f) $B \mid\equiv A \Rightarrow A \leftrightarrow^{Na} B$

g) $A \mid\equiv \#(N_c)$

h) $C \mid\equiv B \mid\equiv <N_c>K$

i) $B \mid\equiv C \mid\equiv <N_c>K, \#<N_c>K$

j) $B \mid\equiv C \Rightarrow <N_c>K, \#<N_c>K$

k) $C \mid\equiv \#(N_c)$

Nota: Las suposiciones i) y j) están basadas en la creencia de que la conexión alámbrica es segura.

Prueba

Mensaje 1:

$B \lhd A \leftrightarrow^{Na} B$

$B \mid\equiv A \mid\equiv A \leftrightarrow^{Na} B$

$B \mid\equiv A \leftrightarrow^{Na} B$ (f, jurisdiction)

Resultado: B obtiene la parte de A de la llave de sesión compartida.

Mensaje 2:

$A \lhd A \leftrightarrow^{Nb} B$

$A \mid\equiv B \mid\equiv A \leftrightarrow^{Nb} B$

$A \mid\equiv A \leftrightarrow^{Nb} B$ (e, jurisdiction)

Resultado: A obtiene la parte de B de la llave de sesión.

Mensaje 3:

$B \lhd \{<N_c>K\}K_{ab}$

$B \mid\equiv A \sim <N_c>K$ (message meaning)

Mensaje 4:

$C \lhd <N_c>K$

$C \mid\equiv A \sim N_c$ (d, message meaning)

$C \mid\equiv A \mid\equiv N_c$ (k, nonce verification)

Resultado: C autentifica a A.

Mensaje 5:

$B \lhd <N_c>K, \#<N_c>K$

$B \mid\equiv C \mid\equiv <N_c>K, \#<N_c>K$ (suposición i)

$B \mid\equiv <N_c>K$ (j, jurisdiction)

$B \mid\equiv \#<N_c>K$ (j, jurisdiction)
$B \mid\equiv A <N_c>K$ (nonce verification)

Resultado: B autentifica a A.

Mensaje 6:

$A \lhd \{<N_c>K\}K_{ab}$
$A \mid\equiv B \sim <N_c>K$ (message meaning)

Resultado: A cree que B existió alguna vez.

Conclusiones

Dado,

$K_{ab} = N_a \oplus N_b$

Por lo tanto,

$A \mid\equiv A \leftrightarrow^{Kab} B$	Llave de sesión compartida
$B \mid\equiv A \leftrightarrow^{Kab} B$	Llave de sesión compartida
$A \mid\equiv B \sim <N_c>K$	A cree que B existió alguna vez.
$B \mid\equiv A \mid\equiv <N_c>K$	B autentifica a A
$C \mid\equiv A \mid\equiv N_c$	C autentifica a A

A.4.5 KERBEROS

Protocolo idealizado

Mensaje 2: $K \rightarrow A \{T_k, (A \leftrightarrow^{Kab} B), \{T_k , A \leftrightarrow^{Kab} B\}K_{bk}\}K_{ak}$
Mensaje 3: $A \rightarrow B \{T_k, A \leftrightarrow^{Kab} B\}K_{bk} , \{T_a , A \leftrightarrow^{Kab} B\}K_{ab}$
Mensaje 4: $B \rightarrow A \{T_b, (A \leftrightarrow^{Kas} S), \{T_b , A \leftrightarrow^{Kas} S\}K_{bk}\}K_{ab}$
Mensaje 5: $A \rightarrow S \{T_b, (A \leftrightarrow^{Kas} S)\}K_{sk} , \{T_a , A \leftrightarrow^{Kas} S\}K_{as}$
Mensaje 6: $S \rightarrow A \{T_a, A \leftrightarrow^{Kas} S\}K_{as}$

Suposiciones de prueba

a) $A \mid\equiv A \leftrightarrow^{Kak} K$
b) $B \mid\equiv B \leftrightarrow^{Kbk} K$
c) $S \mid\equiv S \leftrightarrow^{Ksk} K$
d) $K \mid\equiv S \leftrightarrow^{Ksk} K$
e) $K \mid\equiv A \leftrightarrow^{Kak} K$
f) $K \mid\equiv B \leftrightarrow^{Kbk} K$
g) $A \mid\equiv (B \Rightarrow A \leftrightarrow^{K} S)$
h) $S \mid\equiv (B \Rightarrow A \leftrightarrow^{K} S)$
i) $A \mid\equiv (K \Rightarrow A \leftrightarrow^{K} B)$

j) $B \mid\equiv (K \Rightarrow A \leftrightarrow^{K} B)$
k) $A \mid\equiv \#(k)$
l) $B \mid\equiv \#(T_k)$
m) $B \mid\equiv \#(T_a)$
n) $A \mid\equiv \#(T_b)$
o) $S \mid\equiv \#(T_a)$
p) $S \mid\equiv \#(T_b)$
q) $B \mid\equiv S \leftrightarrow^{Ksk} K$
r) $A \mid\equiv \#(T_a)$

Prueba

Mensaje 2:

$A \lhd \{T_k, (A \leftrightarrow^{Kab} B), \{T_k, A \leftrightarrow^{Kab} B\}K_{bk}\}K_{ak}$
$A \mid\equiv K \sim (T_k, (A \leftrightarrow^{Kab} B), \{T_k, A \leftrightarrow^{Kab} B\}K_{bk})$ (a, message meaning)
$A \mid\equiv K \sim (T_k, A \leftrightarrow^{Kab} B)$
$A \mid\equiv K \mid\equiv (T_k, A \leftrightarrow^{Kab} B)$ (k, nonce verification)
$A \mid\equiv A \leftrightarrow^{Kab} B$ (i, jurisdiction)

Resultado: A obtiene una llave de sesión compartida con B.

Mensaje 3:

$B \lhd \{T_k, A \leftrightarrow^{Kab} B\}K_{bk}, \{T_a, A \leftrightarrow^{Kab} B\}K_{ab}$
$B \mid\equiv K \sim (T_k, A \leftrightarrow^{Kab} B)$ (b, message meaning)
$B \mid\equiv K \mid\equiv (T_k, A \leftrightarrow^{Kab} B)$ (l, nonce verification)
$B \mid\equiv A \leftrightarrow^{Kab} B$ (j, jurisdiction)
$B \mid\equiv A \sim (T_a, A \leftrightarrow^{Kab} B)$ (message meaning)
$B \mid\equiv A \mid\equiv A \leftrightarrow^{Kab} B$ (m, nonce verification)

Resultado: B autentifica a A. B obtiene una llave de sesión compartida con A.

Mensaje 4:

$A \lhd \{T_b, (A \leftrightarrow^{Kas} S), \{T_b, A \leftrightarrow^{Kas} S\}K_{bk}\}K_{ab}$
$A \mid\equiv B \sim (T_b, (A \leftrightarrow^{Kas} S), \{T_b, A \leftrightarrow^{Kas} S\}K_{bk})$ (message meaning)
$A \mid\equiv B \mid\equiv A \leftrightarrow^{Kas} S$ (n, nonce verification)
$A \mid\equiv A \leftrightarrow^{Kas} S$ (g, jurisdiction)

Resultado: A autentifica a B. A obtiene una llave de sesión con S.

Mensaje 5:

$S \lhd \{T_b, (A \leftrightarrow^{Kas} S)\}K_{sk}, \{T_a, A \leftrightarrow^{Kas} S\}K_{as}$
$S \mid\equiv B \sim (T_b, A \leftrightarrow^{Kas} S)$ (q, message meaning rule)
$S \mid\equiv B \mid\equiv A \leftrightarrow^{Kas} S$ (p, nonce verification)
$S \mid\equiv A \leftrightarrow^{Kas} S$ (h, jurisdiction rule)
$S \mid\equiv A \sim (T_a, A \leftrightarrow^{Kas} S)$ (message meaning rule)
$S \mid\equiv A \mid\equiv A \leftrightarrow^{Kas} S$ (o, nonce verification)

Resultado: S autentifica a A. S obtiene una llave de sesión compartida con A.

Mensaje 6:

$A \lhd \{T_a, A \leftrightarrow^{Kas} S\}K_{as}$
$A \mid\equiv S \sim (T_a, A \leftrightarrow^{Kas} S)$ (message meaning)
$A \mid\equiv S \mid\equiv A \leftrightarrow^{Kas} S$ (r, nonce verification)

Resultado: A autentifica a S.

Conclusiones

Por lo tanto,

$A \mid\equiv S \mid\equiv A \leftrightarrow^{Kas} S$ A autentifica a S
$A \mid\equiv A \leftrightarrow^{Kas} S$ Llave de sesión compartida
$S \mid\equiv A \mid\equiv A \leftrightarrow^{Kas} S$ S autentifica a A
$S \mid\equiv A \leftrightarrow^{Kas} S$ Llave de sesión compartida

A. 4.6 PROTOCOLO DE AUTENTIFICACIÓN DE USUARIOS PARA REDES DE COMUNICACIÓN PERSONAL

Protocolo idealizado

Mensaje 1: $A \rightarrow EB \{ \mapsto^{Ka} A\}K_{KCC}^{-1}, \{T_a, A \leftrightarrow^{NR} EB\}K_a^{-1}$
Mensaje 2: $EB \rightarrow KCC \{\mapsto^{Keb} EB\}K_{KCC}^{-1}, \{T_b, \{T_a, EB \leftrightarrow^{NR1} KCC\}K_{eb}^{-1}\}K_{KCC}$
Mensaje 3: $KCC \rightarrow EB \{ \mapsto^{As} KCC\}K_{KCC}^{-1}, \{T_{KCC}, KCC \leftrightarrow^{As} EB\}K_{KCC}^{-1}$
Mensaje 4: $EB \rightarrow A \{ \mapsto^{Ks} A\}K_{EB}^{-1}$

Suposiciones de prueba

a) $A \mid\equiv \mapsto^{Ka} A$
b) $A \mid\equiv \mapsto^{K}_{KCC} KCC$, ó $A \mid\equiv \mapsto^{K}_{EB} EB$
c) $A \mid\equiv (KCC \Rightarrow \mapsto^{K} EB)$, $A \mid\equiv (EB \Rightarrow \mapsto^{K} EB)$
d) $A \mid\equiv \#(T_a)$
e) $A \mid\equiv A \leftrightarrow^{NR} EB$
f) $EB \mid\equiv \mapsto^{Keb} EB$
g) $EB \mid\equiv \mapsto^{K}_{KCC} KCC$
h) $EB \mid\equiv (KCC \Rightarrow \mapsto^{K} A)$
i) $EB \mid\equiv \#(T_b)$
j) $EB \mid\equiv A \leftrightarrow^{RN} EB$
k) $KCC \mid\equiv \mapsto^{Ka} A$
l) $KCC \mid\equiv \mapsto^{K}_{KCC} KCC$
m) $KCC \mid\equiv \mapsto^{Keb} EB$
n) $KCC \mid\equiv (KCC \Rightarrow \mapsto^{K} A)$
o) $KCC \mid\equiv (KCC \Rightarrow \mapsto^{K} EB)$
p) $KCC \mid\equiv \#(T_b)$
q) $KCC \mid\equiv KCC \leftrightarrow^{RN1} EB$
r) $KCC \mid\equiv \#(T_a)$
s) $A \mid\equiv (EB \Rightarrow A \leftrightarrow^{NR} EB)$
t) $EB \mid\equiv (KCC \Rightarrow KCC \leftrightarrow^{NR1} EB)$
u) $A \mid\equiv \#(T_b)$, $A \mid\equiv \#(T_{KCC})$
v) $B \mid\equiv \#(T_a)$, $B \mid\equiv \#(T_{KCC})$
w) $KCC \mid\equiv \#(T_{KCC})$
x) $KCC \mid\equiv (EB \Rightarrow KCC \leftrightarrow^{NR1} EB)$
y) $EB \mid\equiv (KCC \Rightarrow \mapsto^{K} KCC)$

Nota: Los certificados se asume que sean frescos.

Prueba

Mensaje 1:

$EB \lhd \{ \mapsto^{Ka} A\}K_{KCC}^{-1}, \{T_a, A \leftrightarrow^{NR} EB\}K_a^{-1}$

$EB \mid\equiv KCC \sim \mapsto^{Ka} A$ (g, message meaning)
$EB \mid\equiv \mapsto^{Ka} A$ (certificado fresco, h, jurisdiction)
$EB \mid\equiv A \sim (T_a , A \leftrightarrow^{NR} EB)$ (message meaning)
$EB \mid\equiv A \mid\equiv A \leftrightarrow^{NR} EB$ (v, nonce verification)
$EB \mid\equiv A \leftrightarrow^{NR} EB$ (t, jurisdiction)

Mensaje 2:

$KCC \lhd \{\mapsto^{Keb} EB\}K_{KCC}^{-1}, \{T_b , \{T_a , EB \leftrightarrow^{NR1} KCC\}K_{eb}^{-1}\}K_{KCC}$
$KCC \mid\equiv KCC \sim \mapsto^{Keb} EB$ (l, message meaning)
$KCC \mid\equiv \mapsto^{Keb} EB$ (certificado fresco, o, jurisdiction)
$KCC \lhd (T_b, \{T_a, KCC \leftrightarrow^{NR1} EB\}K_{eb}^{-1})$
$KCC \mid\equiv EB \sim (T_a, KCC \leftrightarrow^{NR1} EB)$ (message meaning)
$KCC \mid\equiv EB \mid\equiv KCC \leftrightarrow^{NR1} EB$ (w, nonce verification)
$KCC \mid\equiv KCC \leftrightarrow^{NR1} EB$ (x, jurisdiction)

Mensaje 3:

$EB \lhd \{\mapsto^{KAs} KCC\}K_{KCC}^{-1}, \{T_{KCC}, KCC \leftrightarrow^{As} EB\}K_{KCC}^{-1}$
$EB \mid\equiv KCC \sim \mapsto^{KAs} KCC$ (g, message meaning)
$EB \mid\equiv \mapsto^{KAs} KCC$ (certificado fresco, y, jurisdiction)
$EB \mid\equiv KCC \sim (T_{KCC} , KCC \leftrightarrow^{NR1} EB)$ (message meaning)
$EB \mid\equiv KCC \mid\equiv KCC \leftrightarrow^{NR1} EB$ (v, nonce verification)
$EB \mid\equiv KCC \leftrightarrow^{NR1} EB$ (t, jurisdiction)

Mensaje 4:

$A \lhd \{\mapsto^{Ks} EB\}K_{EB}^{-1}$
$A \mid\equiv KCC \sim \mapsto^{Ks} EB$ (b, message meaning)
$A \mid\equiv \mapsto^{Ks} EB$ (certificado fresco, c, jurisdiction)

Resultado : EB obtiene la llave publica de A. EB autentifica a A a traves de KCC. EB obtiene la parte de A de la llave de sesión compartida. KCC obtiene la llave publica de EB. KCC autentifica a EB y A. A obtiene la parte de EB de la llave de sesión compartida.

Conclusiones

Ks puede ser = NR ⊕ NR1

Por lo tanto,

$A \mid\equiv \mapsto^{Ks} EB$ A obtiene una llave de sesión compartida con EB.
$A \mid\equiv A \leftrightarrow^{Ks} EB$ Llave de sesión compartida
$EB \mid\equiv A \leftrightarrow^{Ks} B$ Llave de sesión compartida
$KCC \mid\equiv EB \mid\equiv KCC \leftrightarrow^{RN1} B$ KCC autentifica a B e indirectamente a A
$EB \mid\equiv A \mid\equiv A \leftrightarrow^{RN} B$ B autentifica a A

A.4.7 PROTOCOLO DE LLAVE PÚBLICA DE NEEDHAM-SCHROEDER

En un articulo, Needham y Schroeder propusieron un protocolo basado en criptografía de llave publica; el cual permitía que dos principales intercambiaran dos números secretos [18]. Una debilidad en el protocolo, es que permite un ataque de replay en las interacciones con la autoridad certificadora si se compromete una llave, como en el protocolo de Needham-Schroeder de llave compartida.
Aquí, S, tiene llave pública K_S, funciona solamente como autoridad de certificación entre A y B, cuyas llaves públicas son K_a, K_b, respectivamente; N_a, y N_b, son nonces. El intercambio del mensaje va como sigue:

Mensaje 1. A → S: A, B.
Mensaje 2. S → A: $\{K_b, B\}_{Ks}^{-1}$
Mensaje 3. A → B: $\{N_a, A\}K_b$
Mensaje 4. B → S: B, A.
Mensaje 5. S → B: $\{K_a, A\}K_s^{-1}$
Mensaje 6. B →A: $\{N_a, N_b\}K_a$
Mensaje 7. A →B: $\{N_b\}K_b$

El protocolo tiene dos componentes algo independientes pero interpolados. Se espera que, inicialmente, A y B mantengan la llave pública K_s de S. por lo tanto, los principales A y B pueden obtener las llaves públicas de cada uno de S. Los mensajes 1,2,4, y 5 logran este propósito. Mientras que en un segundo componente, en los mensajes 3, 6, y 7, A y B utilizan las llaves públicas obtenidas. Comunican los identificadores secretos N_a y N_b. Estos secretos se pueden utilizar más adelante, para los mensajes posteriores de firma. Por ejemplo, si B recibe un mensaje $\{X, N_a\}K_b$, entonces B puede deducir que A envió X.

El protocolo idealizado es como sigue:

Mensaje 2. S → A: $\{\{\mapsto^{Kb} B\}_{Ks}^{-1}$
Mensaje 3. A → B: $\{N_a\}K_b$
Mensaje 5. S → B: $\{\{\mapsto^{Ka} A\}K_s^{-1}$
Mensaje 6. B →A: $\{<A \leftrightarrows^{Nb} B>N_a\}K_a$
Mensaje 7. A →B: $\{<A \leftrightarrows^{Na} B>N_b\}K_b$

Los mensajes 1 y 4 se omiten deliberadamente, puesto que no contribuyen a las características lógicas del protocolo. Los mensajes 2 y 5 son directos, pero los otros requieren una cierta explicación. Es interesante observar la diferencia entre el mensaje 3 y los mensajes 6 y 7. En el mensaje 3, N_a, no es conocido para B y no se está utilizando así para probar la identidad de A; El mensaje 3 se utiliza simplemente para transportar N_a a B. En los mensajes 6 y 7, N_a y N_b son utilizados como secretos, así que la notación $<X>_Y$ es empleada. Estos mensajes también transportan creencia que no tienen ninguna representación en el protocolo concreto, porque los mensajes no serían enviados si la creencia no fue llevada a cabo. De hecho, apenas como en el caso de Kerberos, podríamos agregar a fondo más declaraciones de la creencia a los mensajes 6 y 7.

El protocolo analizado

Primero se asume el estado inicial de creencias asumida de los participantes:

$A \mid\equiv \mapsto^{Ka} A$
$B \mid\equiv \mapsto^{Kb} B$
$A \mid\equiv \mapsto^{Ks} S$
$B \mid\equiv \mapsto^{Ks} S$
$S \mid\equiv \{\mapsto^{Ka} A$
$S \mid\equiv \{\mapsto^{Kb} B$
$S \mid\equiv \{\mapsto^{Ks} S$
$A \mid\equiv (S \Rightarrow \mapsto^{K} B)$
$B \mid\equiv (S \Rightarrow \mapsto^{K} A)$
$A \mid\equiv \#(N_a)$
$B \mid\equiv \#(N_b)$
$A \mid\equiv A \leftrightarrows^{Na} B$
$B \mid\equiv A \leftrightarrows^{Nb} B$
$A \mid\equiv \#(\mapsto^{Kb} B)$
$B \mid\equiv \#(\mapsto^{Ka} A)$

Cada principal sabe la llave pública del agente de certificación S, así como sus propias llaves. Además, S sabe las llaves públicas de A y del principal B. Cada principal confía en el agente de certificación para firmar correctamente los certificados que dan la llave pública del otro. También, cada principal cree que el secreto que él genera está fresco. Las dos suposiciones pasadas están sorprendiendo y representan una debilidad en el protocolo. Cada principal debería asumir que el mensaje contiene la llave pública del otro principal está fresco. La dificultad podía ser resuelta agregando timestamps a los mensajes 2 y 5. Esto es análogo a la manera que los timestamps de Kerberos superan el problema con el protocolo de Needham-Schroeder de llave compartida.

Ahora obtenemos la creencia final:

$A \mid\equiv \mapsto^{Kb} B$
$B \mid\equiv \mapsto^{Ka} A$
$A \mid\equiv B \mid\equiv A \leftrightarrows^{Nb} B$
$B \mid\equiv A \mid\equiv A \leftrightarrows^{Na} B$

Cada principal sabe la llave pública del otro y tiene conocimiento de un secreto compartido que él cree que el otro aceptará como siendo compartido solamente por los dos principales. De este punto, A y B pueden continuar intercambiando mensajes usando N_a, N_b, y encripción de llave publica. De esta manera pueden transferir datos u otras llaves con seguridad.

A.4.8 PROTOCOLO X.509 del CCITT

El "draft" de recomendaciones del estándar CCITT X.509 contiene un sistema de tres protocolos usando entre uno y tres mensajes [19] (El cual ahora se ha convertido en una recomendación oficial del CCITT.) Los protocolos se piensan para firmar, comunicación segura entre dos principales, si se asume que cada uno sabe la llave pública del otro. La versión de tres-mensajes se da a continuación. Los protocolos de dos-mensajes y de uno-mensaje son formados quitando el ultimo o los dos últimos mensajes pasados, respectivamente.

El protocolo publicado contiene dos debilidades, cualquiera de las cuales se puede explotar por un intruso, como se demuestra abajo. Encontramos una de estas debilidades al momento de idealizar el protocolo y la otra durante el análisis subsiguiente.

Mensaje 1. A → B: A, $\{T_a, N_a, B, X_a, \{Y_a,\}K_b\}K_a^{-1}$
Mensaje 2. B →A: B, $\{T_b, N_b, A, N_a, X_b, \{Y_b\}K_a\}K_b^{-1}$
Mensaje 3. A →B: A, $\{N_b\}K_a^{-1}$.

Aquí, T_a, T_b son timestamps, N_a, y Nb son nonces, y X_a, Y_a, X_b, y Y_b son usuarios de datos. El protocolo asegura la integridad de X_a y X_b, asegurando el recipiente del origen, y garantiza la privacidad de Y_a, y Y_b.

Para la idealización del protocolo, simplemente se toma lo siguiente:

Mensaje 1. A → B: $\{T_a, N_a, X_a, \{Y_a,\}K_b\}K_a^{-1}$
Mensaje 2. B →A: $\{T_b, N_b, N_a, X_b, \{Y_b\}K_a\}K_b^{-1}$
Mensaje 3. A →B: $\{N_b\}K_a^{-1}$.

Como de costumbre, los timestamps T_a, y T_b se ven como nonces.

Análisis del protocolo

Se asume que cada principal sabe su propia llave secreta y la otra llave pública, y cree en su propio nonce y el timestamp de l otro sea fresco.

$A \mid\equiv \{\mapsto^{Ka} A\}$
$B \mid\equiv \{\mapsto^{Kb} B\}$
$A \mid\equiv \{\mapsto^{Kb} B\}$
$B \mid\equiv \{\mapsto^{Ka} A\}$
$A \mid\equiv \#(N_a)$
$B \mid\equiv \#(N_b)$
$A \mid\equiv \#(T_b)$
$B \mid\equiv \#(T_a)$

Ahora se puede derivar lo siguiente:

$A \mid\equiv B \mid\equiv X_b$
$B \mid\equiv A \mid\equiv X_a$

Esto representa un resultado más débil que los deseados por autores; en detalle, no se obtiene que: B |≡ A |≡ Y_a, o A |≡ B |≡Y_b. Aunque Y_a y Y_b cada uno se han transferido en un mensaje firmado, no hay evidencia para sugerir que el remitente está realmente enterado de los datos que él envió en la parte privada del mensaje. Esto corresponde a un panorama donde algunos ciertas terceras partes interceptan un mensaje y quitan la firma existente mientras que agregan las propias, ocultando copiando la sección encriptada dentro del mensaje firmado. El arreglo más simple es firmar los datos secretos Y_a y Y_b antes de que sé encripte para privacidad.

Una cierta redundancia es sensible en el segundo mensaje; T_b o N_a, es suficiente para asegurar la puntualidad del mensaje. La descripción del protocolo indica que la comprobación de T_b es opcional en la versión del protocolo de tres-mensajes. De hecho, es perfectamente razonable omitir T_b en conjunto, puesto que es redundante en ambos protocolos de dos y de tres-mensajes.

Desafortunadamente, el documento del CCITT X.509 también sugiere que T_a, no necesita ser comprobado en el protocolo de tres-mensajes. Esto es un problema serio porque la comprobación de T_a, es el único mecanismo que establece la puntualidad del primer mensaje. Lógicamente, si T_a, no se comprueba, no se puede realizar la verificación del nonce en el primer mensaje, y obtenemos solamente el resultado más débil B |≡ A a dicho X_a, en lugar de B |≡ A |≡ X_a.

Esta dificultad explica la intención del tercer mensaje, que es asegurar a B que A generó su primer mensaje recientemente. Los autores parece que esperaron que el uso de N_b sería suficiente con ligar el tercer mensaje al primero, ya que N_b liga los dos mensajes pasados y N_a liga los primeros dos mensajes. El error aquí es que N_b solo no liga los dos mensajes pasados, y ésto puede permitir que un intruso C reenvíe uno de los viejos mensajes de A y personificar después de eso a A.

El sigue intercambio concreto ilustra el defecto. El intruso primero entra en contacto con B

$C \rightarrow B$: A, $\{T_a, N_a, B, X_a, \{Y_a\}K_b\}K_a^{-1}$

Esto es un viejo mensaje enviado originalmente por A. Recordemos que B no está presumido para comprobar el timestamp T_a, en el protocolo de tres-mensajes y así que no descubrirá el reenvío del mensaje original de A. B contesta como si el mensaje viniera de A, y proporciona un nonce nuevo, N_b.

$B \rightarrow A$: B, $\{T_b, N_b, A, N_a, X_b, \{Y_b\}K_a\}K_b^{-1}$

En este punto C hace a A iniciar la autentificación con C, por cualesquiera medios.

$A \rightarrow C$: A, $\{T'_a, N'_a, C, X'_a, \{Y'_a\}K_c\}K_a^{-1}$

C contesta a A, proporcionando el nonce N_b. (el nonce N_b no es secreto, y nada evita que C utilice el mismo valor en un caso del protocolo entre A y C.)

$C \rightarrow A$: C, $\{T_c, N_b, A, N'_a, X_c, \{Y_c\}K_a\}K_c^{-1}$

A contesta a C, firmando el mensaje exacto necesitado para que C convenza a B que el primer mensaje fuera enviado recientemente por A y no sea un replay de un viejo mensaje, por lo tanto, ésto puede permitir que C personifique A.

$A \rightarrow C$: A, $\{N_b\}K_a^{-1}$.

Una solución es incluir el nombre de B en el ultimo mensaje. Puesto que B garantiza la unicidad de sus propios nonces, él puede estar seguro que este mensaje está ligado a esta instancia del protocolo. La versión idealizada del mensaje 3 podía entonces incluir cualquier creencia transmitida en el mensaje 1, asegurando a B de su puntualidad.

El protocolo X.509 utiliza realmente funciones hash para reducir la cantidad de encriptado: Para firmar un mensaje m, una función hash H(m) de m es calculada y firmada. Esto no se ha demostrado en la descripción de arriba. La lógica y el análisis de este protocolo particular son cambiados solo levemente por la introducción del hashing [3]

Tabla A.1. Resumen de resultados

	Needham Schroeder shared key	**Otway Rees**	**Kerberos**	**Yahalom**	**Aziz/ Diffie**	**GSM**	**CDPD**	**Beller-Chang-Yacobi**	**PAUM-RCP**	**Needham Schroeder Public key**	**CCITT X.509**
Caracteristica	Distribuir llave	Distribuir llave	Distribuir llave	Distribuir llave	Distribuir llave	Distribuir llave	Distribuir llave	Distribuir llave	Distribuir llave	Establecer secretos públicos	Transferir datos públicos
Llaves	Compartidas	Compartidas	Compartidas	Compartidas	Publicas	Compartidas	Compartidas	híbridas	Publicas	Públicas	Públicas
Uso de secretos				Si	Si	Si	Si	Si	Si	Si	
Nonces / Clocks	Nonces	Nonces	Clocks	Nonces	Nonces	Nonces	Nonces	Nonces	Ambos	Nonces	Ambos
Pruebas	A y B	B	A y B*	A y B	A y B	B	A	A y B	A y B*	A y B	A y B*
Presencia de redundancia	Si	Si	Si	Si	Si	Si	Si	Si	Si		Si

* La presencia de B es garantizada a A solamente si los pasos del protocolo óptimos son utilizados.

CONCLUSIONES

La literatura reciente ha acentuado la importancia de razonar sobre el conocimiento para poder entender el cómputo distribuido [20]. Además, ha habido algunas descripciones formales de los protocolos criptográficos [21, 22, 23]. Los protocolos aquí analizados nos sirven como fundamentos para nuestro análisis más específico de los protocolos de autentificación propuestos. En este apartado hemos descrito algunas lógicas para razonar sobre protocolos de autentificación y para tratar varios ejemplos. La Tabla A.1 enumera algunos de los protocolos estudiados con la lógica, incluyendo algunos discutidos anteriormente, y se resumen sus cualidades. Dicha tabla resume algunas características bien conocidas:

- la meta de cada protocolo
- el tipo de criptosistema utilizado (llave secreta o pública)
- si los secretos (con excepción de llaves) son utilizados y
- si la puntualidad del mensaje está garantizada con nonces o relojes sincronizados.

Además, se incluyen aspectos del formalismo que ayuda a ver:

- si el protocolo prueba la presencia de cada parte a la otra
- redundancia, y

- problemas de seguridad.

En la misma tabla 4.1 los principales implicados en los protocolos son A y B; el iniciador es A. Los ejemplos demuestran cómo una lógica simple puede capturar diferencias sutiles entre los protocolos. Para una variedad de protocolos, nos permite exhibir gradualmente cómo la creencia se aumenta hasta el punto de la autentificación mutua. Para otros protocolos, nos dirige en identificar errores y en sugerir correcciones

A.5 LÓGICA SIMPLE PARA DISEÑO DE PROTOCOLOS DE AUTENTIFICACIÓN

La lógica que se presenta aquí, en cierta medida pertenece a la familia de la lógica BAN [3]. La lógica BAN es la mejor conocida y más influyente lógica en la verificación de protocolos. Nos permite describir la creencia de las partes dignas de confianza involucradas en los protocolos de autentificación y la evolución de estas creencias como una consecuencia de la comunicación. Se ha aplicado con éxito para descubrir defectos en una variedad de protocolos de autentificación y también ha ayudado en el entendimiento de los conceptos básicos de la autentificación. La lógica BAN es simple (tiene alrededor de 15 reglas de inferencia) lo cuál pudo ser una de las razones de su popularidad. La necesidad de muchas suposiciones universales en el modelo, sin embargo, es una desventaja de menor importancia.
En la lógica BAN, se asume que los principales del sistema son dignos de confianza y no lanza los secretos (esto no se ha entendido siempre con su significado completo [24]). La aplicación de los esquemas de encripción son fuertes, es decir, los mensajes encriptados pueden ser desencriptados solamente con la llave apropiada, cada mensaje encriptado contiene suficiente redundancia para permitir al principal que desencripta verificar que esta utilizando la llave correcta, y los principales pueden reconocer e ignorar sus propios mensajes.

La lógica de BAN se ha extendido en muchas direcciones (por ejemplo, [7], [25]). Cada extensión esta basada en un modelo más general (que necesita menos suposiciones) que el modelo de la lógica original, pero como consecuencia, cada uno tiene una lógica más compleja. En la lógica GNY [9], por ejemplo, no se requiere asumir que los principales son confiables y la redundancia esta siempre explícitamente presente en los mensajes encriptados. La lógica GNY distingue entre lo que puede poseer un principal y lo que puede creer adentro. Esto nos permite expresar diferentes niveles de confianza y condiciones implícitas atrás de los pasos del protocolo. Sin embargo, la lógica GNY tiene mas de 40 reglas de inferencia.

La lógica utilizada en nuestro caso, preserva la simplicidad de la lógica BAN y adopta algunos conceptos de la lógica GNY.

A continuación describiremos esta lógica simple, aunque no por esto deja de ser formal. Esta lógica utiliza la noción de canales que son generalizaciones de enlaces de comunicación con varias propiedades de seguridad. La naturaleza abstracta de canales nos permite tratar el protocolo en un nivel más alto de abstracción que la mayoría de las lógicas conocidas para la autentificación, y por lo tanto podemos dirigir las propiedades funcionales de alto nivel del sistema, sin tener que involucrarse con los problemas de implementación actual. La mayor ventaja de esta lógica es que también es conveniente para el diseño de protocolos de autentificación como ya se había comentado.

Más adelante se dará un conjunto de reglas sintéticas que pueden ser utilizadas por los diseñadores de protocolos para construir un protocolo en un sentido sistemático.

A.5.1 EL MODELO

Se considera un sistema distribuido que sea un grupo de principales (usuarios, hosts y procesos) y canales. Los principales pueden interactuar con cualquier otro de acuerdo con las reglas de algunos protocolos predefinidos, a fin de lograr una tarea común (ejemplo para proporcionar un servicio.). Las interacciones son basadas en mensajes que son transportados vía los canales. Un canal es una abstracción de una facilidad de comunicación que tiene ciertas propiedades de acceso. El canal puede representar un enlace físico, tan bueno como una conexión lógica segura criptográficamente entre principales. Los canales son la principal abstracción que se considera en este apartado. Su naturaleza abstracta nos permite diseñar y analizar protocolos sin dirigirnos, al problema y la complejidad de la implementación.

Un canal es caracterizado por el grupo de lectores y el grupo de escritores (por ejemplo, el grupo de principales que pueden recibir mensajes vía el canal y el grupo de principales que pueden enviar mensajes vía el canal) El grupo de lectores y escritores del canal "C" son denotados por "r(C)" y "w (C)" respectivamente.

Como un ejemplo, se consideran 2 principales:

"A y B" que tienen un secreto "K" compartido y establecido entre ambos, además ambos conocen un algoritmo de encripción de llave simétrica (AE.). Dado el algoritmo AE y la llave K, A y B pueden seguramente intercambiar mensajes vía C. Estos mensajes son incomprensibles para otros principales en el sistema. Se dice, que existe un canal C entre A y B, para el cual r(C) = w(C) = {A,B} (esto es, solamente A y B pueden enviar y recibir mensajes vía C).

Para utilizar un canal, un principal necesita información acerca de cómo leer y/o escribir desde/hacia el canal. Esta información es denotada por C^r y C^w respectivamente. Se dice, que el principal P es un lector del canal C, denotado por $P \in r(C)$, si P posee C^r, y similarmente el principal P es un escritor de canal C, denotado por $P \in w(C)$, si P posee C^w.

En el ejemplo anterior $C^r = C^w$ = (K, AE) nos indicaría que el conocimiento del secreto compartido y el algoritmo de encripción es suficiente para que A y B puedan comunicarse en un sentido seguro.

Se asume que un principal siempre puede detectar la llegada de un mensaje en cualquier canal que este pueda leer. Para ser más precisos, se asume que si un mensaje llega en un canal legible, entonces el principal detecta la llegada y el principal esta disponible para determinar en cual canal el mensaje a llegado. Se hace esta consideración para mantener este modelo simple. Removiendo el problema del reconocimiento desde el alcance del modelo nos permite concentrarnos en las propiedades funcionales importantes del sistema. Por razones similares, también se asume que los principales pueden reconocer sus propios mensajes (si el mensaje llega en un canal que el principal puede leer) e ignorar a estos. Implementaciones reales de los protocolos pueden soportar estas características poniendo suficiente redundancia dentro de los mensajes y utilizando llaves de identificación.

A continuación veremos los tipos básicos de canales:

Canal público. C es un canal público si cualquiera en el sistema puede escribir y leer en este. Esto es, $r(C) = w(C) = \Omega$, donde Ω es el grupo de todos los principales.

Canal autentico. C es un canal autentico si cualquiera puede leerlo, pero solamente un principal P puede escribirlo. Esto es, $r(C) = \Omega$ y $w(C) = \{P\}$. Canales auténticos pueden ser implementados utilizando firmas digitales.

Canal confidencial. C es un canal confidencial si cualquiera puede escribir en este, pero solamente un principal P puede leerlo. Esto es $r(C) = \{P\}$ y $w(C) = \Omega$. Canales confidenciales pueden ser establecidos por encripción con la llave publica de P.

Canal certificado. C es un canal certificado si

Canal dedicado. C es un canal dedicado si un principal P puede leerlo y un principal Q puede escribirlo. Esto es $r(C) = \{P\}$ y $w(C) = \{Q\}$. Los canales dedicados pueden ser construidos por la combinación de propiedades del canal autentico con las propiedades del canal confidencial.

Canal de grupo cerrado. C es un canal de grupo cerrado si un grupo de principales pueden escribir en este y el mismo grupo de principales pueden leerlo. Esto es, $r(C) = w(C) = \beta$, donde β es un grupo de principales. Los canales de grupo cerrado pueden ser implementados por el uso de encripción de llave simétrica y distribuyendo la llave para un grupo de principales. Si el tamaño del grupo de principales es 2, entonces llamaremos al resultado de canal de grupo cerrado, canal secreto convencional.

Canal secreto convencional. C es un canal secreto convencional si este puede ser utilizado solamente por 2 principales P y Q. esto es $r(C) = w(C) = \{P,Q\}$.

Canal de grupo certificado y cerrado. C es un canal de grupo certificado y cerrado si un grupo dc principales pueden escribir en este y el mismo grupo de principales pueden leerlo. Esto es, $r(C) = w(C) = \delta$, donde δ es un grupo de principales. Los canales de grupo certificado y cerrado pueden ser implementados por el uso de certificados, canales auténticos y confidenciales y distribuyendo el certificado para un grupo de principales. Si el tamaño del grupo de principales es 2, entonces llamaremos al resultado de canal de grupo certificado y cerrado, canal secreto convencional certificado.

Canal secreto convencional certificado. C es un canal secreto convencional certificado si este puede ser utilizado solamente por 2 principales P^C y Q^C. Esto es $r(C) = w(C) = \{P^C,Q^C\}$. Donde P^C y Q^C son dos principales certificados.

A.5.2 LA LÓGICA

Esta es una lógica modal clasificada que nos permite razonar respecto a protocolos de autentificación. Esto es, modela el comportamiento de los principales y los canales en el sistema descrito en la sección previa. Este consiste de un lenguaje y un número pequeño de reglas de inferencia. El lenguaje es empleado para describir suposiciones y eventos tan buenos como las metas del protocolo. Las reglas de inferencia son usadas para derivar nuevos estados respecto al sistema.

Uno puede utilizar la lógica para analizar protocolos de autentificación existentes. Por esta razón, es conveniente que primero las descripciones del protocolo deban ser traducidas dentro de una descripción en esta notación. Esto normalmente significa que todas las operaciones de encripción y desencripción deben ser eliminadas de la descripción y deberán ser reemplazadas por los canales apropiados. Esto es también requerido para interpretar algunos mensajes y para reemplazar ciertas partes de un mensaje por formulas (las cuales sé definirán más adelante) cuando esto es necesario para capturar suposiciones implícitas de pasos del protocolo anteriores.

El objetivo del análisis es para construir una deducción de prueba, la cual es una derivación de los objetivos, desde las suposiciones y la descripción del protocolo formal. Se considera que el protocolo es correcto si tal deducción existe. La falta de una deducción de prueba significa que el protocolo no podría ser correcto. Aunque, en este caso, no podemos generar un escenario de ataque completo con esta lógica, el análisis puede revelar con amplia certeza la posible debilidad en el protocolo y así, poder construir un ataque después más fácilmente y enfocado directamente a las posibles vulnerabilidades.

La técnica de análisis es la misma que en [3] Se anota el protocolo con formulas lógicas. Se escriben formulas antes del primer mensaje (esto representa las suposiciones iniciales y después de cada mensaje. En esta lógica, un mensaje "m" que es recibido por Q en el canal C es representado por la formula lógica $Q \lhd C$ (m). No utilizamos la notación clásica $P \rightarrow Q$: m, por que cuando un principal recibe un mensaje este no necesariamente conoce de donde viene el mensaje. Sin embargo, ciertos canales podrían permitir al principal derivar esta información después. La notación de la recepción de mensaje también nos previene de estar atribuyendo el mismo significado del mensaje m para ambos el emisor y el receptor, asumiendo que la intención o propósito del emisor y la interpretación del receptor son la misma puede conducir a un sutil error en el análisis como se puntualiza en [26].

Se considera importante la interpretación del receptor. Después de recibir un mensaje, la formula que representa el mensaje comienza a ser valida. Desde esta reciente formula valida y desde las formulas validas derivadas previamente se pueden derivar nuevas formulas validas. En este sentido podemos seguir la evolución de varios estados respecto al sistema.

Como ya se ha venido mencionando uno puede también utilizar esta lógica en el proceso de diseño de protocolos de autentificación. Por lo cual, se han construido algunas reglas sintéticas que están basadas en las reglas de inferencia de la lógica. Uno puede utilizar estas reglas para generar la descripción formal del protocolo y las suposiciones de inicio desde las metas del protocolo. Debido al número pequeño de las reglas de inferencia se hace posible mantener el número de las pequeñas reglas sintéticas también pequeño.

Durante la síntesis, el diseñador debe escoger reglas sintéticas que concuerdan lo mejor con las metas deseadas y las suposiciones que ya hallan sido generadas. Es posible que para alcanzar una meta dada, diferentes reglas sintéticas puedan ser utilizadas. Esto puede dar lugar a diversos protocolos para las mismas metas y suposiciones. Aplicando una regla sintética, el diseñador obtiene nuevas metas a ser alcanzadas. Ciertas metas pueden ser alcanzadas simplemente, asumiendo que las declaraciones que las representan son verdaderas. Otras metas podrían requerir un mensaje de protocolo a ser enviado. Si no hay más metas a alcanzar, entonces la síntesis es terminada. Este proceso resulta en un numero de suposiciones y mensajes de protocolo, de lo cual la construcción de la descripción del protocolo formal es entonces directa.

Así, utilizando esta lógica, se puede describir y analizar los protocolos con un nivel de abstracción alta, sin batallar con los problemas de implementación. Si el protocolo trabaja en ese nivel, entonces se puede remplazar la construcción de alto nivel (los canales) por sus implementaciones. Separando la fase de diseño y la de implementación reduce significativamente la complejidad de la construcción de protocolos de autentificación.

A.5.2.1 EL LENGUAJE

De aquí en adelante se utilizaran las siguientes notaciones: P y Q extendidos sobre los principales.

C es un canal, X representa un mensaje, el cual puede ser dato, formula o ambos. C(X) denota el mensaje X en el canal C, ϕ será utilizado para denotar una formula.

Las formulas básicas son las siguientes:

P $\lhd$ C(X): P ver C(X). Alguien ha enviado un mensaje X vía el canal C y P puede observar este. Si P no puede leer el canal C, entonces P no puede reconocer y entender este mensaje (Esto es, P no puede determinar cual canal fue utilizado y que fue el mensaje).

P $\lhd$ X|C: P ver X vía C. P recibió el mensaje X vía el canal C. Esto es posible solamente si alguien ha enviado este mensaje, y P puede leer este canal.

P $\lhd$ X: P ver X.. Alguien ha enviado un mensaje conteniendo X vía un canal que P puede leer.

(X): X es fresco. X nunca había sido dicho antes de la corrida actual del protocolo. Esto es generalmente valido para los nonces.

P $|\sim$ X : P alguna vez dijo X. P en algún momento envió un mensaje que contiene X. Aunque no sabemos exactamente cuando el mensaje fue enviado.

P $||\sim$ X : P recientemente ha dicho X. Esto significa que P dijo X en la actual corrida del protocolo.

Si ϕ es una formula entonces lo siguiente es también una formula:

P $|\equiv \phi$: P Believes ϕ. P confía que ϕ es verdadero. Esto no significa que ϕ es realmente verdadero, pero P actúa como si lo fuera.

Nuevas formulas pueden ser derivadas utilizando los operadores lógicos convencionales desde la lógica proposicional. Si ϕ_1 y ϕ_2 son formulas, entonces las siguientes son también formulas:

$\phi_1 \wedge \phi_2$: ϕ_1 y ϕ_2

$\phi_1 \vee \phi_2$: ϕ_1 o ϕ_2

$\phi_1 \rightarrow \phi_2$: ϕ_1 implica ϕ_2

También utilizamos notaciones del conjunto de teoría. El significado de estas notaciones es directo en este lenguaje (Esto es, $P \in \lambda$, donde λ es un conjunto de principales, es una formula lo cual significa que el principal P es un elemento del conjunto λ.)

Podemos demostrar el poder expresivo de este lenguaje mostrando como ciertos aspectos de veracidad pueden ser descritos con las construcciones definidas anteriormente. En este lenguaje la creencia o convicción del principal P en la honestidad del principal Q es modelado por la siguiente formula:

$P \mid\equiv ((Q \mid\sim \phi) \rightarrow (Q \mid\equiv \phi))$

Esto significa que P cree que si Q ha dicho recientemente ϕ, entonces Q cree en ϕ. En otras palabras, P cree que "Q dice solamente que confía que ϕ es verdadero". P no necesariamente cree que lo que Q dice es verdadero, este solamente cree que Q cree esto. Sin embargo, es posible que P no comparta todas las creencias de Q. Otro aspecto de confianza, llamado competencia puede ser expresado por la siguiente formula:

$P \mid\equiv ((Q \mid\equiv \phi) \rightarrow \phi)$

Esto significa que P cree que si Q cree en ϕ, entonces ϕ es verdadero. Si combinamos estas 2 formulas, entonces obtenemos la siguiente:

$P \mid\equiv ((Q \mid\sim \phi) \rightarrow \phi)$

Esto significa que P cree que si Q ha dicho recientemente ϕ, entonces ϕ es verdadero. En resumen, P cree " lo que Q dice es verdadero". Esto es, P cree que Q es honesto y competente.

Podemos restringir el alcance de estas creencias dando más detalles acerca de ϕ. P puede creer, por ejemplo, que Q es un servidor de autentificación y este es competente solamente en el establecimiento de nuevos canales secretos entre principales. Podemos expresar esto por la siguiente formula.

$P \mid\equiv ((Q \mid\equiv (r(C) = w(C) = \{A, B\})) \rightarrow (r(C) = w(C) = \{A, B\}))$

Al mismo tiempo, P podría no creer, por ejemplo, que Q es competente en el reconocimiento de frescura, lo cual significa que no es capaz :

$P \mid\equiv ((Q \mid\equiv \#(X)) \rightarrow \#(X))$

A.5.2.2 REGLAS DE INFERENCIA

Reglas de visión:

(S1) Si un principal P recibe un mensaje X vía un canal C, y P puede leer este canal, entonces P reconoce que el mensaje ha llegado en C y P puede ver el mensaje.

$$\frac{P \lhd C(X),\ P \in r(C)}{P \mid\equiv (P \lhd X \mid C),\ P \lhd X}$$

(S2) Si un principal P ve un mensaje compuesto (X, Y), entonces este también ve las partes del mensaje (esto es, X y Y).

$$\frac{P \lhd (X,Y)}{P \lhd X,\ P \lhd Y}$$

Reglas de interpretación

(I1) Si un principal P cree que un canal C puede ser escrito solamente por un grupo de principales W, entonces P cree eso si recibe un mensaje vía C, entonces alguien del grupo W excepto el mismo P (W \ {P}) dijo X.

$$\frac{P \mid\equiv (w(C) = W)}{P \mid\equiv ((P \lhd X \mid C) \rightarrow \bigvee_{\forall Qi \in W \setminus \{P\}} (Q_i \mid\sim X))}$$

(I2) Si un principal P cree que otro principal Q ha dicho un mensaje compuesto (X, Y), entonces cree que Q ha dicho las partes del mensaje también (esto es, X y Y).

$$\frac{P \mid\equiv (Q \mid\sim (X, Y))}{P \mid\equiv (Q \mid\sim X),\ P \mid\equiv (Q \mid\sim Y)}$$

Reglas de frescura

(F1) Si un principal P cree que otro principal Q dijo un mensaje X y P también cree que X es fresco, entonces P cree que Q ha dicho recientemente X..

$$\frac{P \mid\equiv (Q \mid\sim X)\ ,\ P \mid\equiv \#(X)}{P \mid\equiv (Q \mid\mid\sim X)}$$

(F2) Si un principal P cree que una parte de un mensaje compuesto X es fresco, entonces este cree que todo el mensaje (X, Y) es fresco.

$$\frac{P \mid\equiv \#(X)}{P \mid\equiv \#(X,Y)}$$

Regla de racionalidad

(R1) Esta regla de racionalidad es el bien conocido axioma K de las lógicas modales [27] en forma de regla:
Si un principal P cree que ϕ_1 implica ϕ_2 y el principal cree que ϕ_1 es verdadero, entonces este cree que ϕ_2 es también verdadero.

$$\frac{P \mid\equiv (\phi_1 \rightarrow \phi_2),\ P \mid\equiv \phi_1}{P \mid\equiv \phi_2}$$

A.5.2.3 TEOREMAS

En lo siguiente, se dan 3 teoremas simples de la lógica que se utilizan en la construcción de las reglas sintéticas.

Teorema 1. Si un principal P recibe un mensaje X vía un canal C, entonces cree que el mensaje fue dicho por alguien, en quien el principal cree que sea posible que escriba el canal C, excepto sí mismo.

$$\frac{P \lhd C(X),\ P \in r(C),\ P \mid\equiv (w(C) = W)}{P \mid\equiv (V_{\forall Q_i \in W \setminus \{P\}} (Q_i \mid\sim X))}$$

Prueba. La prueba puede ser construida empleando la primera regla de visión (S1), la 1ª regla de interpretación (I1) y la regla de racionalidad (R1.).

Teorema 2. Si un principal P cree que otro principal Q es honesto y P cree que Q ha dicho recientemente ϕ, en donde ϕ es una formula, entonces P cree que Q cree en ϕ.

$$\frac{P \mid\equiv (Q \mid\!\sim \phi) \rightarrow (Q \mid\equiv \phi)),\ P \mid\equiv (Q \mid\!\sim \phi)}{P \mid\equiv (Q \mid\equiv \phi)}$$

Prueba. Esta es una instancia de substitución de la regla de racionalidad.

Teorema 3. Si un principal P cree que otro principal Q es honesto y competente, y P cree que Q ha dicho recientemente ϕ, donde ϕ es una formula, entonces P cree en ϕ.

$$\frac{P \mid\equiv ((Q \mid\!\sim \phi) \rightarrow \phi),\ P \mid\equiv (Q \mid\!\sim \phi)}{P \mid\equiv \phi}$$

Prueba. Esta es una instancia de substitución de la regla de racionalidad.

A.5.2.4 REGLAS SINTÉTICAS.

Invirtiendo las reglas de inferencia de la lógica y los teoremas derivados anteriormente se construyen las reglas sintéticas que pueden ser utilizados en un sentido sistemático cuando estemos diseñando protocolos. La construcción de las reglas sintéticas es directa, excepto (Sint4) y el primer par de (Sint5) que sé discutirán a mayor detalle más adelante. La lista de las reglas sintéticas dada aquí no es exhaustiva; uno puede derivar nuevas reglas desde nuevos teoremas. Para nuestros propósitos, es decir para mostrar como utilizar las reglas sintéticas para construcción de protocolos. La lista dada abajo es suficiente.

Las reglas sintéticas tienen la forma general siguiente:

G

$\hookrightarrow G_1$

$\hookrightarrow G_2$

$\hookrightarrow$ ....

Lo cual significa que a fin de alcanzar el objetivo G todas las nuevas metas dadas G_1, G_2,...tienen que ser alcanzadas. Una meta G puede tener la forma G′ / G′′, lo cual significa que también G′ o G′′ tienen que ser alcanzadas.

(Sint1) (Derivada de (S1)) Para reconocer que un mensaje X ha llegado vía un canal C, un principal P tiene que recibir C(X) y tiene que poder leer C.

$P \mid\equiv (P \lhd X \mid C)$

$\hookrightarrow P \lhd C(X)$

$\hookrightarrow P \in r(C)$

(Sint2) (Derivado de (S1) y (S2)) Para ver un mensaje X, un principal tiene que ver un mensaje (X, Y) que contiene a X o tiene que recibir X vía un canal C.

$P \lhd X$

$\hookrightarrow P \lhd (X, Y) / P \mid\equiv (P \lhd X \mid C)$

(Sint3) (Derivado de (I2)) Para creer que un principal Q dijo X, un principal P tiene que creer que Q dijo un mensaje (X, Y) que contiene X..

$P \mid\equiv (Q \mid\sim X)$

$\hookrightarrow P \mid\equiv (Q \mid\sim (X, Y))$

(Sint4) (derivado del teorema 1) Para creer que un principal Q dijo X, un principal P tiene que recibir X vía un canal C que pueda leer y que cree se puede escribir solamente por Q o P y Q. además, Q tiene que ver X.

$P \mid\equiv (Q \mid\sim X)$

$\hookrightarrow P \lhd C(X)$

$\hookrightarrow P \in r(C)$

$\hookrightarrow P \mid\equiv (w(C) = \{Q\}) /$
$P \mid\equiv (w(C) = \{P,Q\})$

$\hookrightarrow Q \lhd X$

(Sint5) (derivado del teorema 1 y (F1)) Para creer que un principal Q ha dicho recientemente X, lo siguiente es requerido:
Si X es una formula y P cree que Q es honesto. Esto es, $P \mid\equiv ((Q \mid\sim \phi) \rightarrow (Q \mid\equiv \phi)))$, entonces P tiene que recibir X vía un canal C que pueda leer y que cree se puede escribir solamente por Q, o P y Q. Además, P tiene que creer que X es fresco y Q tiene que creer en X.

$P \mid\equiv (Q \mid\sim X)$

$\hookrightarrow P \lhd C(X)$

$\hookrightarrow P \in r(C)$

$\hookrightarrow P \mid\equiv (w(C) = \{Q\}) /$
$P \mid\equiv (w(C) = \{P,Q\})$

$\hookrightarrow P \mid\equiv \#(X)$

$\hookrightarrow Q \mid\equiv X$

Por otra parte P tiene que creer que Q dijo X y X es fresco.

$P \mid\equiv (Q \mid\sim X)$

$\hookrightarrow P \mid\equiv (Q \mid\sim X)$

$\hookrightarrow P \mid\equiv \#(X)$

(Sint6) (Derivado de (F2)) Para creer que un mensaje X es fresco, un principal P tiene que creer que alguna parte X′ de X es fresca.

$P \mid\equiv \#(X)$

$\hookrightarrow P \mid\equiv \#(X')$

(Sint7) (derivado del teorema 2) Para creer que un principal Q cree en una formula ϕ, un principal P tiene que creer que Q recientemente ha dicho ϕ y que Q es honesto.

$P \mid\equiv (Q \mid\equiv \phi)$

$\hookrightarrow P \mid\equiv (Q \mid\sim \phi)$

$\hookrightarrow P \mid\equiv ((Q \mid\sim \phi) \rightarrow (Q \mid\equiv \phi))$

(Sint8) (Derivado de teorema 3) Para creer en una formula ϕ, un principal P tiene que creer que Q recientemente dijo ϕ y que un principal Q es honesto y competente.

$P \mid\equiv \phi$

$\hookrightarrow P \mid\equiv (Q \mid\sim \phi)$

$\hookrightarrow P \mid\equiv ((Q \mid\sim \phi) \rightarrow \phi)$

(Sint9) (Derivado de (R1)) para creer en una formula ϕ, un principal P tiene que creer en una formula ϕ' y la implicación $\phi' \rightarrow \phi$

$P \mid\equiv \phi$

$\hookrightarrow P \mid\equiv \phi'$

$\hookrightarrow P \mid\equiv (\phi' \rightarrow \phi)$

Las reglas (Sint4) y la primera parte de (Sint5) requieren alguna explicación. Comenzando por Sint4, la cual esta basada en la reversa del teorema 1. Los primeros 3 de los nuevos objetivos en (Syn4) siguen directamente del teorema 1. Requerimos el ultimo para asegurar que en el protocolo que es generado usando esta regla sintética, los principales dicen solamente mensajes que ellos ven (esto es, que ellos poseen). En [27] donde una suposición similar es tomada para derivar el grupo restrictivo ultimo de suposiciones para un objetivo y protocolo dado, esto es llamado "precondición requerida".Por razones similares, en la primer parte de (Sint5), requirió él ultimo de los objetivos nuevos, para asegurar que el protocolo que es generado utilizando esta regla sintética, no se impone o anula la suposición de honestidad, es decir, principales honestos dicen solamente lo que ellos creen.

A.5.3 DISEÑO Y ANÁLISIS DEL PROTOCOLO DE AUTENTIFICACIÓN DE USUARIOS MÓVILES PARA REDES DE COMUNICACIÓN PERSONAL.

El protocolo se divide básicamente en la inicialización tanto de la red para ambas partes involucradas en la comunicación, como los usuarios móviles (precálculos). Posteriormente viene una etapa de autentificación usuario a estación base (formación de usuario) y por ultimo la seguridad al nivel de usuario. Con fines de simplificar el análisis del protocolo, y poder confirmar su fortaleza y seguridad se platea a continuación analizar el diseño del protocolo de manera simplificada (idealizada) utilizando la lógica antes mencionada en la parte de formación de usuario, donde se puede observar que un principal A se autentifica con la estación base (EB) donde esta puede ser VLR a través de HLR, con la ayuda de una entidad certificadora (KCC).

1) A → EB: $\{idA, TE, \{S....)_{Kp}{}^{A}\}_{KpEB}$ o $K_{A\text{-}EB}$
2) EB → KCC: $\{idA, idd_R, TE, N_R, DSAs \{N_R, \{S...\}, idd_R\}\}K_{p\,KCC}$ o $K_{EB\text{-}KCC}$
3) KCC → EB: $\{\{CertAs (KCC, As, ASR, U)\}\ K_{pEB}$ o $K_{EB\text{-}KCC}$ Donde U es A en este caso.
4) EB → A: $\{\{CertAs (ASR, Ks,U)\}\ K_{p\,A}$ o $K_{A\text{-}EB}$

Pasando a la lógica simplificada

Cabe mencionar que en este análisis sé utilizan canales dedicados, por que estos contemplan tanto las firmas digitales como los criptosistemas de llave publica, y ambos son empleados en el protocolo a analizar.

EB $\triangleleft C_{A\text{-}EB}\{X,Y\}$
KCC $\triangleleft C_{EB\text{-}KCC}((X,Y), C_{A\text{-}KCC}\{X',Y'\})$
EB $\triangleleft C_{EB\text{-}KCC}\{Cert\}$
A $\triangleleft C_{A\text{-}EB}\{Cert'\}$

Suposiciones

(Sup1) $A \in r(C_{A\text{-}EB})$	A puede leer del canal A-EB
(Sup2) $A \in r(C_{A\text{-}KCC})$	A puede leer del canal A-EB
(Sup3) $KCC \in r(C_{A\text{-}KCC})$	A puede leer del canal A-KCC
(Sup4) $EB \in r(C_{A\text{-}EB})$	EB puede leer del canal A-EB
(Sup5) $EB \in r(C_{EB\text{-}KCC})$	EB puede leer del canal EB-KCC
(Sup6) $KCC \in r(C_{EB\text{-}KCC})$	KCC puede leer del canal EB-KCC
(Sup7) $A \mid\equiv (w(C_{A\text{-}EB}) = \{A, EB\})$	A cree que solamente A y BB pueden escribir en $C_{A\text{-}EB}$
(Sup8) $KCC \mid\equiv (w(C_{A\text{-}KCC}) = \{A, KCC\})$	KCC cree que solamente A y KCC pueden escribir en $C_{A\text{-}KCC}$
(Sup9) $A \mid\equiv (w(C_{A\text{-}KCC}) = \{A, KCC\})$	A cree que solamente A y KCC pueden escribir en $C_{A\text{-}KCC}$
(Sup10) $EB \mid\equiv (w(C_{A\text{-}EB}) = \{A, EB\})$	EB cree que solamente A y EB pueden escribir en $C_{A\text{-}EB}$
(Sup11) $EB \mid\equiv (w(C_{EB\text{-}KCC}) = \{EB, KCC\})$	EB cree que solamente EB y KCC pueden escribir en $C_{EB\text{-}KCC}$
(Sup12) $KCC \mid\equiv (w(C_{EB\text{-}KCC}) = \{KCC, EB\})$	KCC cree que solamente KCC y EB pueden escribir en $C_{EB\text{-}KCC}$
(Sup13) $A \mid\equiv ((KCC \mid\sim \phi) \rightarrow (KCC \mid\equiv \phi))$	A cree que KCC es honesto
(Sup14) $A \mid\equiv (KCC (EB \mid\sim (Cert\cdot)) \rightarrow (EB \mid\sim (Cert\cdot)))$	A cree que KCC es competente en decidir si EB dijo el mensaje (Cert′)
(Sup15) $EB \mid\equiv ((KCC \mid\sim \phi) \rightarrow (KCC \mid\equiv \phi))$	B cree que KCC es honesto
(Sup16) $EB \mid\equiv (KCC (A \mid\sim (X, Y)) \rightarrow (A \mid\sim (X, Y)))$	EB cree que KCC es competente en decidir si A dijo el mensaje (X, Y)
(Sup17) $KCC \mid\equiv \#(X, Y)$	KCC cree que (X, Y) es reciente
(Sup18) $EB \mid\equiv \#(X, Y)$	B cree que (X, Y) es reciente
(Sup19) $A \mid\equiv \#(Cert')$	A cree que (Cert′) es reciente

Utilizando reglas sintéticas

Si expresamos como objetivo de esta parte del protocolo de la siguiente forma:

A |≡ (EB ||~ (A, (Cert′)) Lo cual se puede interpretar como: A cree que EB recientemente ha dicho el mensaje (Cert′) para A. O bien, A recibió los certificados temporales y autorizados por la estación base remota abalados por KCC y HLR. Esto es, a ha sido autentificada

1) Utilizando la segunda parte de (sint5) tenemos:

A |≡ (EB ||~ (A, (Cert′))

↪A |≡ (EB |~ (A, (Cert′))

↪ A |≡ # (A, (Cert′))

2) Utilizando (sint6) tenemos:

A |≡ # (A, (Cert′))

↪ A |≡ # (Cert′)

Donde esta ultima se puede ver que es la suposición 19, por lo que esta meta es alcanzada.

3) Continuando con: A |≡ (EB |~ (A, (Cert′)), para este caso utilizamos (sint8)

A |≡ (EB |~ (A, (Cert′))

↪ A |≡ (KCC ||~ (EB |~ (A, (Cert′))))

↪ A |≡ (KCC ||~ (EB |~ (A, (Cert′))) → (EB |~ (A, (Cert′))))

Aquí se puede ver que el segundo nuevo objetivo se ha alcanzado por la suposición (sup13) y (sup14). Es decir, A cree que KCC es honesto y competente en decidir si EB dijo un mensaje compuesto.

4) Continuando con la meta A |≡ (KCC ||~ (EB |~ (A, (Cert′)))), utilizando la primer parte de (sint5):

A |≡ (KCC ||~ (EB |~ (A, (Cert′)))

↪ A ◁ $C_{A\text{-}KCC}$ (EB |~ (A, (Cert′)))

↪ A ∈ r($C_{A\text{-}KCC}$)

↪ A |≡ (w($C_{A\text{-}KCC}$) = {A, KCC})

↪ A |≡ #((EB |~ (A, (Cert′)))

↪ KCC |≡ (EB |~ (A, (Cert′)))

De lo anterior tenemos que el segundo y tercer nuevos objetivos son alcanzados por las suposiciones (2 y 9) respectivamente. El cuarto objetivo es alcanzado por (sint6) y la suposición 19. Mientras que el primer objetivo puede ser alcanzado, enviando la formula EB |~ (A, (Cert′)) vía el canal $C_{A\text{-}KCC}$

5) Y continuando con la ultima meta, tenemos:
KCC $|\equiv$ (EB $|\sim$ (A, (Cert'))), para lo cual podemos utilizar (sint4)

KCC $|\equiv$ (EB $|\sim$ (A, (Cert')))

$\hookrightarrow$ KCC $\lhd C_{EB\text{-}KCC}$ (A, (X,Y)), es decir KCC $\lhd C_{B\text{-}KCC}$ ((X,Y), $C_{A\text{-}KCC}${X',Y'})

$\hookrightarrow$ KCC $\in r(C_{EB\text{-}KCC})$

$\hookrightarrow$ KCC $|\equiv (w(C_{EB\text{-}KCC}) = \{EB, KCC\})$

$\hookrightarrow$ EB $\lhd$ (A, (X, Y))

En los objetivos anteriores podemos ver que la primera nueva meta es un mensaje del protocolo. Las siguientes 2 metas pueden ser alcanzadas por la (sup6) y (sup12) respectivamente. La ultima puede ser alcanzada enviando el mensaje (A, (X,Y)) a EB que también es uno de los mensajes del protocolo, por lo tanto el protocolo de esta sección quedaría como:

EB $\lhd$ (A, (X, Y))
KCC $\lhd C_{EB\text{-}KCC}$ (A, (X, Y))
A $\lhd C_{A\text{-}KCC}$ (EB $|\sim$ (A, (Cert')))

Este protocolo descrito en forma lógica, puede ser implementado en varias formas, aquí se da una implementación que es similar a la original.

Conclusiones

Presentamos una manera de asegurarse de estar creando un protocolo seguro al momento del diseño, apoyándonos de una lógica simple la cual puede ser utilizada, tanto para el proceso de diseño del protocolo de autentificación, como para comprobar su diseño propiamente. La lógica esta basada en la noción de canales que son de alguna manera las vistas abstractas de diversos tipos de facilidades de comunicaciones seguras. Los canales son bastante generales para describir muchos de los bien conocidos esquemas de comunicación protegidos criptográficamente. Debido a esta noción, esta lógica resulta simple y efectiva ya que solo cuenta con no más de 7 reglas de inferencia principales. Esto hace una lógica económica y clara mientras que logra capturar todos los aspectos importantes de los protocolos de autentificación.

ANEXO B

B.1 CRIPTOGRAFIA DE CURVAS ELÍPTICAS (CCE)

B.1.1. INTRODUCCIÓN

Todos los sistemas de llave pública conocidos en la actualidad basan su seguridad en la resolución de algún problema matemático que, por su gran magnitud, es casi imposible de resolver en la practica. Pero estos problemas matemáticos en los que se basan los métodos tradicionales de criptografía de llave pública son jóvenes. Aun cuando estos han sido estudiados por varios cientos de años por matemáticos como Euclides, Fermat o Euler, estos problemas se estudiaban para obtener otro tipo de resultados, como ecuaciones que revelen propiedades de los problemas.
Puesto que el estudio de estos problemas no estaba destinado para uso criptográfico, no se puede decir que tenemos tanta experiencia en las matemáticas subyacentes en que se basa la criptografía, como suele pasar de tanto en tanto, que se descubren nuevas propiedades o algoritmos para "romper" los criptosistemas en un tiempo menor al que se había calculado inicialmente. En efecto, a menudo los resultados que se requieren en criptografía son del tipo que cierta propiedad no vale o que ciertos algoritmos eficientes no se puedan desarrollar. Claro queda que este no era el objetivo de aquellos antiguos matemáticos, puesto que no solamente no existía la criptografía de llave pública, presentada al mundo en 1976 por Diffie y Hellman, sino que ni siquiera existían las computadoras, de modo que tratar de desarrollar un algoritmo que resuelva un problema donde los números involucrados eran de 400 cifras decimales, simplemente no era de interés.
En estos momentos, los métodos de llave pública más utilizados son el RSA, para encriptación y firma digital, el Diffie-Hellman para acuerdo de llaves, y el DSA para firmas digitales.

Debido a la aparición en los últimos años de métodos que resuelven el problema matemático en que se basan los algoritmos arriba mencionados en un tiempo menor al que se había pensado (en el actual estado del conocimiento de dichos problemas cuando éstos fueron primero concebidos como método criptográfico), se necesita agrandar el espacio de claves para "emparchar" dicho sistema. Como una opción, en 1985, Neil Koblitz y Victor Miller (independientemente) propusieron el Elliptic Curve Cryptosystem (CCE), o Criptosistema de Curva Elíptica (CCE), cuya seguridad descansa en el mismo problema que los métodos de Diffie-Hellman y DSA, pero en vez de usar números enteros como los símbolos del alfabeto del mensaje a encriptar (o firmar), usa puntos en un objeto matemático llamado Curva Elíptica. CCE puede ser usado tanto para encriptar como para firmar digitalmente. Hasta el momento, no se conoce ataque alguno cuyo tiempo de ejecución esperado sea sub exponencial para poder romper los CCE, esto hace que para obtener el mismo nivel de seguridad que brindan los otros sistemas, el espacio de claves de CCE sea mucho más pequeño, lo que lo hace una tecnología adecuada para utilizar en ambientes restringidos en recursos (memoria, costo, velocidad, ancho de banda, etc.)

En este anexo supondremos que el lector esta familiarizado con los conceptos de los sistemas de llave pública, confidencialidad, firmas digitales, espacios de llaves, clases de complejidad, teoría de números y álgebra abstracta, así como un cierto conocimiento de los sistemas más utilizados en la práctica como RSA, Diffie-Hellman o DSA.

Es otro tipo de criptografía de llave pública, el cual emplea curvas elípticas definidas en un campo finito. La diferencia que existe entre este sistema y RSA es el problema del cual basan su seguridad, mientras RSA razona de la siguiente manera: te doy el número 15 y te reta a encontrar los factores primos. Mientras que el problema del cual están basados los sistemas que usan curvas elípticas es el problema del logaritmo discreto elíptico, en este caso su razonamiento con números sería algo como: te doy el número 15 y el 3 y te reta a encontrar cuantas veces tienes que sumar el mismo 3 para obtener 15.

A continuación nos dedicaremos a explicar de manera general, nada formal un poco mas los puntos más importante de los **CCE**

1) Entenderemos como una curva elíptica a una ecuación de la forma siguiente:

$$y^2 + axy + by = x^3 + cx^2 + dx + e$$

Donde las constantes *a,b,c,d* y e pertenecen a cierto conjunto llamado campo **F,** que para propósitos de la criptografía o es un campo primo ($\mathbf{Z_p}$) o un campo de característica 2, o sea donde los elementos son n-adas de ceros y unos ($\mathbf{F_2^n}$)

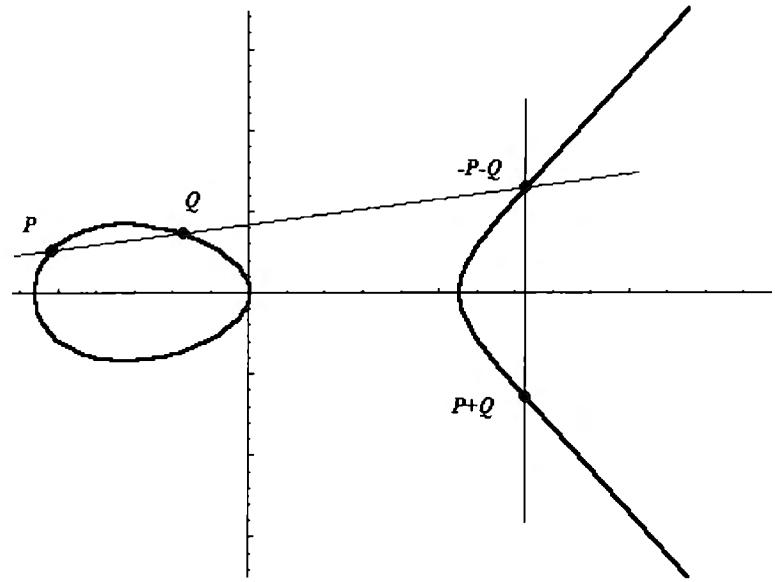
2) A un punto que satisface la ecuación anterior se le llama punto racional. Si el campo es finito, entonces el conjunto de puntos (x,y) que satisfacen la ecuación es finito y es llamado conjunto de puntos racionales de la curva **E** sobre el campo **F**. Al conjunto de puntos racionales lo podemos representar como

$$E : O, P_1, P_2, P_3, \ldots, P_n$$

E representa la ecuación y ***O*** es un punto que no tiene coordenadas y hace el papel de cero (llamado punto al infinito) ya que en este conjunto los puntos pueden sumarse y tiene las mismas propiedades que la suma de los números enteros, es decir lo que se conoce como un grupo abeliano.

Ejemplo: veamos una curva elíptica simple, si la ecuación es $y^2=x^3+4x+3$ y el campo $\mathbf{Z_5}$, es decir el conjunto {0,1,2,3,4}, entonces las parejas que satisfacen la ecuación son {(2,2), (2,3}, por lo tanto la curva elíptica es ***E***: {***O***, (2,2), (2.3)}. En este caso ***E*** tiene 3 puntos racionales.

3) La suma de estos puntos tiene una explicación geométrica simple, si la gráfica representa a todos los puntos que satisfacen la ecuación de la curva elíptica, y queremos sumar a *P* y *Q*, trazamos una línea recta que pase por *P* y *Q*, la ecuación de la curva es de grado 3 y la línea de grado 1, entonces existen siempre tres soluciones, en este caso la tercera solución esta dibujada como el punto *-P-Q*, enseguida se procede a dibujar una línea recta paralela al eje Y que pase por *-P-Q*, esta línea vertical también intercepta tres veces a la recta, todas las líneas verticales interceptan al punto especial llamado infinito y que geométricamente esta en el horizonte del plano, el tercer punto es por definición *P+Q*, como se muestra en la figura



2) No es difícil obtener fórmulas para calcular las coordenadas del punto $P+Q$ a partir de las coordenadas del punto P y del punto Q. Por ejemplo si el campo de definición de la curva es un campo primo $\mathbf{Z_p}$, entonces las fórmulas de suma son las siguientes

$$x_3 = \lambda^2 - x_1 - x_2$$
$$y_3 = \lambda(x_1 - x_3) - y_1$$
$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1}, & P \neq Q \\ \dfrac{3x_1^2 + a}{2y_1}, & P = Q \end{cases}$$

3) La anterior forma de sumar puntos de una curva elíptica es un poco extraña sin embargo, es esta extrañeza lo que permita que sea un poco más difícil romper los CCE. En el área de las matemáticas conocida como teoría de grupos se sabe que estos grupos son muy simples llamados grupos abelianos finitos lo que permite también que los CCE sean fácil de implementar, llamaremos al número de puntos racionales de la curva como el orden de la curva. En nuestro ejemplo $P_0=O$, $P_1=(2,2)$, $P_2=(2,3)$, donde $2P_1=P_2$.

4) Los CCE basan su seguridad en el Problema del Logaritmo Discreto Elíptico (PLDE), esto quiere decir que dados P,Q puntos de la curva hay que encontrar un número entero x tal que $xP = Q$ ($xP = P+P+\ldots+P$, x veces). Obsérvese que a diferencia del PFE (Problema de Factorización Entera) el PLDE no maneja completamente números, lo que hace más complicado su solución.

5) La creación de un protocolo con criptografía de curvas elípticas requiere fundamentalmente una alta seguridad y una buena implementación, para el primer punto se requiere que la elección de la curva sea adecuada, principalmente que sea no-supersingular y que el orden del

grupo de puntos racionales tenga un factor primo de al menos 163 bits, además de que este orden no divida al orden de un número adecuado de extensiones del campo finito, para que no pueda ser sumergido en él, si el campo es Z_P, se pide que la curva no sea anómala o sea que no tenga p puntos racionales. Todo esto con el fin de evitar los ataques conocidos.
Para el caso de la implementación hay que contar con buenos programas que realicen la aritmética del campo finito, además de buenos algoritmos que sumen puntos racionales, tanto en el caso de Z_p como F_2^n, en este último se toma una base polinomial que tenga el mínimo de términos por ejemplo un trinomio para generar los elementos del campo finito esto si la implementación es en software, y se toma una base normal si es en hardware. Además de contemplar que las operaciones de puntos racionales pueden hacerse en el espacio proyectivo, esto elimina el hacer divisiones, ahorrando tiempo.

6) Lo anterior se ve reflejado en las ventajas que ofrecen los CCE en comparación con RSA, la principal es la longitud de la llave secreta. Se puede mostrar que mientras en RSA se tiene que usar una llave de 1024 para ofrecer una considerable seguridad, los CCE solo usan 163 bits para ofrecer la misma seguridad, así también las llaves RSA de 2048 son equivalentes en seguridad a 210 de CCE. Esto se debe a que para resolver el PLDE el único algoritmo conocido toma tiempo de ejecución totalmente exponencial, mientras que el algoritmo que resuelve PFE incluso también el PLD en Z_p toma un tiempo subexponencial.

7) Otra buena noticia sobre CCE es que los elementos de los puntos racionales pueden ser elementos de un campo finito de característica 2, es decir pueden ser arreglos de ceros y unos de longitud finita (01001101110010010111), en este caso es posible construir una aritmética que optimice la rapidez y construir un circuito especial para esa aritmética, a esto se le conoce como Base Normal Optima.

8) Lo anterior permite con mucho que los CCE sean idóneos para ser implementados en donde el poder de cómputo y el espacio del circuito sea reducido, donde sea requerida una alta velocidad de procesamiento o grandes volúmenes de transacciones, donde el espacio de almacenamiento, la memoria o el ancho de banda sea limitado. Lo que permite su uso en Smart Cards, Teléfonos celulares, Fax, Organizadores de Palma, PCs, etcétera.

9) En la actualidad existen varios estándares que permiten el uso adecuado y óptimo de los CCE, entre los cuales se encuentran: **IEEE P1363** (Institute of Electrical and Electronics Engineers), el **ANSI X9.62**, **ANSI X9.63**, **ANSI TG-17**, **ANSI X12** (American National Standards Institute), **UN/EDIFACT**, **ISO/IEC 14888**, **ISO/IEC 9796-4**, **ISO/IEC 14946** (International Standards Organization), **ATM** Forum (Asynchronous Transport Mode), **WAP** (Wireless Application Protocol). En comercio electrónico: **FSTC** (Financial Services Technology Consortion), **OTP 0.9** (Open Trading Protocol), **SET** (Secure Electronic Transactions). En internet **IETF** (The Internet Engineering Task Force), **IPSec** (Internet Protocol Security Protocol)

Los CCE son el mejor candidato para reemplazar a las aplicaciones que tienen implementado RSA, estas definen también esquemas de firma digital, Intercambio de llaves simétricas y otros.

ANEXO C

C.1 EL PRINCIPIO BÁSICO DEL CELULAR.

Durante mucho tiempo las personas soñaron con poderse comunicar con cualquier otra persona en cualquier parte donde esta se encontrara, sin la necesidad de contar con una instalación telefónica o con el trabajo de encontrar una caseta telefónica. Esto comenzó a hacerse realidad tan pronto como la tecnología avanzaba y se comenzaron a utilizar las ondas de radio para lograr una comunicación, a fines del siglo XIX.

Actualmente las transmisiones inalámbricas constituyen una eficaz y poderosa herramienta que permite la transferencia de voz, datos y video, sin la necesidad de utilizar cables para establecer la conexión.

Esta transferencia de información es lograda a través de la emisión de ondas de radio, permitiendo así tener dos grandes ventajas las cuales son la movilidad y flexibilidad del sistema en general.

En este anexo se destaca como se ha dado la evolución de los sistemas inalámbricos desde sus comienzos, sus precursores, y muestra además cual es la tendencia actual y los desarrollos llevados a cabo hasta el momento. Por otra parte se menciona los aspectos principales del sistema radio celular y como están conformados, para poder entrar en contexto.

C.1.1 RESEÑA HISTÓRICA DE LAS COMUNICACIONES PERSONALES INALÁMBRICAS

El primer sistema de telefonía móvil (Mobile Telephone Service, St. Louis, 1946) utilizó una estación base sencilla, la cual cubría el área de servicio entera. En este sistema el numero de conexiones simultaneas es limitada por el numero de canales de radio disponibles.

Si revisamos la historia, encontramos que las comunicaciones inalámbricas comenzaron con:

- La postulación de las ondas electromagnéticas por James Cleck Maxwell durante el año de 1860 en Inglaterra.
- La demostración de la existencia de estas ondas por Heinrich Rudolf Hertz en 1880 en Inglaterra.
- La invención del telégrafo inalámbrico por Guglielmo Marconi.

Durante 1890 eminentes científicos como Jagdish Chandra Bosé de la India, Oliver Lodge en Inglaterra y Augusto Righi de la Universidad de Bologna, se encargaron del estudio de los fundamentos naturales de las ondas electromagnéticas.

La noción de la transmisión de información sin el uso de cables fue vista por nuestros ancestros como algo mágico.

En 1896 la primera patente de comunicaciones inalámbricas fue concedida a Guglielmo Marconi en el Reino Unido. Desde aquel momento, entonces el número de desarrollos en el campo de las comunicaciones inalámbricas tomó ese sitio. Como se puede ver en la tabla C.1[2].

En 1980 comienza la era celular. Diferentes desarrollos y nuevas tecnologías tomaron lugar durante los años de 1990 al 2000.

Tabla C.1 La Era Inalámbrica

ERA PIONERA
• 1860, Postulación de las ondas EM por James Maxwell • 1880, Demostración de la existencia de las ondas por Henry Rudolf Hertz. • 1890, Primera patente de los sistemas inalámbricos por Guglielmo Marconi. • 1905, Primera transmisión de voz y música vía enlace inalámbrico por Reginald Fessenden • 1912, Hundimiento del Titanic destacando la importancia de la comunicación inalámbrica sobre las vías marítimas, en los años siguientes la marina comenzó a establecer los radios de telegrafía.
ERA PRE-CELULAR
• 1921, El Dpto. de la Policía de Detroit dirige maniobras militares con radios móviles. • 1933, En EEUU, existen 4 canales en los 30-40 Mhz. • 1938, En EEUU, se reglamenta el servicio regular. • 1946,Primer comercio de los sistemas de teléfonos móviles operados por el sistema Bell, en EEUU. • 1948, Primer comercio plenamente automático de teléfonos móviles en EEUU. • 1950, Los teléfonos y los enlaces de microondas son desarrollados. • 1960, Introducción de líneas interurbanas a los sistemas de radio con canales automáticos en EEUU. • 1970, Los sistemas de teléfonos móviles operan en muchas ciudades. Lo utilizaban varios millones de vehículos.
ERA CELULAR
• 1980, Distribución de los sistemas celulares analógicos por el mundo • 1990, Distribución de los celulares digitales y modo de operación dual de los sistemas digitales. • 2000, Distribución de los servicios multimedia a través de FPLMTS, IMT-2000, UMTS • 2010, Ancho de banda para Comunicación inalámbrica que soporten redes B-ISDN y ATM • 2010+, Radio sobre fibra (así como microceldas sobre fibra óptica)

[2] Esta tabla solo contiene comunicaciones inalámbricas en términos de tecnologías de radio.

El nombre completo de este sistema se conoce como Sistema de Radio Telefonía Móvil en Tecnología Celular, el cual describe sus tres características más importantes como:

Radiotelefonía. Telefonía a través de ondas de radio.
Móvil. Capacidad para dar servicio a teléfonos en movimiento inclusive a altas velocidades.
Tecnología celular. Técnica que permite reutilizar un número limitado de frecuencias para aumentar "ilimitadamente" la capacidad del sistema, mediante el uso de células.

C.1.2 SISTEMA DE RADIOTELEFONÍA

Radiotelefonía. El desarrollo de radiotransmisores y receptores de muy alta frecuencia después de la segunda guerra mundial propició la aparición del servicio de radiotelefonía. Sus primeras aplicaciones fueron en la policía, bomberos y taxis. Posteriormente, se utilizó como verdadero *radioteléfono*, conectado a la RTPC (Red Telefónica Publica Conmutada), para la recepción y generación de llamadas telefónicas ordinarias.

La tecnología permitió el empleo de una antena de radio, sobre una montaña, asociada a un potente transceptor de radio multicanal.

Para llamadas que se hacen hacia o desde el usuario de radiotelefonía, la RTPC se conecta vía Centros de Conmutación Móviles (CCM), con varias estaciones base de transmisión que emiten y reciben señales de radio desde los teléfonos móviles. Para conectar cada teléfono durante la conversación se necesitan dos canales de radio, uno para cada dirección de conversación. La figura C.1, muestra un CCM único y una sola estación base de transmisión, que atienden a 3600 usuarios móviles dentro de un área de 1200 Km. cuadrados, con un radio de aproximadamente 20 Km. desde la estación base.

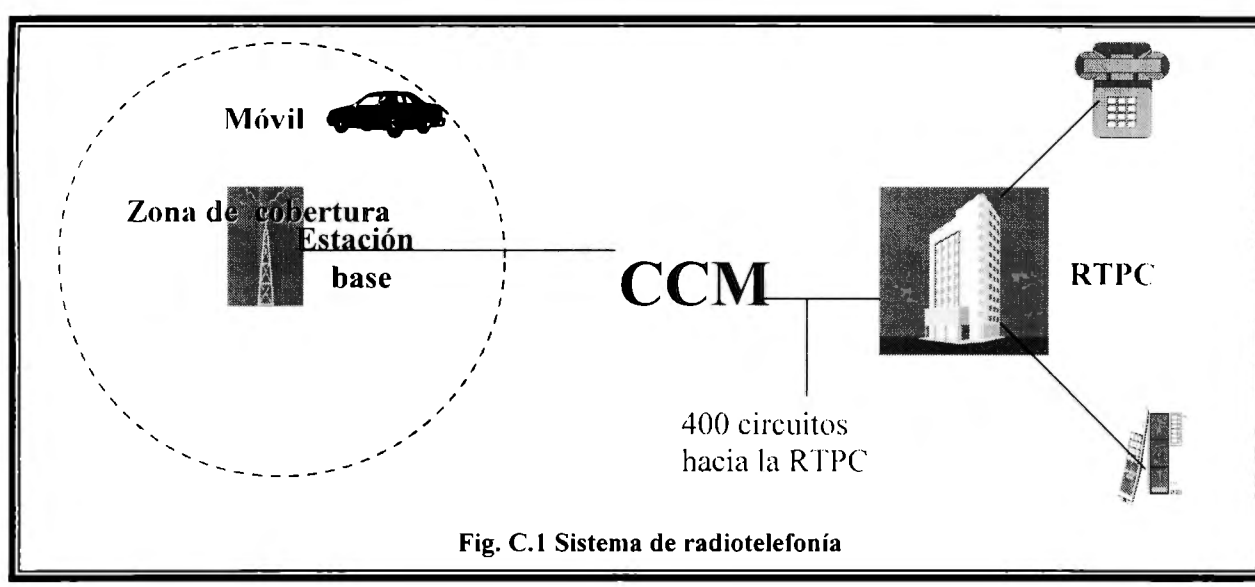


Fig. C.1 Sistema de radiotelefonía

Las llamadas se hacen o se reciben en la estación telefónica móvil, monitoreando un canal particular de radio, llamado canal de señalización o de llamada. Si se inicia una llamada, la

estación móvil envía un mensaje a la estación base de manera muy semejante a como un teléfono común envía la señal de descuelgue y el tren de dígitos de destino. Enseguida, la estación base asigna un par de canales de radio libres para emplearse entre la estación base y el móvil durante el período de conversación subsiguiente y los dos se conmutan sobre los canales asignados simultáneamente. Al final de la llamada, los canales de radio se liberan para emplearse en otra llamada.

Las llamadas entrantes al radioteléfono se conectan de la misma manera. El usuario común de la RTPC marca un número telefónico ordinario. Generalmente, se asigna un código de área particular dentro del plan normal de numeración telefónica para identificar la red móvil y todas las llamadas con este código de área sé enrutan hacia el CCM que vía la estación base apropiada de transmisión, se envían al receptor móvil deseado.

El problema a superar en el diseño de sistemas de radiotelefonía, es la necesidad de enrutar las llamadas entrantes sólo hacia la estación base más cercana.

La insatisfacción de este requisito provoca la falla de llamadas entrantes, pues el móvil no puede garantizar su estancia dentro del área de cobertura de una estación base particular de transmisión.

Los primeros sistemas de radiotelefonía resolvieron el problema pidiendo que el solicitante supiera el paradero del móvil al que quería llamar. El mismo número de usuario se usaría en todas las llamadas entrantes, pero se tenía que emplear un código de área diferente, dependiendo de la estación base sobre la cual el solicitante deseaba que su llamada se enrutara (es decir, de acuerdo con la zona en la que el solicitante creía que se encontraba el móvil). Si el número de usuario es "12345", entonces cuando el solicitante cree que el móvil estaba en la zona 1, se debería marcar el número 0331 12345, de igual manera, para las zonas 2 y 3, los números 0332 12345 y 0333 12345 respectivamente.

Se desarrollaron algunos sistemas más avanzados de radiotelefonía con la habilidad de recordar la ultima ubicación a partir de la cual se hizo una llamada y enrutarla hacia ese punto. Otros emplearon los métodos de prueba y error, sin embargo, estos sistemas fueron reemplazados paulatinamente por nueva tecnología.
Los sistemas automáticos no sólo se basaron en el hecho de que el solicitante supiera dónde estaba el móvil, sino que algunos requerían la acción por parte del usuario del móvil de presionar para hablar.

El sistema permitía que el usuario se moviera durante la llamada, pero si se salía del área de cobertura, la llamada se perdía. No había facilidades de transferencia a otras estaciones base durante el curso de la llamada. Estos inconvenientes, provocaron la muerte de la radiotelefonía en su concepción original y su reemplazo por la telefonía celular.

C.1. 3 SISTEMAS CELULARES

La dificultad con los sistemas anteriores de radiotelefonía, fue la pequeña capacidad en el número de usuarios. Esto se debió a que fueron diseñados para tener una cobertura de área muy grande, pero con un número limitado de canales de radio disponibles dentro de cada zona. El reempleo de canales de radio en otras zonas se prohibía por el riesgo de interferencia, excepto donde las estaciones base estaban separadas por grandes distancias. Debido a la gran demanda de

radiotelefonía y también, que la disponibilidad de canales de radio era y aún es un recurso limitado, se desarrollaron sistemas que hicieran uso más eficiente del espectro de radiofrecuencias, incrementando sustancialmente, la capacidad disponible de usuarios.
El sistema que surgió se conoce como radio celular, cuya principal ventaja sobre otros sistemas móviles es su capacidad para manejar mayores cargas de tráfico.
Pero primero definamos que es una célula.

Célula. Es la unidad básica de cobertura en que se divide un sistema celular, o bien es el área de cobertura de una estación base. Las células se clasifican según su tamaño en: (ver tabla C.2)

Tabla C.2. Clasificación de las células según su tamaño

TIPO	RADIO	COBERTURA
Célula gigante	> 50 Km	Continentes
Célula grande	10-50 Km	Rural
Célula	1-10 Km	Urbana
Mini-célula	100-1000 m	Urbana densa
Micro-célula	10-100 m	Oficina/Campus
Pico-célula	2-10 m	Habitación
Femto-célula	< 2 m	Privado

Las redes de radio celular hacen uso eficiente del espectro, re-utilizando las mismas frecuencias de canal de radio en un gran número de células. Estructuradas en forma de panal de abejas, cada célula se mantiene pequeña para que la potencia de transmisión que se requiere en la estación base se pueda mantener baja. Esto limita el área sobre la cual, la señal de radio es efectiva, reduciendo así el área sobre la cual puede ocurrir la interferencia de señales. Fuera de la zona de interferencia del transmisor, es decir, en una célula no adyacente, a suficiente distancia de la primera, se pueden volver a emplear las mismas frecuencias de canal de radio. La figura C.2 ilustra la zona de interferencia de una estación base de célula, junto con otra célula que emplea las mismas frecuencias de canal de radio.

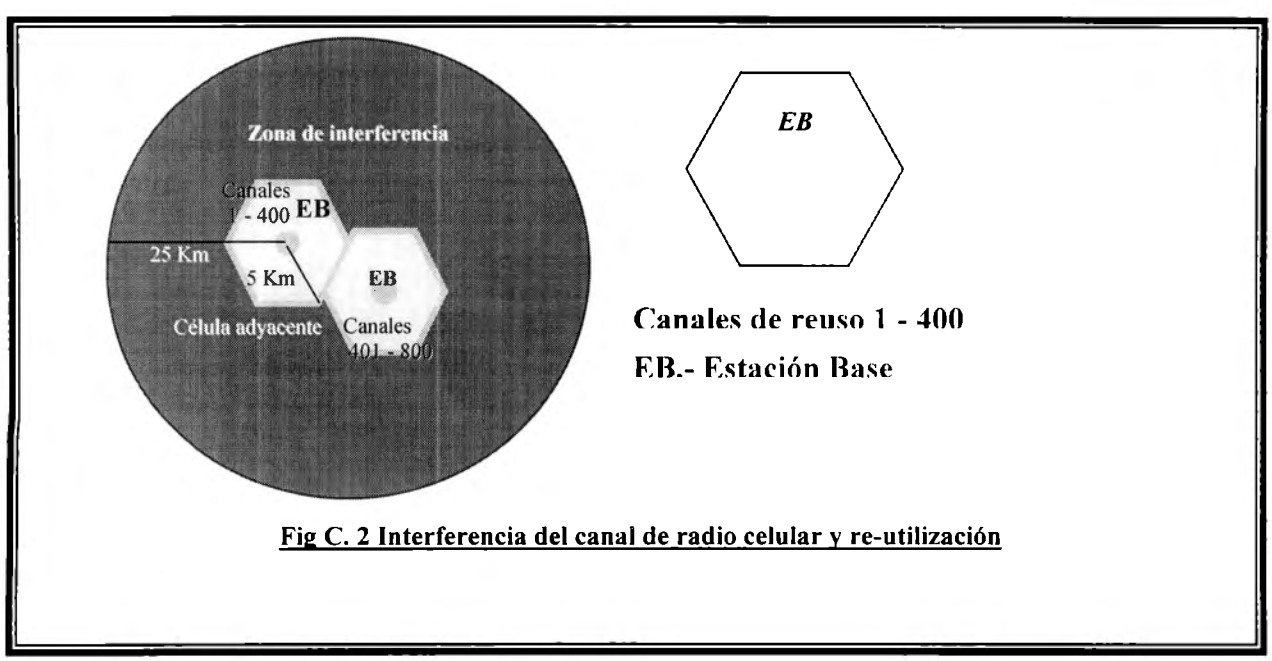


Fig C. 2 Interferencia del canal de radio celular y re-utilización

Las células que conforman el sistema se encuentran agrupadas en "clusters"[3]. El número de células en un cluster tiene que ser determinado de manera que pueda repetirse de forma no interrumpida en el área de cobertura. Los clusters típicos se basan en 4, 7, 12 o 21 células, a continuación se presentan los clusters mencionados (ver figura C.3).

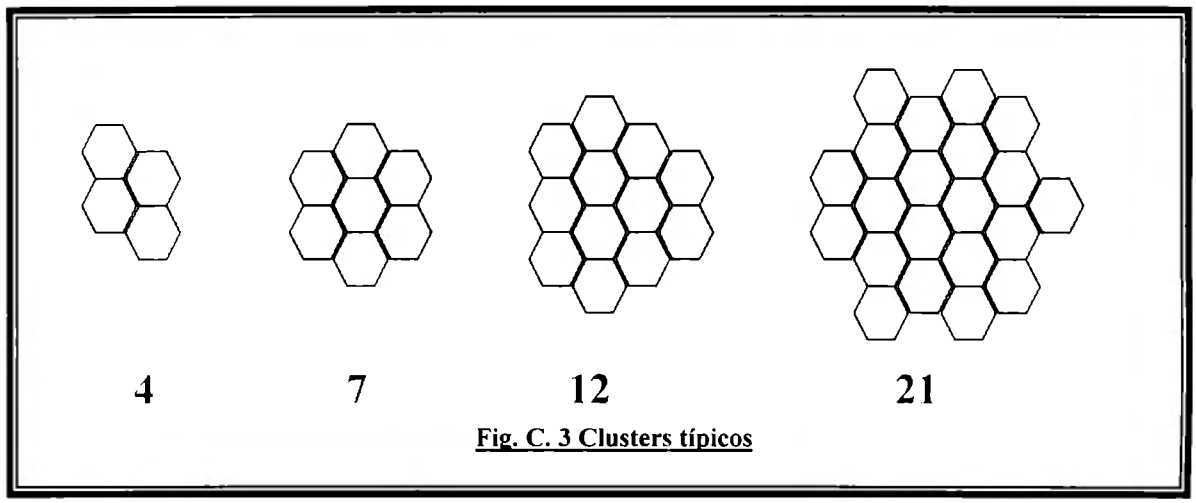


Fig. C. 3 Clusters típicos

Se puede establecer un panal completo de células, reutilizando canales de radio entre las diferentes células de acuerdo a un plan determinado. En la figura C.4 se muestra un patrón de reutilización de 7 células. Siete esquemas diferentes de frecuencias de canal de radio se repiten sobre cada grupo de células hexagonales, cada célula emplea un conjunto diferente de frecuencias. Con tal planeación, la misma frecuencia de radio se puede emplear para diferentes conversaciones dos o tres células más adelante.

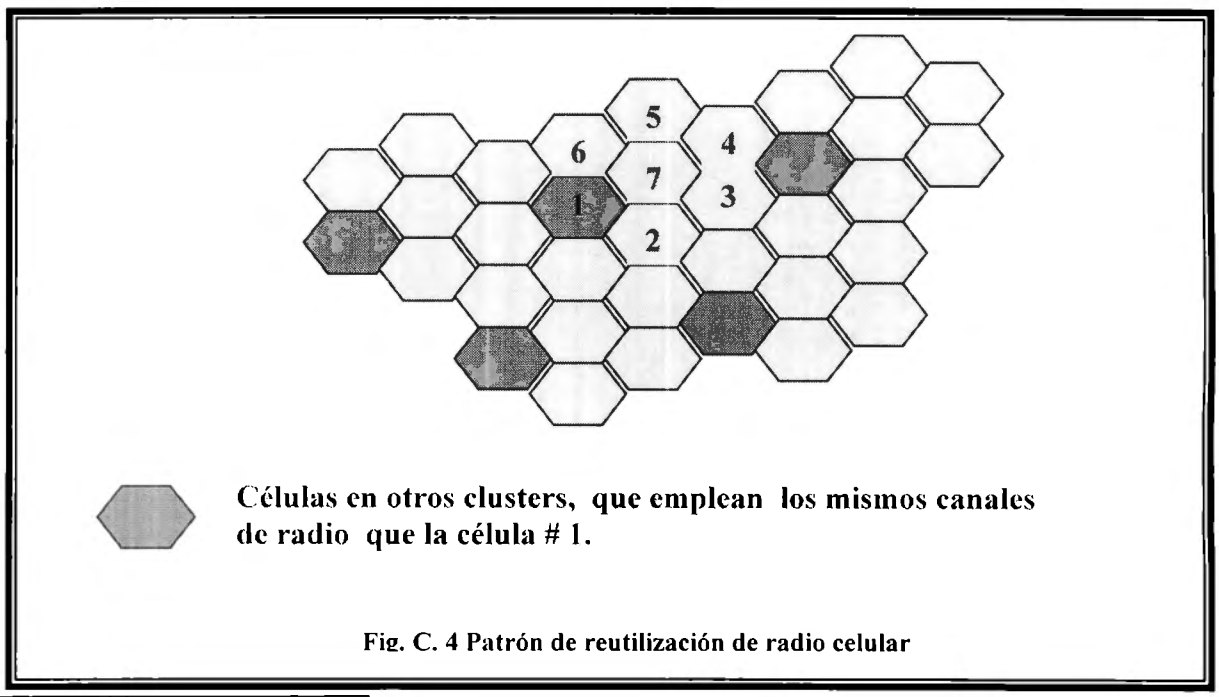


Fig. C. 4 Patrón de reutilización de radio celular

[3] También conocido como cluster, racimo o agrupamiento, el cuál es un grupo de células dentro de estos, los cuales no son reutilizados. Sumando varios clusters es como se alcanza la cobertura final del sistema celular, reutilizando las mismas frecuencias de todos los clusters.

El patrón de reutilización de la figura anterior es de 7 células, pero pueden emplearse otros patrones conocidos. Se necesitan grandes patrones de repetición para satisfacer la demanda de alto tráfico en áreas donde las células pequeñas no adyacentes, aún pueden interferirse entre sí.
Para que un sistema celular funcione correctamente, es necesario considerar que si se utilizan más canales por célula y el tamaño del cluster es menor (menos células), la distancia entre las células que utilizan los mismos canales es menor, teniendo como consecuencia que la interferencia entre clusters adyacentes aumenta (interferencia cocanal.).
El número total de canales por célula, depende del número de canales disponibles y del tipo de cluster, por lo tanto: **# de canales por célula = # total de canales / cluster**

Donde el cluster puede tener 4, 7, 12, 21,....células.
La característica de las redes de radio celular es su habilidad de acoplarse al nivel creciente de demanda, primero usando más canales de radio y después reduciendo el tamaño de las células.
Por otra parte, reducir el tamaño de la célula tiene como efecto la reducción del número probable de unidades móviles que se encuentren en ella a la vez, disminuyendo así la congestión.

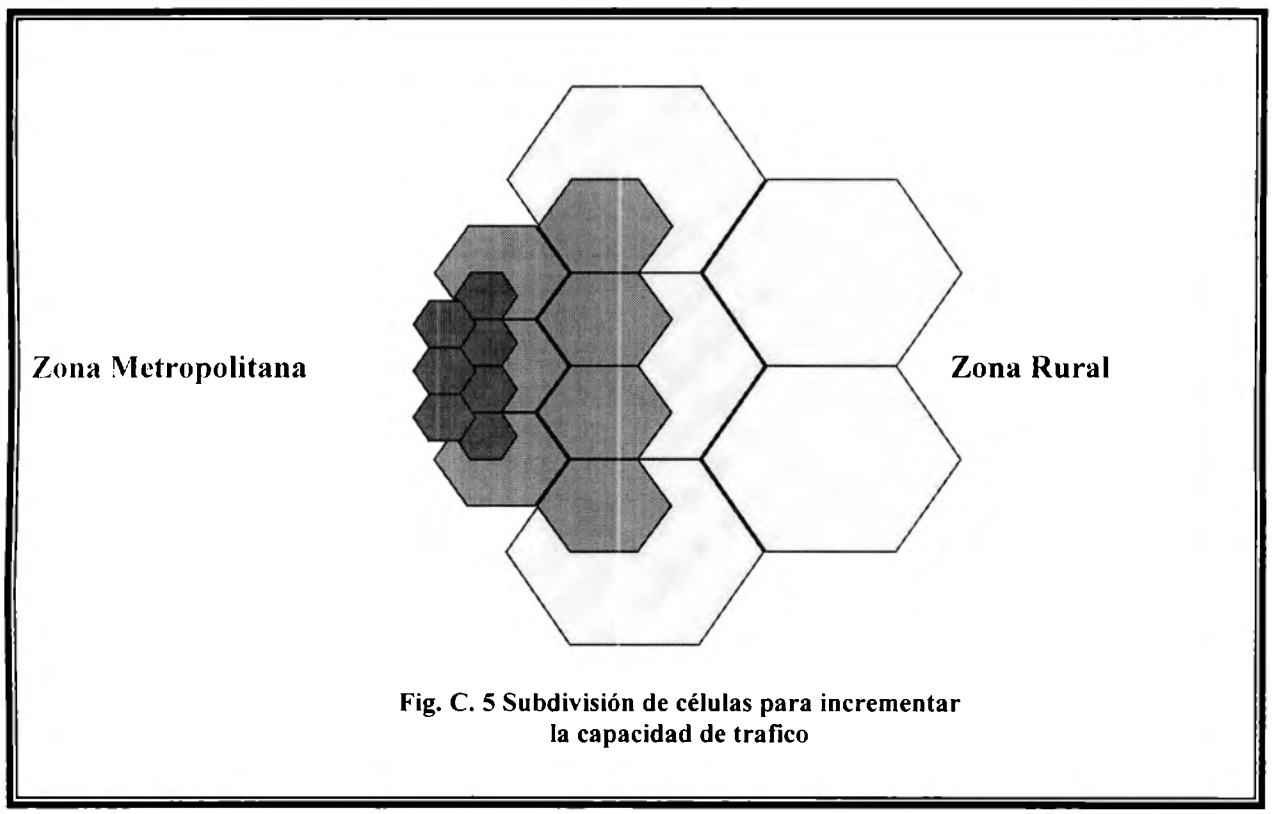


Fig. C. 5 Subdivisión de células para incrementar la capacidad de trafico

La figura C.5 ilustra la división simple de células y la reducción gradual en el tamaño de la célula en la región de transición entre un área rural de bajo tráfico y la región de alto tráfico que rodea a una zona metropolitana. Cuando las células se dividen de esta manera, se debe tener mucho cuidado al asignar las frecuencias de radio a las nuevas células, ya que inclusive se puede necesitar un nuevo plan de reutilización para prevenir la interferencia entre células.
Mediante las técnicas de reutilización de frecuencia y subdivisión de células, se logra que los sistemas celulares atiendan un gran número de clientes en un área grande, usando un ancho de banda relativamente pequeño.

Un sistema celular básico esta constituido por una serie de células, cubiertas cada una por un sistema de radio que permite la conexión de las unidades móviles con la Estación Base (EB). Así mismo, se tienen Centros de Conmutación Móvil, que permiten la interconexión entre las EB y la RTPC.
El esquema siguiente (figura C.6) muestra la configuración básica de un sistema celular.

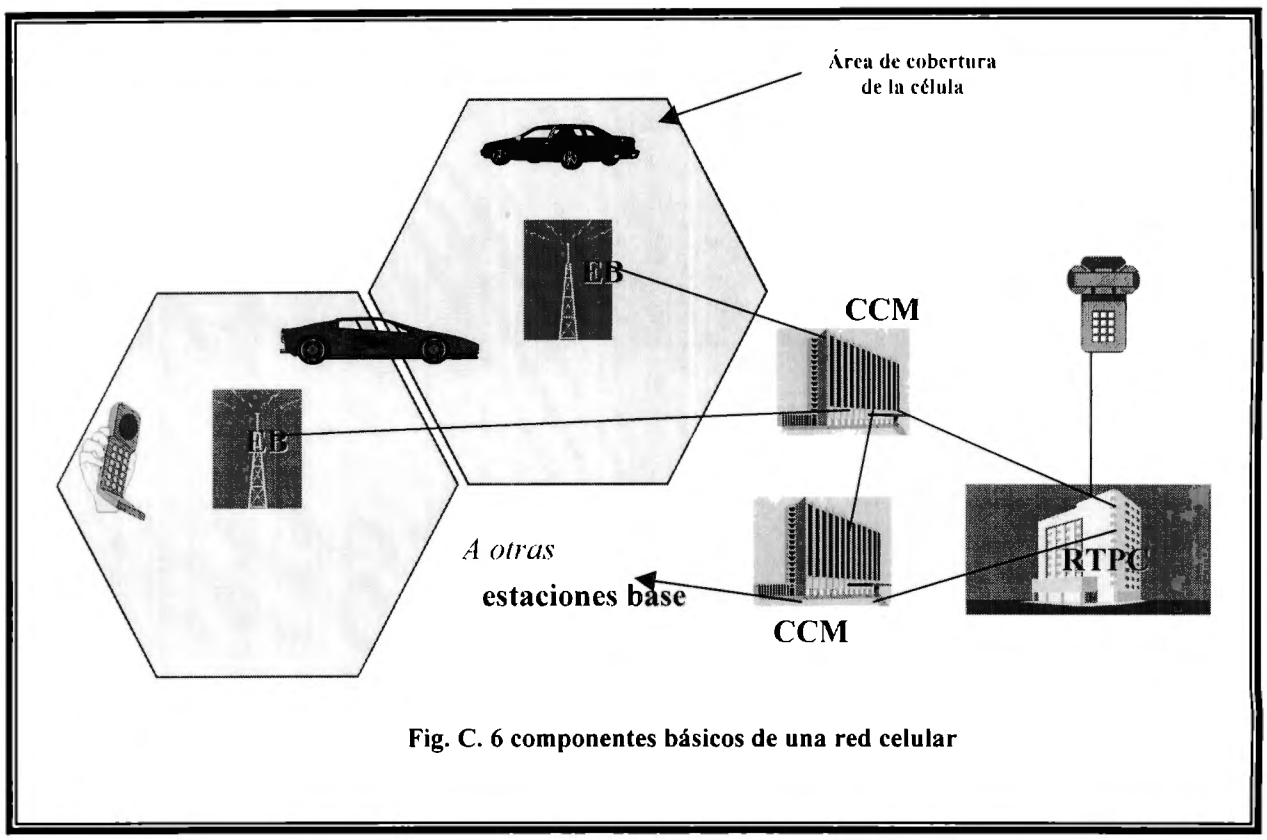


Fig. C. 6 componentes básicos de una red celular

Como se muestra en el esquema anterior, las estaciones base en el centro de cada célula están unidas a un centro de conmutación móvil, que es una central telefónica de conmutación modificada para el sistema celular. Una red celular consiste de varios centros de conmutación móvil interconectados entre sí. La configuración anterior permite la realización completa de los distintos tipos de llamada, tales como, llamadas de móvil a fijo, de fijo a móvil, y de móvil a móvil.

Un sistema celular podría utilizar células cuadrangulares o rectangulares, o circulares, sin embargo, al utilizar el circulo como forma para las células, se dieron cuenta, que entre ellas existían áreas de sombra y de traslape como se ve en la figura C.7, por lo que se han escogido células hexagonales ya que permiten cubrir un área mayor con un número menor de estaciones base. Aunque es importante tener en cuenta que en realidad las células no son hexagonales, sino que tienen una forma irregular determinada por parámetros como la propagación de las ondas de radio en el terreno, obstáculos y las restricciones de la estación base debidas a factores geográficos, como se muestra en la figura C.7.

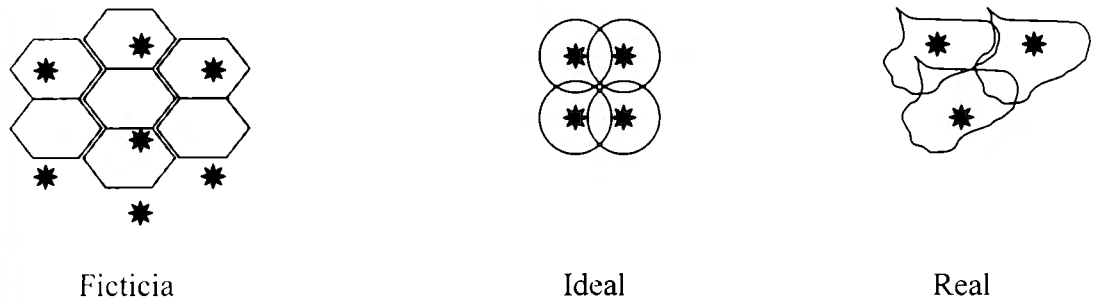


Fig. C.7 Relación entre coberturas

La estación base puede colocarse en el centro o en cada esquina alterna de la célula. Cuando se sitúan en el centro utilizan antenas omnidireccionales, mientras que si se sitúan en las esquinas, se requiere usar antenas direccionales.

La utilización de la estación base en el centro de la célula es apropiada para ciudades pequeñas, mientras que para ciudades grandes se utilizan estaciones base en las esquinas, pues se elimina la interferencia entre canales de la misma frecuencia, como se muestra a continuación en la figura C.8.

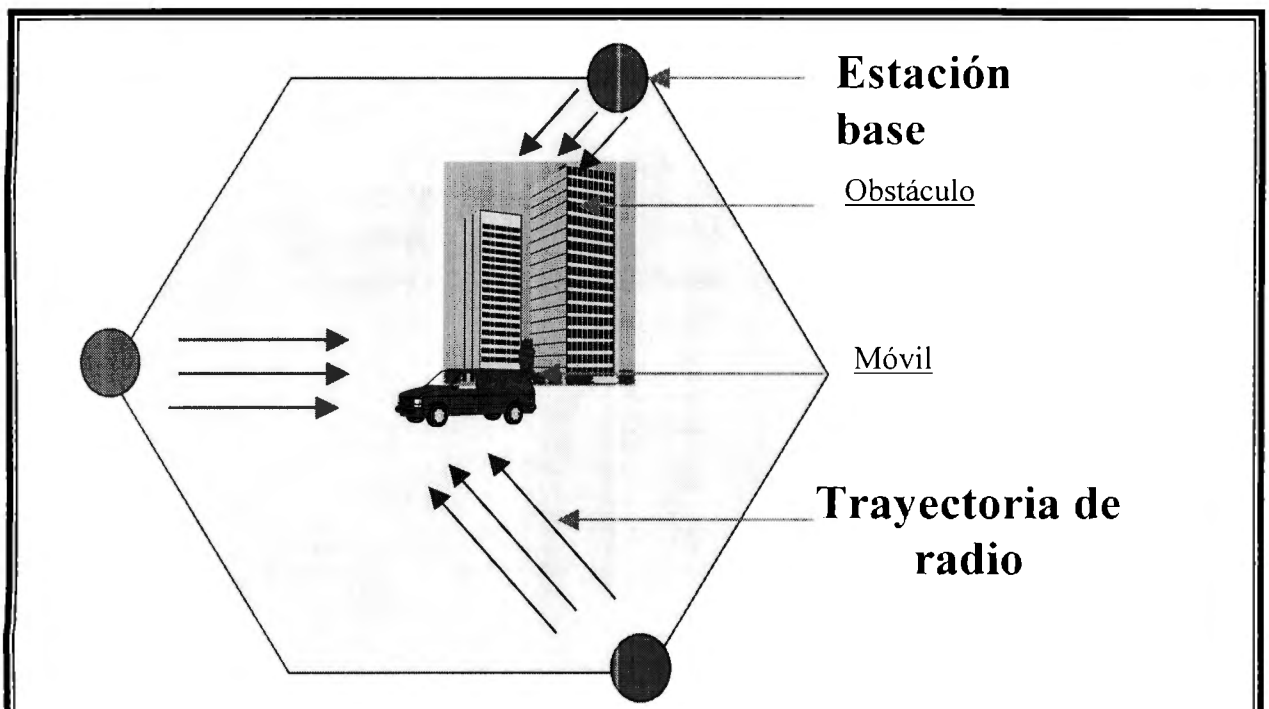


Fig. C. 8 Estaciones base en esquinas alternadas de la célula

La telefonía celular incluye algunas características importantes:

- El registro, cuando un móvil notifica a la red, que se encuentra presente y esta disponible para realizar y recibir llamadas, esta llevando a cabo el proceso de registro, esto es, trata de la facultad del sistema para conocer la ubicación del móvil en todo momento dentro de la zona de cobertura.

- "Hand-over" o traspaso, un de los servicios principales de la telefonía móvil, es como su nombre lo dice, permitir la comunicación aún cuando estamos en movimiento, de ahí y según las características de nuestro sistema, el handover es el proceso de pasar la comunicación de un mismo móvil de un canal a otro, es decir, es la capacidad del sistema para realizar el cambio de célula sin perder la comunicación. En función de la relación entre los canales origen y destino, los handover se clasifican en:

    - Handover intercelular, si el canal destino se encuentra sobre otra frecuencia distinta a la del origen, pero en la misma célula.
    - Handover inter-BS, cuando existe cambio de células, pero ambas BS se encuentran dentro del mismo sistema controlado de estaciones base.
    - Handover inter-CCM, cuando existe un cambio de células y de controlador de estaciones base pero ambos controladores de BS dependen de la misma central de conmutación móvil.
    - Handover entre CCMs cuando existe un cambio de células y ambas dependen de CCMs distintas.

- "Hand-Off", Es la manera de mantener una conversación celular sin interrupciones mientras el usuario esta desplazándose de una celda a otra dentro de una red celular, en algún momento, el nivel de energía de su señal se comenzará a recibir débil o atenuada y podría perder la conexión, para evitar eso, la BS coordina las decisiones de handoff, como el tipo de hand-off necesario, la nueva celda requerida y el momento de llevar a cabo el hand-off. Normalmente el CCM no desempeña un papel de control del hand-off excepto en algunos casos, pero siempre será informado cuando ocurra un hand-off.

- "Roaming", es la capacidad que ofrece una red móvil para poder registrarse en cualquier VLR de la red. Este concepto esta comúnmente saciado al registro de un móvil en una red distinta de la propia. Esta capacidad solo se puede ofrecer si se cumplen ciertos requisitos técnicos y administrativos que lo permitan.

Dentro del sistema, se reserva un número de canales para señalización. Adicionalmente, la red se divide en un número de áreas de tráfico, cada área consiste en un grupo de células. La estación base genera un código de identificación correspondiente con el área de tráfico a la que pertenece, como parte de la información transmitida por los canales de señalización.

La transferencia de llamada tiene como propósito asegurar una relación señal a ruido adecuada durante todo el tiempo de la llamada.
La transferencia de llamada tiene un efecto mínimo en la interferencia entre canales de la misma frecuencia, ya que ésta se controla principalmente por la separación física de las células que

utilizan el mismo grupo de frecuencias. Un móvil puede trasladarse por toda el área de servicio y su comunicación no debe interrumpirse. Para lograr esto, cada vez que el móvil pase de una célula a otra, la llamada debe transferirse de la estación base de la célula que sirve actualmente al móvil, a la estación base de la célula a la que se esté pasando.

C.1.4 TELÉFONOS CELULARES

Existen varios tipos de teléfonos celulares, incluyendo teléfonos de automóvil, portátiles y de bolsillo, que cuentan con tres componentes esenciales:

1) Microteléfono
2) Transceptor de radio
3) Antena

El microteléfono contiene todos los elementos de interacción con el usuario, tales como, transductores acústicos, teclado de marcación y de funciones (envío, borrado, etc.) y visualizador. El microteléfono contiene también la unidad de control de todo el radioteléfono y comanda al resto del equipo.

El transceptor de radio, consiste de un transmisor y receptor, que sintoniza cualquier canal de radio del sistema celular. El transceptor se comunica con las estaciones base para establecer las conexiones, determinar las frecuencias adecuadas y coordinar el cambio de célula si fuera necesario.

La antena es un elemento critico y su tipo determina la calidad de recepción y transmisión.

Los teléfonos celulares no se utilizan de la misma manera que los teléfonos convencionales. En primer lugar, los usuarios que llaman por la red celular no escuchan el tono de invitación a marcar cuando se inicia una llamada de salida. También, la conexión no se establece hasta que el número telefónico es marcado y la tecla de envío se presiona. La única señal audible que el usuario percibe es la de llamada u ocupado una vez que la conexión se ha establecido; el procedimiento posterior es idéntico al de una llamada telefónica convencional.

Para el establecimiento o recepción de llamadas entre usuarios móviles, se emplea uno o más canales de radio de control o búsqueda. Si un usuario móvil desea hacer una llamada, el aparato portátil móvil explora los canales predeterminados para seleccionar el canal de control más fuerte y lo monitorea para recibir información acerca del estado y disponibilidad de la red. Cuando el número telefónico de destino se ha marcado y se ha presionado la tecla de envío, el aparato portátil móvil encuentra un canal de control libre y radia una solicitud de canal de usuario.

Todas las estaciones base emplean los mismos canales de control y los monitorean para determinar las solicitudes de llamada.Cuando alguna estación base recibe tal solicitud, se envía un mensaje al centro de conmutación móvil (CCM) más cercano que indica tanto el deseo del usuario móvil de establecer una llamada, como la intensidad de la señal de radio que se recibe desde el móvil. El CCM determina que estación base ha recibido la mayor intensidad de señal y basándose en esto determina en que célula se encuentra el móvil. Entonces, solicita al aparato

portátil móvil identificarse con un número autorizado que se puede emplear para el cobro de la llamada, "eliminándose también cualquier posibilidad de fraude".

Después de la autorización de una llamada de salida, se asigna un canal libre de radio en la célula apropiada para el transporte de la propia llamada extendiéndose hasta su destino sobre la RTPC. La célula apropiada no necesariamente incluye a la estación base más cercana, redes celulares más sofisticadas pueden también escoger estaciones base adyacentes, si esto ayuda a aliviar la congestión de canales.

Al final de la llamada, el usuario móvil genera la señal de terminación de llamada, que provoca la liberación del canal de radio y regresa al aparato portátil al estado de monitoreo del canal de control y búsqueda.

Cada CCM, controla a cierto número de estaciones base, y si durante el curso de una llamada, el usuario móvil pasa de una célula a otra, entonces el CCM está capacitado para transferir la llamada y enrutarla vía una estación base diferente apropiada para la nueva posición. El proceso de cambio a la nueva célula ocurre sin perturbar la llamada y es lo que se conoce como transferencia automática, (hand off) anteriormente definida. Se inicia ya sea mediante la estación base activa o mediante el usuario móvil, dependiendo del tipo de sistema.

Las intensidades relativas de señales que se reciben en todas las estaciones base más cercanas, se comparan entre sí continuamente y cuando la intensidad de señal de la estación en curso cae abajo de un umbral preestablecido, o es sobrepasada por la intensidad de señal disponible vía una estación base adyacente, entonces se inicia la transferencia automática. El CCM establece un canal duplicado de telefonía y radio en la nueva célula y una vez establecido, la llamada se transfiere a la nueva trayectoria de radio mediante un mensaje de control al aparato portátil. Cuando el nuevo canal y estación base se confirman, la conexión original se libera.

Después de que se prende un aparato portátil, a intervalos regulares de tiempo emplea el canal de control para indicar su presencia en el CCM más cercano. Esto permite que el CCM local tenga cuando menos una idea de la localización del usuario móvil. Si está fuera de su CCM de residencia, el CCM local inicia un procedimiento de registro, en el que se interroga al CCM de residencia sobre los detalles del móvil, incluyendo el número de autorización y otra información. La información está contenida en el CCM de residencia en una base de datos que se conoce como Registro de Ubicación de Residencia (RUR).

Este registro, contiene la información de mapeo necesaria para completar llamadas con el usuario móvil desde la RTPC (identidad de red, autorización e información de cobro). El CCM local duplica algo de esta información en un Registro de Ubicación de Visitante (RUV), hasta que él llamante abandona el área del CCM. Una vez que el RUV se ha establecido en el CCM local, el usuario móvil puede efectuar llamadas de salida. El procedimiento de registro es parte crucial del mecanismo para el rastreo del paradero de los usuarios móviles para que las llamadas entrantes se entreguen.

Las llamadas de entrada sé enrutan primero vía el CCM más cercano al punto de origen. Este CCM interroga al RUR sobre la ultima ubicación del usuario móvil. Entonces, la llamada se puede dirigir hacia el CCM en el que fue detectado por ultima vez, con lo que un mecanismo de búsqueda, empleando los canales de control de la estación base, puede localizar la célula exacta

en la cual se encuentra actualmente el móvil. Se puede entonces seleccionar un canal de radio libre apropiado para completar la llamada.

C.1.5 TELEFONÍAS CELULARES ACTUALES

En todo el mundo, existen dos formas en las cuales la información puede ser transmitida por los medios de telecomunicación, incluida la telefonía celular.

1) Transmisión analógica
2) Transmisión digital

Desde su inicio, la transmisión celular ha sido analógica. Este sistema genera una onda similar a aquella que produce la voz humana. En el sistema analógico, una conversación ocupa totalmente un canal y limita la capacidad de procesar llamadas simultáneamente.

En contraste, la tecnología digital convierte la transmisión analógica en códigos binarios o datos que representan los sonidos de la voz. El beneficio del uso del sistema digital para el usuario, se refleja en la facilidad para realizar y recibir llamadas, gracias a que esta red procesa varias llamadas simultáneamente. El sistema digital "es más seguro porque evita la posibilidad de que su llamada sea interferida". Esta tecnología es la plataforma para los nuevos y diversos sistemas y servicios de telecomunicaciones.

Sistema Analógico.- Hace referencia a la onda radial por la cual se transmite, es una señal frágil transmitida a baja potencia. La llamada es convertida en impulsos eléctricos que viajan en forma de ondas de radio "análogas" al sonido de la voz original. Estas ondas se distorsionan fácilmente por factores naturales (lluvias), o por obstáculos (árboles, edificios, líneas eléctricas), que ocasionan pérdida de calidad en la transmisión e inclusive la perdida de la misma. Adicionalmente estas redes no se pueden ampliar.

Sistema Digital.- Toma las señales analógicas (sonidos), y las traduce a códigos binarios que pueden ser transmitidos a alta velocidad, para después ser reconvertidos en el sonido de la voz original.
La llamada se transmite de forma similar a los datos de una computadora, lo que permite que las transmisiones sean de alta calidad, velocidad y resistencia a los inconvenientes usuales.
Este sistema de redes de transmisión permite subdividirse con el objeto de ampliarse.

Tabla C.3.

Cuadro comparativo	
Sistema analógico	**Sistema Digital**
• Una llamada por canal • Posibilidad de interceptación de llamadas • Tecnología de ayer y todavía hoy	• Varias llamadas simultaneas por canal • Mayor privacidad y menor riesgo de intercepción de llamadas • Tecnología de hoy y del futuro

C.2 USO DEL ESPECTRO

Los sistemas de comunicación inalámbrica hacen frente al problema común de la escasez del espectro. Debido a la disponibilidad limitada de la capacidad de radio, el índice total del tráfico transmitido simultáneamente es en cualquier momento limitado. No semejante a los sistemas de comunicaciones alambrados que pueden aumentar la taza de datos soportada en el intercambio en función del costo del equipo agregado, los limites de la capacidad inalámbrica son difíciles en el diseño del sistema.

Muchos sistemas celulares actualmente en operación trabajan en la banda de los 800 - 900 Mhz. y poseen 832 canales duplex disponibles. Sin embargo, muchas administraciones de telecomunicaciones en el mundo han considerado adecuado concesionar el servicio a dos compañías de telefonía celular por zona, con la mitad de canales del espectro de frecuencias para cada una de ellas. Las tablas siguientes (D.1 y D.2), muestran diversos parámetros de los tres sistemas celulares analógicos más utilizados en el mundo.

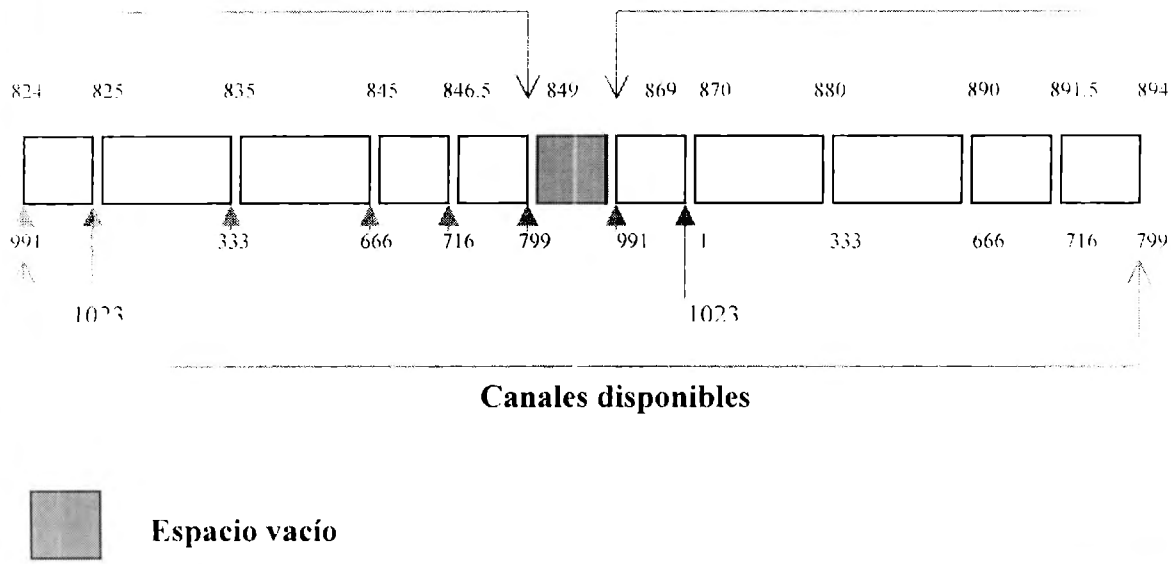
Tabla C2.1.

SISTEMA AMPS (Advanced Mobile Phone Systems, utilizado en Norte América.		
Parametro	Compañía A	Compañía B
Estación Base (Mhz)	869-880, 890-891.5	880-890, 891.5-894
Estación Móvil (Mhz)	824-835, 845-846.5	835-845, 846.5-849
Potencia máxima(Watts)	3	3
Radio de célula (Kms)	2-20	2-20
Número de canales duplex	416	416
Ancho de banda de canal (Khz)	30	30
Modulación de voz	FM	FM

Tabla C.2.2

SISTEMA ETACS (Extended Total Access Communications Systems, utilizado en el Reino Unido.) **JTACS**(Japan Total Access Communications Systems, utilizado en Japón.)		
Parametro	ETACS	JTACS
Estación Base (Mhz)	872-905	860-870
Estación Móvil (Mhz)	917-950	915-925
Potencia máxima(Watts)	3	2.8
Radio de célula (Kms)	2-20	2-20
Número de canales duplex	1320	399
Ancho de banda de canal (Khz)	25	25
Modulación de voz	FM	F

Fig. C 2.1 Estructura espectral del sistema AMPS



Sin embargo, los sistemas analógicos de banda ancha de FM no tienen oportunidad de aumentar el número de usuarios. Consecuentemente, sistemas de FM de banda angosta con mayor capacidad de usuarios, mostrados en la tabla D.3 siguiente, han reemplazado los sistemas de banda ancha. Mas aún, para aumentar esta capacidad de usuarios, sistemas celulares digitales se muestran en la tabla D.4.

Tabla C2.3

SISTEMAS ANALÓGICOS DE BANDA ANGOSTA		
Parámetro	NAMPS	NTACS
Estación Base (Mhz)	869-894	843-870
Estación Móvil (Mhz)	824-849	898-925
Potencia máxima(Watts)	3	2.8
Radio de célula (Kms)	2-20	2-20
Número de canales duplex	2496	2160
Ancho de banda de canal (Khz)	10	12.5
Modulación de voz	FM	FM

NAMPS.- Narrowband Advanced Mobile Phone System, utilizado en Norte America.
NTACS.- Narrowband Total Access Communications System, utilizado en japón.

Tabla C2.4

SISTEMAS CELULARES DIGITALES		
Parámetro	GSM	NADC
Estación Base (Mhz)	935-960	869-894
Estación Móvil (Mhz)	890-915	824-849
Potencia máxima(Watts)	20	3
Número de canales duplex	125	832
Ancho de banda de canal (Khz)	200	30
Método de acceso al canal	TDMA	TDMA
Usuario por canal	8	3

GSM.- Global System for Mobile Communications, utilizado en Europa.
NADC.- North American Digital Celular, Utilizado en Norte América.

C2.2 TÉCNICAS DE ACESO MÚLTIPLE

Él término acceso múltiple se refiere a que "simultáneamente" los usuarios puedan accesar a un mismo sistema. En otras palabras, un gran número de usuarios utiliza un sistema común en un rango de canales. Cada usuario puede tener acceso a cualquier canal, pero no el mismo canal al mismo tiempo. Las diferentes técnicas de acceso múltiple ofrecen diferentes formas de lograr acceso a un canal. Los sistemas de comunicación inalámbrica emplean diferentes técnicas de acceso múltiple, entre las que destacan: FDMA, TDMA y CDMA.

FDMA (Frequency Division Multiple Access).- Sistemas analógicos como AMPS, NAMPS, TACS, utilizan esta técnica. En esta técnica, cada usuario es asignado a una parte discreta del rango de frecuencia conocida como canal. Solamente un usuario en un tiempo determinado es asignado a un canal. Ningún otro usuario puede accesar este canal hasta que la llamada del usuario inicial haya terminado o hasta que dicha llamada se cambie a otro canal. En el sistema AMPS cada canal tiene un ancho de banda de 30 Khz, en el NAMPS de 10 Khz y en el sistema TACS de 25 Khz.

TDMA (Time Division Multiple Access).- Los sistemas digitales GSM, PDC y D-AMPS, utilizan esta técnica. El sistema TDMA, al igual que el FDMA, emplea partes discretas del rango del espectro de frecuencia referido como una portadora. Cada portadora es dividida en espacios de tiempo, y cada usuario es asignado a un espacio de tiempo conocido como canal.
En el sistema TDMA, cada usuario alterna el uso de la portadora en la porción del tiempo y puede enviar o recibir información en un tiempo determinado. De esta forma, el flujo de información no es continuo para ningún usuario, pero es enviado y recibido en espacios de tiempo.

En TDMA, ningún otro usuario puede accesar una porción de tiempo asignada, hasta que la llamada original haya terminado, o bien, sea cambiada a otra frecuencia. En particular, el sistema GSM divide a una portadora de 200 Khz en 8 porciones de tiempo, así mismo, el sistema D-AMPS, divide una portadora de 30 Khz en tres canales. Es importante hacer notar que el sistema GSM no ofrece eficiencia espectral sobre los sistemas analógicos TACS, ya que GSM permite 8

usuarios en 200 Khz (200 Khz/8 = 25 Khz), lo mismo que emplea TACS por usuario. De manera similar D-AMPS no ofrece eficiencia espectral sobre el sistema NAMPS (30 Khz/3 = 10 Khz).

CDMA (Code División Multiple Access).- Es una tecnología digital que se diferencia de los sistemas FDMA y TDMA por utilizar códigos digitales únicos para identificar a los usuarios. CDMA emplea códigos digitales únicos en lugar de la separaración de frecuencias RF o espacios de tiempo. Todos los usuarios emplean el mismo rango del espectro de frecuencias. CDMA es una tecnología de banda ancha en la cual el espectro de comunicación es dividido en portadoras de aproximadamente 1.23 Mhz.

En CDMA el tráfico de canales es creado mediante la asignación de un código único para cada usuario dentro de la portadora. Cada código es posteriormente empaquetado y transmitido simultaneamente a lo largo de la portadora. Uno de los aspectos de CDMA es que aunque existe un límite en el número de llamadas telefónicas atendidas por la portadora, el número no es fijo. Además la capacidad del sistema dependerá de diferentes factores que el operador pueda controlar. CDMA es una tecnología de espectro esparcido, lo que significa que dispersa la información contenida en una señal particular de interés en un ancho de banda mucho más grande que el de la señal original.

En CDMA una señal comienza con un estándar de 9.6 khz, y es esparcida posteriormente en una banda de frecuencia de 1.23 Mhz. Para obtener la señal CDMA de espectro esparcido, el receptor decodifica la señal, ya que conoce el código digital que se empleó para codificar la señal original. Cuando la señal es recibida, los códigos de la señal deseada son removidos, separando la información del usuario. La técnica de espectro esparcido de CDMA ha sido utilizada para las comunicaciones militares por más de 50 años.

Los militares emplean esta técnica por muchas de las mismas razones por las que hoy los operadores de sistemas inalámbricos están cambiando a CDMA, algunos de estos aspectos son:

- CDMA reduce significativamente el tráfico de señales debido a su espectro esparcido.
- CDMA proporciona seguridad en la comunicación, ya que es muy difícil de detectar y descifrar tanto el espectro esparcido como los códigos digitales.

El número máximo de usuarios por portadora depende de la actividad que exista en dicha portadora y por lo tanto no es preciso. Depende del operador de red el decidir si un usuario más es aceptado en la portadora que se encuentre en sobrecarga, corriendo el riesgo de perder calidad en la comunicación.

C.3 REVISIÓN DE LA TECNOLOGÍA DE PCS

La asociación de Industrias de Telecomunicaciones (TIA) la especificación IS-136 es la base del Acceso Múltiple por División de Tiempo (TDMA) de la tecnología de interfase aérea de PCS. IS-136 esta diseñada para operar en ambas bandas de frecuencia (800 Mhz y 1900 Mhz.).

C.3.1 DIGITAL CONTROL CHANNEL (DCCH)

El DCCH forma el corazón de la especificación IS-136 y es el realce primario para la tecnología inalámbrica digital TDMA. Es un nuevo mecanismo de control de canal adicional al canal de control analógico (ACC), el canal de voz analógico (AVC), y el canal de trafico digital (DTC) de la interfase aérea TDMA. La tecnología IS 136 DCCH TDMA proporciona la plataforma para los PCS, introduciendo nuevas funcionalidades y soportando características sobresalientes que hacen de los PCS un sistema digital poderoso.

C.3.2 OPERACIÓN DE "DUAL-BAND DUAL-MODE" BANDA-DUAL MODO-DUAL

Los teléfonos PCS de banda dual operan en 800 MHz. y 1900 Mhz. permiten a usuarios recibir las características completas de los PCS y servicios para sistemas IS-136 dondequiera que vaguen. La capacidad del Modo-Dual proporciona continuamente servicio e interoperabilidad entre redes digitales y analógicas. Como resultado, un teléfono PCS puede proporcionar acceso a todos los servicios inalámbricos al aire libre, ser utilizado en un sistema edificio privado, y servir como un teléfono sin cables digital en el hogar.

C.3.3 CAPACIDADES Y CARACTERÍSTICAS

La siguiente tabla E.1 muestra las capacidades y características más importantes de los PCS.

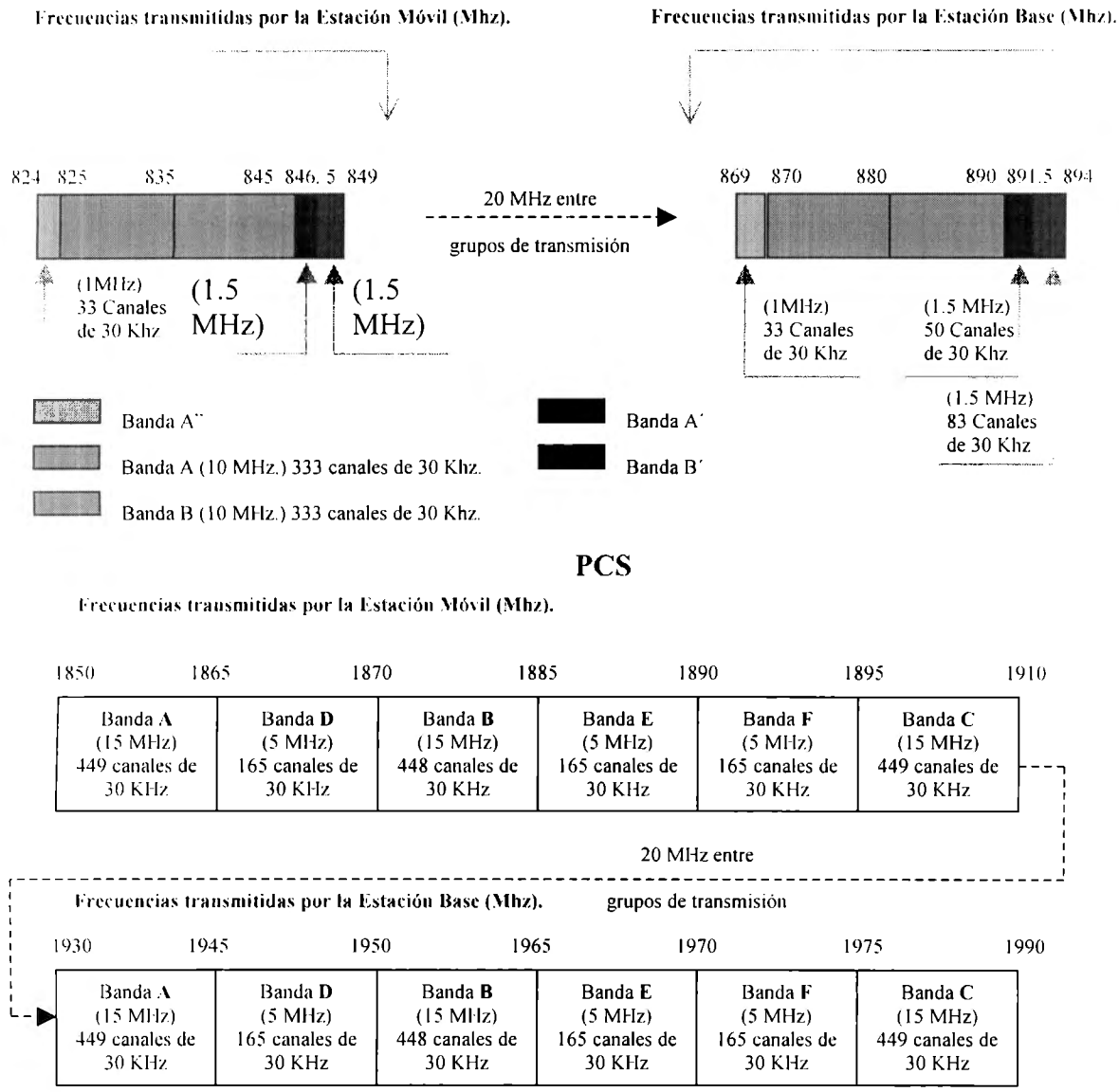
Tabla C3.1

Característica	**capacidad**
Modo dormido	prolonga el tiempo espera del teléfono y realza la vida de la batería
Servicio de Mensajes Cortos (SMS)	Transfiere mensajes alfanuméricos a y desde teléfonos PCS y celulares.
Privacidad de datos y voz	Incrementa la resistencia al escucha no autorizada
Calidad de voz superior	Resulta en menos ruido de fondo y pocas caídas de llamadas
Ambiente jerárquico	Proporciona soporte para operación de microceldas-macroceldas
Pre-exploración inteligente	Permite control ajustado de la selección del sistema
Sistemas residenciales y privados de IDs	Proporciona Servicio de Oficina Inalámbrica (WOS) mas simplificado y controlado y características de Estación Base Personal (PBS)
"Seamless roaming"	Permite vagar entre frecuencias usando teléfonos de banda dual y proporciona soporte para roaming internacional
Datos de circuitos-conmutados	Proporciona transmisión de datos altamente confiables
Soporte	E-mail inalámbrico, telefax, y acceso a Internet
Autentificación	Incrementa la seguridad del teléfono y la resistencia a clonación
Identificador de Numero de Llamada (CNI)	Permite que los llamadores sean identificados antes el contestar
Indicador de Espera de Mensaje (MWI)	Notifica a los usuarios que tienen mensajes de correo de voz
Servicio de envío de texto	Los operadores toman mensajes del llamador y envían mensajes de texto al teléfono PCS

C.3.4 COMPARACIÓN DE LOS ESPECTROS DE PCS Y CELULAR

La figura C3.1 ilustra el espectro celular inalámbrico de 800 MHz, y el espectro de 1900 MHz de PCS.

Fig. C3.1 Comparación de los espectros de PCS y Celular



PCS

C3.2 EL AMBIENTE DEL DCCH

Un canal de radio consiste de dos frecuencias dentro del espectro de radio-frecuencia (RF) que están separados por una distancia fija. Estas dos frecuencias permiten a un sitio de la célula y el teléfono inalámbrico transmitir y recibir señales simultáneamente. Los sitios de la célula se comunican con los teléfonos inalámbricos usando dos diferentes canales de radio:

- El canal de voz y
- El canal de control

En los sistemas TDMA, cada canal de radio-digital puede llevar hasta tres llamadas de voz por trafico de voz multiplexado en tiempo dentro de ranuras de tiempo. Un DCCH es introducido dentro del sistema TDMA por la reprogramación de uno de esos canales de trafico, llamado DTCs, para convertirse el DCCH en una frecuencia que contiene el DTCs existente.

La figura C3.2 representa el par de ranuras DTC (1,4) empleadas para un DCCH, y muestra cada célula dividida en sectores (A, B, C). Solamente un par de ranuras es requerido para un DCCH en cada sector de la célula sin importar él numero de radios digitales en el sector.

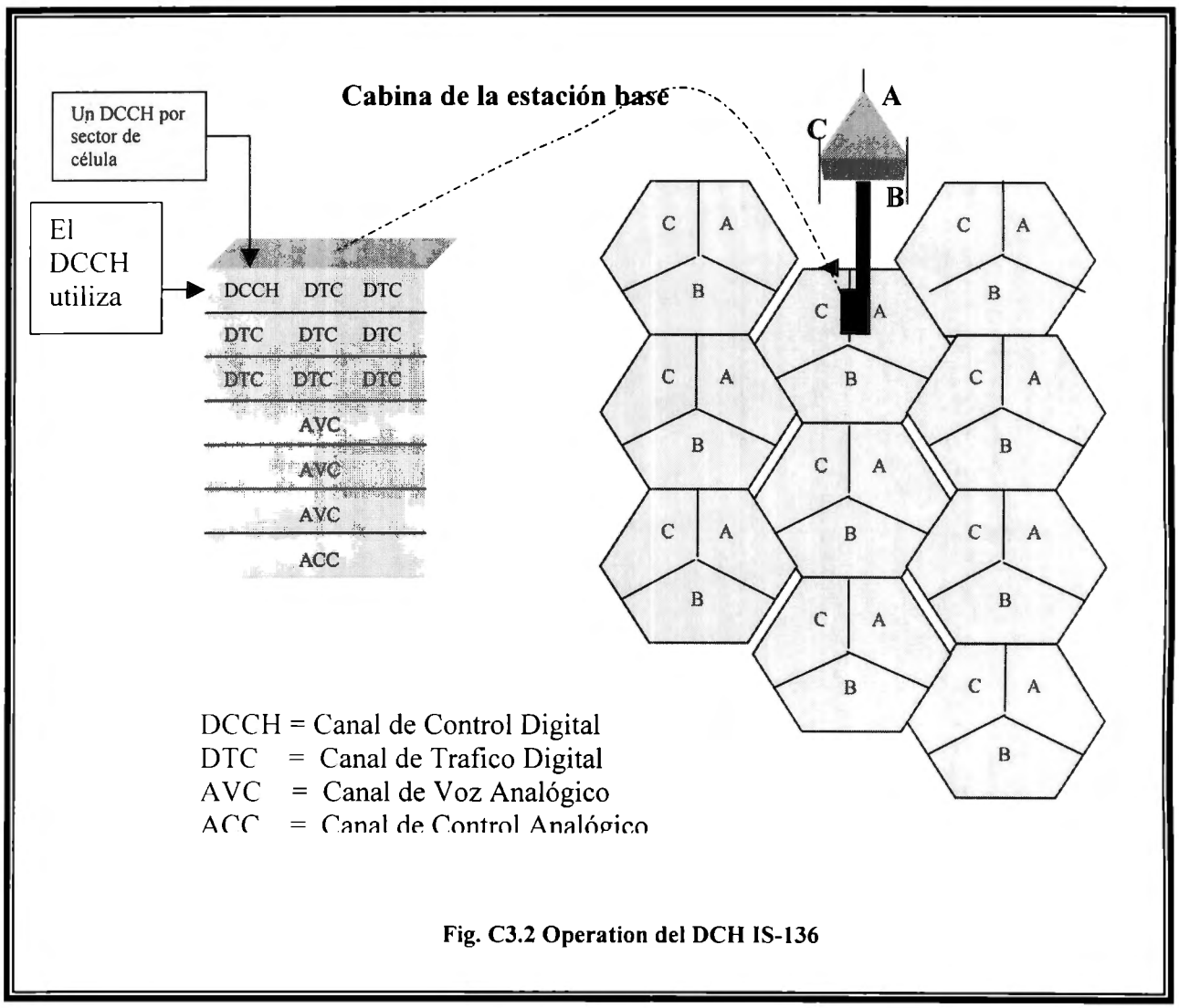


Fig. C3.2 Operation del DCH IS-136

C3.2.1 PRINCIPIO DE OPERACIÓN

La información es transmitida en el flujo del DCCH en dos direcciones sobre la interfase de aire: desde el sistema al teléfono (downlink), y desde el teléfono al sistema (uplink). En la figura anterior, la estación base representa al sistema.

La capacidad del DCCH y los teléfonos PCS vigilan a un DCCH en cada sector de un sistema inalámbrico que soporta servicios IS-136. Un teléfono PCS explorará para este canal, sincronización de ganancia, y comienza a decodificar la información proporcionada por medio de

un canal de control de "broadcast" en el DCCH. El DCCH sirve como el canal de control de teléfonos hasta que el teléfono encuentra otra célula que sea mas apropiada.

Los teléfonos PCS reciben páginas, que envían los iniciadores, y se comunican con el sistema en el DCCH. Después de recibir una pagina o presentar una iniciación de llamada, un canal de trafico es entonces designado para la llamada, y el teléfono mantendrá una conversación celular sin interrupciones mientras el usuario esta desplazándose de una celda a otra dentro de una red celular. En la terminación de la llamada, el teléfono regresa al DCCH para aguardar interacciones adicionales.

C3.3 LA INTERFASE DE AIRE: PROTOCOLO DE VARIAS CAPAS

La interfase de aire utilizada en los PCS esta estructurada en diferentes capas, cada una con propósitos específicos. Esta fractura conceptual hace más fácil entender las interacciones entre la estación base y el teléfono a través de la interfase de aire. Existen 4 capas:

- Capa física (capa 1). Repartida con la interfase de radio, "bursts", las ranuras, las tramas, y las supertramas
- Capa de enlace de datos (capa 2). Maneja el empaquetado de los datos, la corrección de error, y transporte de mensaje.
- Capa de mensaje (capa 3). Crea y maneja envío y recepción de mensajes a través del aire.
- Capas de aplicación superior. Representan el teleservicio que es utilizado actualmente, tal como voz y transacciones de mensajería, o servicios futuros como programación en-aire.

C3.3.1 EL MODELO DE LA INTERFASE-AIRE

La figura C3.3 muestra el modelo de la interfase-aire. Esta estructura simplifica la introducción de servicios actuales y futuros empleando la plataforma DCCH IS-136 por que las capas inferiores en el protocolo de interfase-aire (la interfase de radio, manejo de datos, mensajes, etcétera) permanezcan sin cambiar.

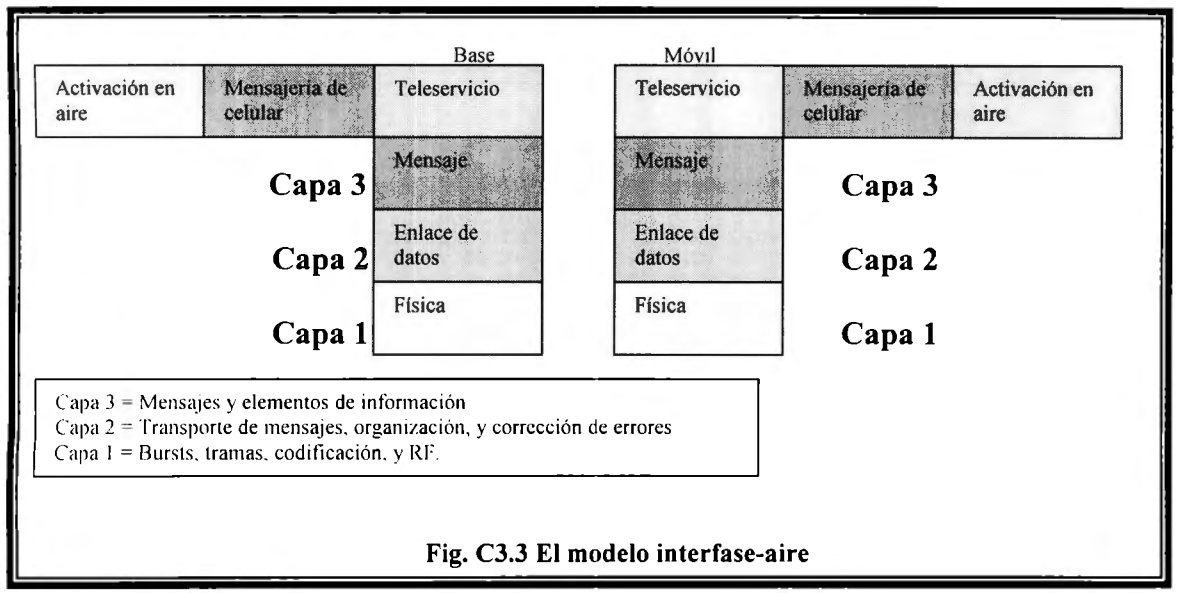


Fig. C3.3 El modelo interfase-aire

C3.3.2 PRINCIPIO DE OPERACIÓN

En la figura C3.4 se muestra como un mensaje de capa 3 es vinculado dentro de varias mas de la capa 2 y como la trama de tiempo de la capa 2 es trazada sobre una ranura de tiempo. La ranura de tiempo es más tarde trazada sobre un canal DCCH. La figura muestra como la información es pasada de capa en capa hacia abajo a través de la pila hasta que se crea un "burst", listo para la transmisión. En el extremo de recepción, la información es desenvuelta tanto como sea necesario pasar el mensaje hasta la aplicación.

El mensaje de capa 3 mostrado en la figura C3.4 puede ser un registro de "uplink", un mensaje de PCS de "downlink", una respuesta de la paginación, o un mensaje de difusión. El mensaje de capa 3 es empaquetado dentro de una trama de la capa 2 donde un campo de encabezado y corrección de error son agregados. El paquete entonces se codifica y los dígitos binarios individuales se interpolan (mezclados y distribuidos) para contrariar errores introducidos en el ambiente de radio.

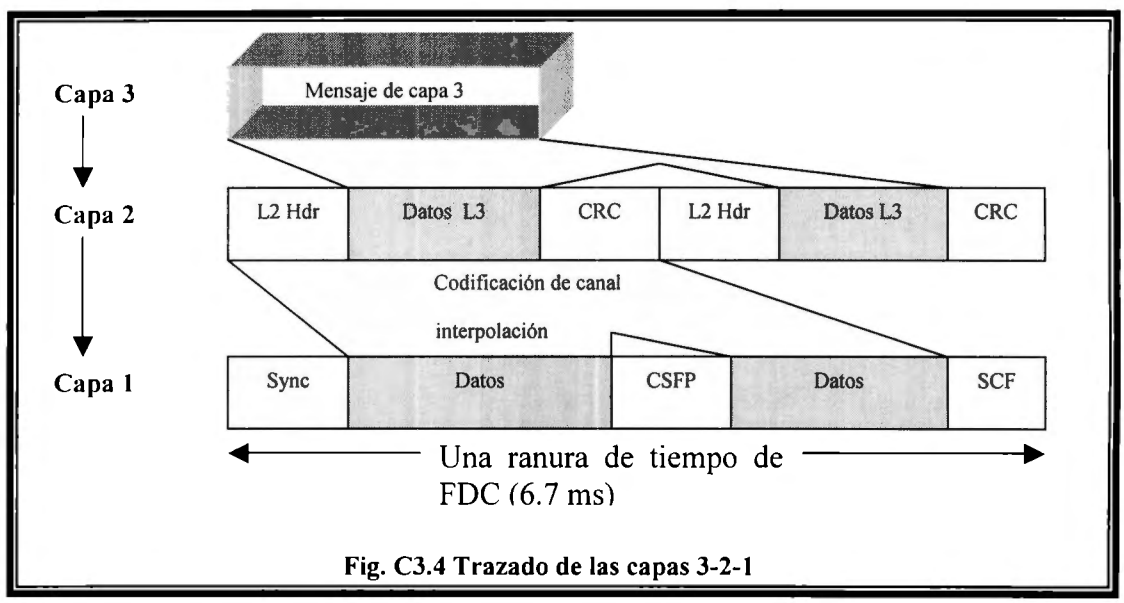


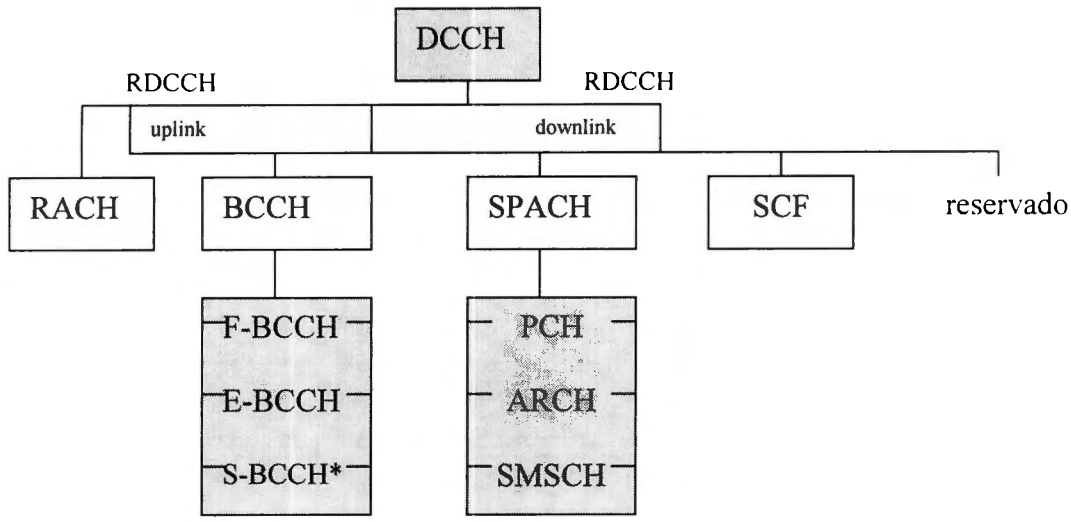
Fig. C3.4 Trazado de las capas 3-2-1

C3.4 CANALES LÓGICOS

Los canales lógicos fueron desarrollados en la tecnología DCCH IS-136 para organizar las PCS y otra información digital fluyendo a través de interfases de aire.

C3.4.1 CONFIGURACIÓN DE CANALES-LÓGICOS

Los canales lógicos son representados gráficamente en la figura C3.5. La figura muestra como el FDCCH consiste de muchos canales lógicos llevando información desde el sistema hacia el teléfono. El DCCH de reversa (RDCCH), transmitiendo información desde el teléfono hacia el sistema, consiste de un canal lógico.



* para versiones de sistemas futuros

DCCH = Canal de control digital
RACH = Canal de acceso aleatorio
BCCH = Canal de difusión

SPACH = SMS mensajería punto-a-punto, paginación Y canal de respuesta de acceso
SCF = Regeneración compartida del canal

F-BCCH = canal de difusión rápida
E-BCCH = Canal de difusión extendida
S- BCCH = Servicio de mensajería de difusión (para versiones futuras)
PCH = Canal de paginación
ARCH = Canal de respuesta al acceso
SMSCH = Canal de mensajería punto-a-punto SMS

Fig. C3.5 Configuración de canales lógicos

C3.4.2 PRINCIPIO DE OPERACIÓN

Los canales lógicos clasifican y prioritizan la información de señalización por el uso funcional. Los datos son entonces proyectados sobre un DCCH, el cual es un canal físico. Los canales físicos son las porciones reales del ancho de banda electromagnético, consistiendo de frecuencias y divisiones de tiempo. Los datos del canal lógico fluyen en el DCCH en ambas direcciones:

- Desde el sistema hacia el teléfono (downlink)
- Desde el teléfono hacia el sistema (uplink)

C3.4.3 FUNCIONES DE LOS CANALES LÓGICOS

El canal de difusión multiplexado (BCCH) mostrado en la figura C3.5 es designado para llevar información respecto a la configuración del sistema y las reglas que los teléfonos deben seguir en el acceso al sistema. Estos fundamentos de los canales lógicos son los siguientes:

- Canal de difusión rápida (F-BCCH). Lleva información que los teléfonos necesitan inmediatamente, tal como el ID del sistema e información de registro.
- Canal de difusión extendida (E-BCCH). Lleva información que no es de tiempo critico. tal como la lista de células vecinas.

El sistema utiliza mensajería punto-a-punto multiplexada SMS, paginación, y canal de respuesta de acceso (SPACH) mostrado también en la figura C3.5 para comunicarse con un teléfono especifico. Estos canales lógicos son los siguientes:

- Canal de servicio de mensaje corto (SMSCH). Lleva mensajería de PCS y programación y activación sobre-el-aire (OAA/P)-PCS la información es llevada en los canales lógicos en ambas frecuencias 800 y 1900 MHz.
- Canal de paginación (PCH). Lleva paginas del sistema hacia el teléfono.
- Canal de respuesta al acceso (ARCH). Proporciona respuesta del sistema a las preguntas del teléfono y la información de administración.

La tabla C3.2, delinea los canales lógicos.

Tabla C3.2. Descripción de los canales lógicos

Canal Lógico	Descripción
BCCH	Este es un canal multiplexado de enlace de bajada contenido de F-BCCH y E-BCCH
SPACH	Este es un canal multiplexado de enlace de bajada contenido de SMSCH, PCH, y ARCH
RACH	Este es un canal de enlace de subida sencillo con todos las ranuras de tiempo usadas por el acceso al sistema.
SCF	El campo SCF en el enlace de bajada es utilizado para proveer un mecanismo de prevención de colisión para el enlace de subida.

C3.5 MODO DORMIDO Y TIEMPO DE ESPERA

Los PCS utilizan el DCCH para proporcionar un modo dormido durante el cual los teléfonos pueden apagar mucha de su circuitería hasta que ellos necesiten despertar, en intervalos predeterminados, para recibir mensajería del sistema. Esta característica incrementa la vida de la batería, con lo cual incrementa el tiempo de espera de los teléfonos. El tiempo de espera, es aquel en el que un teléfono inalámbrico esta ocioso; esto es, el teléfono esta encendido, pero ninguna llamada esta siendo realizada o recibida.

C3.5.1 PRINCIPIO DE OPERACIÓN

Un teléfono ocioso se encuentra en el DCCH. El teléfono revisa si hay llamadas entrantes cada pocos milisegundos y luego entra en el modo dormido. Esto difiere de un teléfono que emplea un ACC, donde un teléfono ocioso debe vigilar el canal de control constantemente, desgastando de esta manera la batería.

Los mensajes del sistema recibidos por el teléfono pueden ser paginaciones(por cualquiera de los dos, una llamada de voz o mensajería de PCS) o mensajes de difusión (por ejemplo, actualizaciones respecto a cambios de célula o listas de vecinos) llevados en el DCCH de enlace de bajada. El teléfono necesita decodificar la información del enlace de bajada solamente en intervalos en sus ranuras de paginación predeterminadas o en las ranuras de difusión si la información de difusión cambia. De esta manera, el teléfono ha prolongado períodos del tiempo

en los cuales puede accionar abajo algo de su circuitería y dormir entre las oportunidades de la paginación

C3.5.2 CONSUMO DE CORRIENTE Y PERIODOS DORMIDOS

La figura C3.6 muestra el consumo de corriente de batería de DCCH contra ACC e indica los periodos del modo-dormido de los teléfonos en el DCCH. El punto del tiempo en el segmento del DCCH del área representativa de las ranuras de paginación predeterminadas.

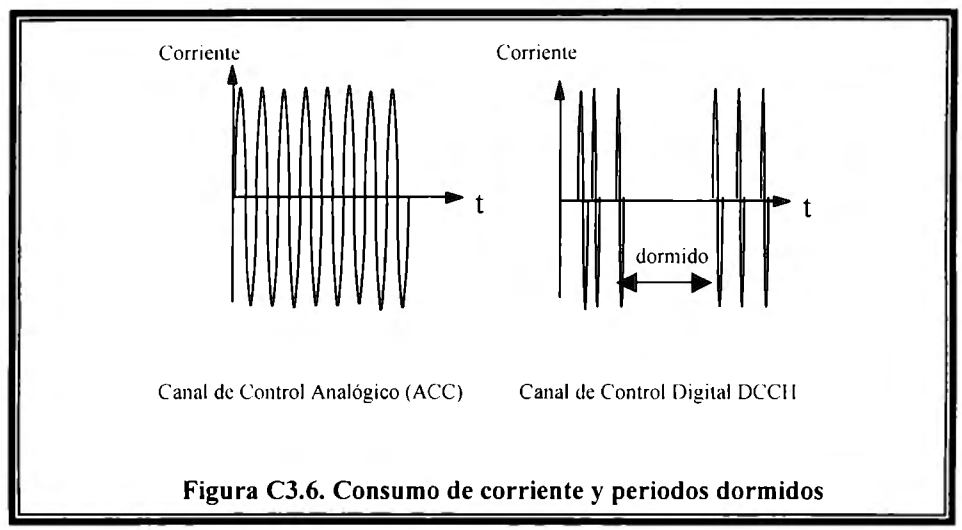


Figura C3.6. Consumo de corriente y periodos dormidos

C3.6 MENSAJERÍA DE PCS

La mensajería de los PCS es una característica SMS digital que permite a los teléfonos inalámbricos recibir páginas numéricas y mensajes de texto cortos. Esto permite a un dispositivo hacer el trabajo de dos, "pager" y teléfono. Los usuarios pueden recibir mensajes en sus pantallas de visualización del teléfono desde diferentes fuentes: computadoras, teléfonos, e-mail, correo de voz, y envío de texto.

Los PCS utilizan el DCCH y DTCs para entregar los mensajes alfanuméricos hacia y desde el teléfono inalámbrico. Los mensajes son enviados y recibidos vía un centro de mensaje, el cual es un nodo en la red inteligente inalámbrica. Los mensajes contienen una variedad de cualidades que controlan su entrega, almacenaje, y comportamiento de la exhibición.

C3.6.1 ARQUITECTURA DEL MENSAJE

Cada Mensaje de PCS originado en la red consiste de los siguientes tres elementos básicos:

- Dirección de la información. Le dice al sistema a cual teléfono el mensaje será entregado.
- Texto alfanumérico. Los caracteres que confeccionan el mensaje real de texto
- Atributos del mensaje. Le dice al teléfono cómo dirigir y exhibir el mensaje cuando este es recibido.

C3.6.2 TIPOS DE MENSAJES

La mensajería de PCS puede entregar mensajes "numéricos-callback" desde un teléfono y envía mensajes alfanuméricos vía modem y computadora. Ejemplos de mensajería de PCS incluyen el "paging" y a notificación de un nuevo mensaje de voz y mensajes de correo electrónico. Mensajes de hasta 239 caracteres pueden ser enviados sobre la interfase aérea.

C3.6.3 PRINCIPIO DE OPERACIÓN

Las características de mensajería de PCS utilizan una terminal de paginación dedicada. Cuando la red recibe un mensaje PCS, localiza el teléfono destino y entrega el mensaje. El teléfono notifica al usuario con un icono de mensaje, un "bip", o ambos. El mensaje puede entonces ser desplegado y leído. Si los usuarios salen de una área de mensajería de PCS, la red guarda cualquier mensaje hasta que ellos regresen. La red tratara repetidamente de dejar un mensaje hasta que el teléfono este disponible para recibir este.

C3.6.4 GENERACIÓN DEL MENSAJE

Las siguientes entidades pueden ser utilizadas para la generación de mensajes PCS:

- Establecimiento de una red desde los terminales de paginación existentes
- Unidad de respuesta de voz
- Servicio de entrega de texto de operador
- Modem dial-up
- E-mail gateway
- Fuente de información de datos
- Sistema de correo de voz

La figura C3.7 muestra un esquema de mensajería de teleservicio en el cuál un mensaje es formulado en una computadora personal (PC) y enviado hacia el teléfono o dispositivo receptor del mensaje. Las pantallas de visualización del dispositivo o teléfono difieren dependiendo del modelo y fabricante, pero todos ellos muestran él numero de mensajes nuevos.

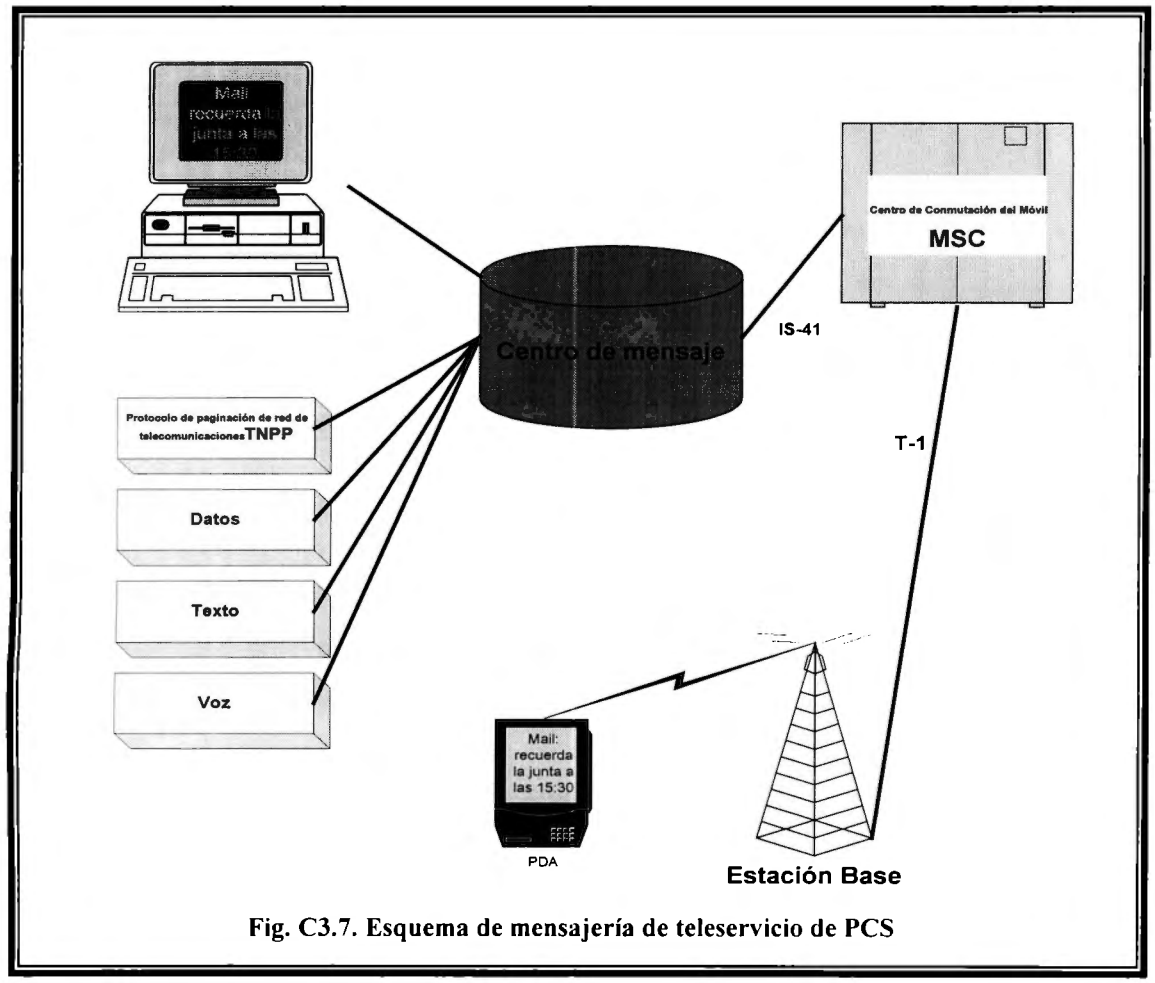


Fig. C3.7. Esquema de mensajería de teleservicio de PCS

C3.6.5 ENTREGA DEL MENSAJE

La mensajería de los PCS esta designada para operar en prácticamente, situaciones diarias.

- Encendido.
- Teléfono ocupado
- Apagado.
- Correo de voz.
- Roaming

C3.7 RELACIONES JERÁRQUICAS DE LA CÉLULA

Los sitios de las células han existido tradicionalmente como macrocélulas en torres que cubren áreas por arriba de varios miles en diámetro. Las macrocélulas son típicamente células publicas, sirviendo a todos los usuarios de teléfonos inalámbricos. La tecnología TDMA DCCH IS-136 habilita el uso de muchas células pequeñas llamadas microcélulas. Las microcélulas típicamente proporcionan características de WOS para teléfonos específicos dentro de un edificio privado o un ambiente de campus.

C3.7.1 COBERTURA DE LA JERARQUÍA DE CÉLULA

La cobertura combinada de ambas; macrocélulas y microcélulas son llamadas cobertura de jerarquía de célula, con la microcélula se crea un segundo nivel de cobertura bajo el nivel existente. Aunque las macrocélulas son generalmente publicas y las microcélulas privadas, ellas pueden invertir sus roles.

La figura C3.8, muestra una microcélula de sistema privado dentro de una macrocélula publica.

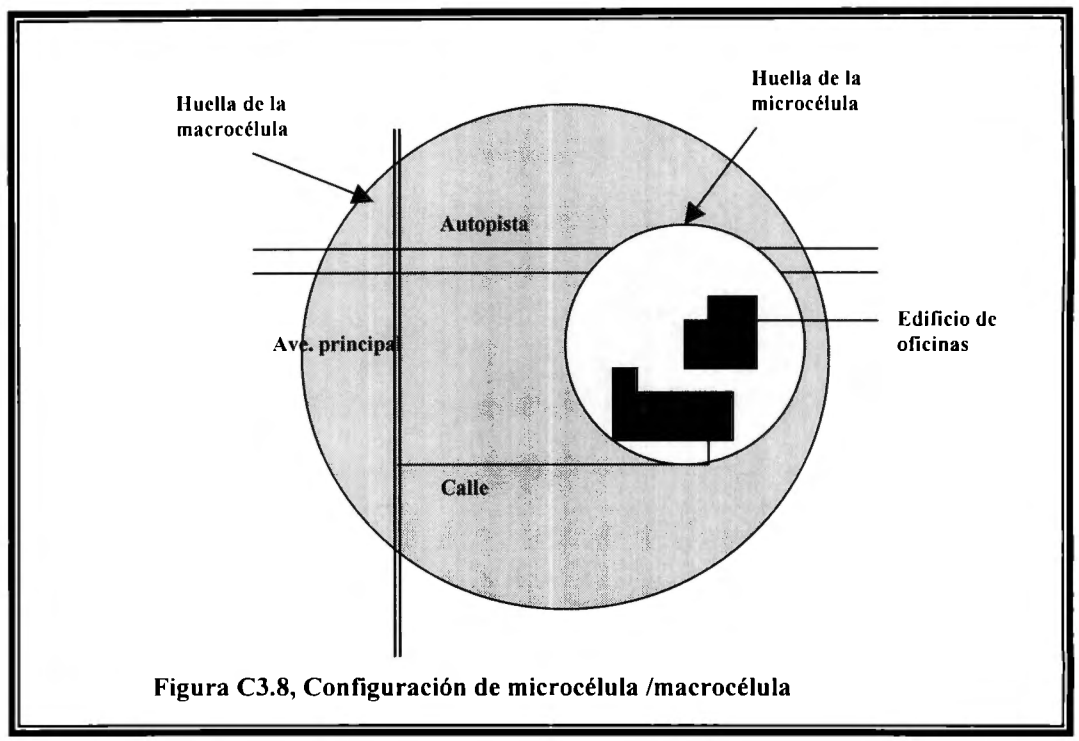


Figura C3.8, Configuración de microcélula /macrocélula

C3.7.2 ESTRUCTURAS JERÁRQUICAS DE LA CÉLULA.

En ambientes PCS, una área geográfica debe ser cubierta por una mezcla de macrocélulas y microcélulas así como sistemas privados y públicos. Un teléfono PCS por lo tanto debería determinar el canal de control más conveniente en el que proporcionara servicio, incluso si la fuerza de la señal de una célula vecina no es la más alta siendo recibida por el teléfono, pero es de un nivel suficiente para proporcionar calidad de servicio. Los PCS utilizan estructuras de jerarquía de célula (HCS) para lograr esto por la identificación de células vecinas como preferidas, regulares, o no-preferidas.

- Célula vecina preferida. Tiene una alta preferencia. El teléfono reelige esta incluso si la fuerza de la señal es baja en comparación con la célula que le sirve actualmente. El criterio principal es la calidad del servicio.
- Célula vecina regular. Tiene la preferencia más alta en segundo lugar. El teléfono reelige esta si la fuerza de la señal de la célula es más grande que la que le sirve actualmente (más un valor de histéresis) y no hay disponible una célula preferida.

- Célula vecina no-preferida. Esta tiene la preferencia más baja. El teléfono reelige esta solamente si la fuerza de la señal de la célula que sirve actualmente comienza a ser insuficiente para proporcionar servicio y la fuerza de la señal del vecino no preferido es más grande que esta ultima. (Más un valor de histéresis).

C3.7.3 PRINCIPIO DE OPERACIÓN

Las HCSs permiten al DCCH identificar y designar una célula vecina como preferida, regular, o no-preferida. Un teléfono PCS utiliza esta información jerárquica para reelegir a una célula vecina particular sobre otras basadas en el tipo de relación definida entre la célula que sé esta utilizando (Célula servidora) y la célula vecina adyacente. Cada destinación de células vecinas dictamina cual tipo de algoritmo el teléfono utilizara cuando este considere la célula como una candidata a re-seleccionar.

La figura C3.9 muestra la re-selección basada en la designación del tipo de célula HCS.

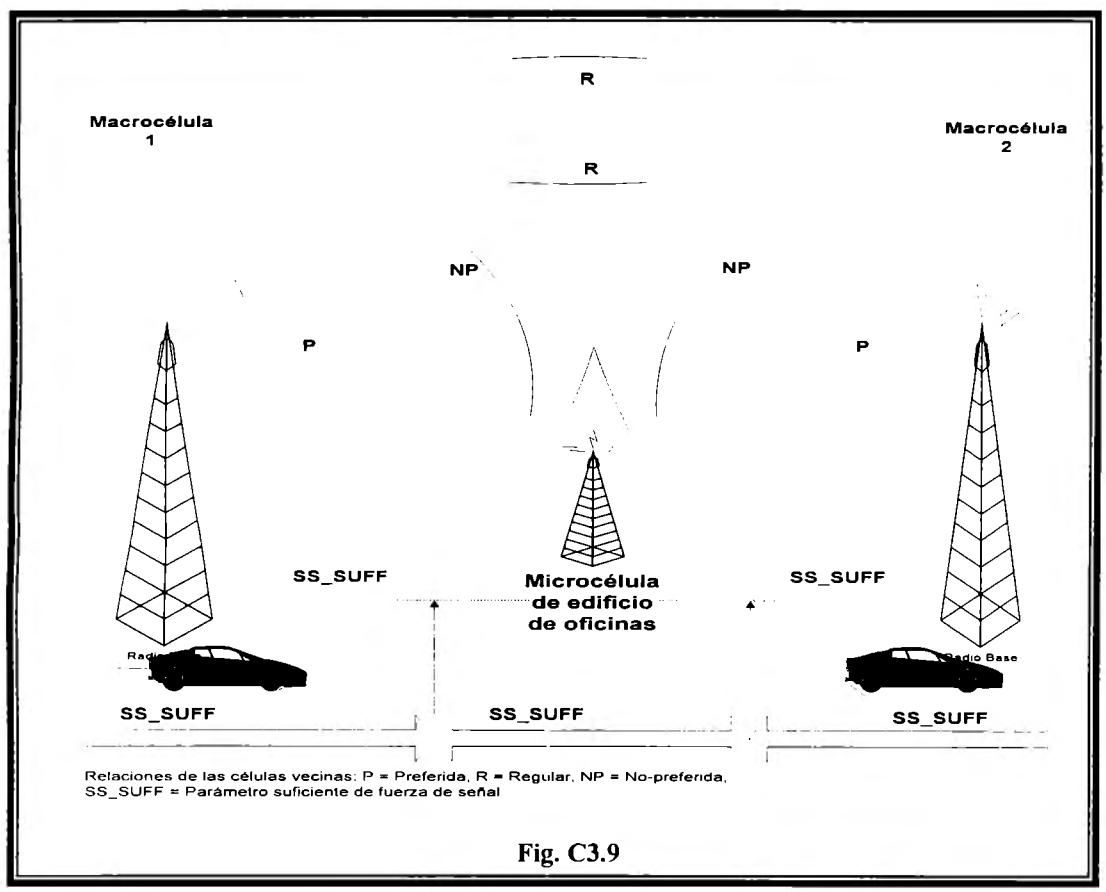


Fig. C3.9

C3.8 SISTEMAS PUBLICO, PRIVADO Y RESIDENCIAL

Los telefotos PCS pueden comportarse de manera diferente de acuerdo con el tipo de sistema que proporciona servicio al usuario. Por ejemplo, los teléfonos que proporcionan solamente servicio básico no podrán re-seleccionar o ponerse en células privadas, de tal modo improvisando su tiempo de servicio. Similarmente, los teléfonos que proporcionan servicio en un sistema residencial, como los PBS, podrán presentar diferentes rutinas de búsqueda para encontrar su sistema local.

C3.8.1 PRINCIPIO DE OPERACIÓN

La siguiente figura C3.10, muestra algunas configuraciones de sistemas de red.

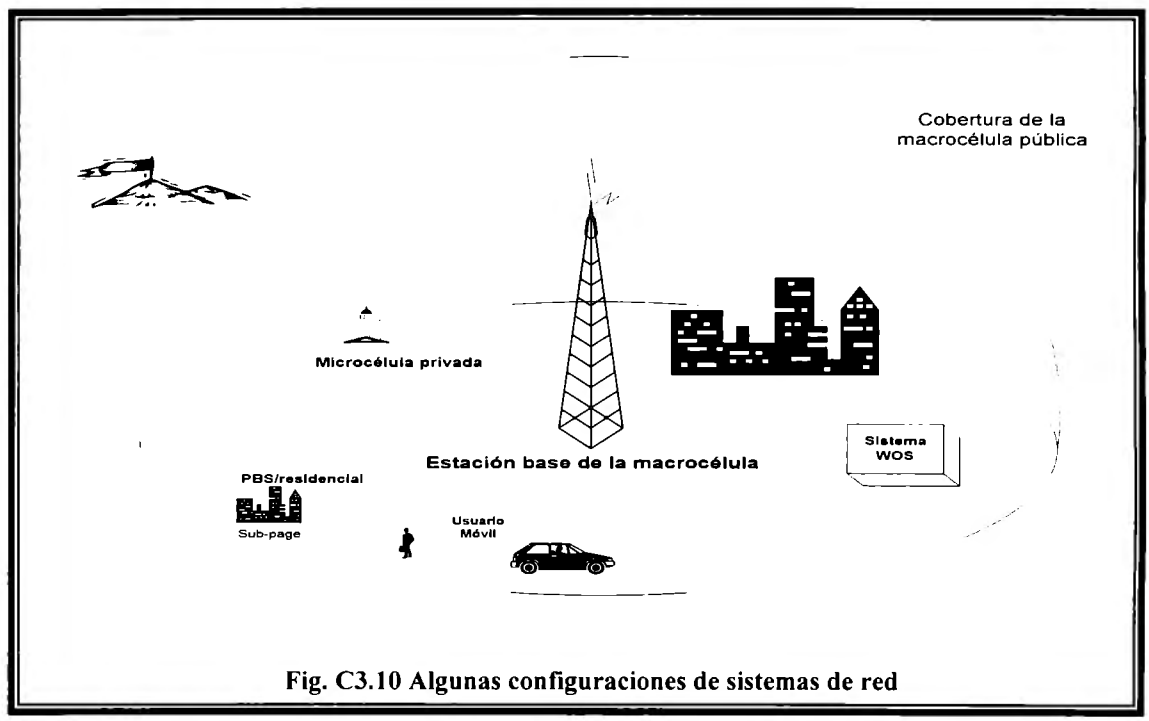


Fig. C3.10 Algunas configuraciones de sistemas de red

C3.8.2 TIPOS DE REDES

Las designaciones para los principales tipos de redes y los subsistemas incluyen lo siguiente:

- Pública. Esta se refiere a las células que proporcionan el mismo servicio celular básico a todos los clientes.
- Privada. Estas células proporcionan servicios especiales para un grupo predefinido de privado o solamente clientes WOS, y que no soportan uso público de esa célula. La designación privada es empleada para sistemas de compañías dentro de edificios con características especificas.
- Semi-privada. Es un subtipo, estas células proporcionan servicio básico a todos los clientes y también proporcionan servicios especiales para un grupo predefinido de clientes privados. Un ejemplo podría ser una célula proveyendo servicio para un sistema WOS así como los usuarios públicos.

- Residencial. Estas células proporcionan servicios especiales para solamente un grupo de clientes residenciales, y no soportan uso publico de la célula. El PBS que permite a un teléfono celular comportarse como un teléfono inalámbrico es clasificado como sistema residencial.
- Semi-residencial. Es un subtipo, estas células proporcionan servicio básico a todos los clientes y también proporcionan servicios especiales para un grupo predefinido de clientes residenciales. Este tipo es empleado en un vecindario donde la macrocélula pública es también proveedor de servicio celular residencial.
- Autónoma. Estas son células que difunden un DCCH en la misma área geográfica como otros sistemas DCCH pero no están listados como un vecino en la lista de vecinos del sistema público. Ejemplos de sistemas autónomos incluyen el PBS y sistemas privados que no son coordinados con el sistema público. Los teléfonos deben presentar algoritmos de búsqueda de frecuencia especial para encontrar células autónomas.

C3.9 OTRAS TECNOLOGÍAS DE APOYO PARA LOS PCS

W-OFDM (Wide-band Orthogonal Frequency Division Multiplexing)
Es un esquema de transmisión que codifica la información en múltiples radio frecuencias simultáneamente. Dando como resultado, mayor seguridad y mayor velocidad. Este método como otros codifica los datos dentro de una señal de radio frecuencia (RF). Transmisiones convencionales como AM/FM envían solamente una señal a la vez sobre una frecuencia de radio, mientras que OFDM envía una señal de alta velocidad concurrentemente sobre frecuencias diferentes. Esto nos permite hacer un uso muy eficiente del ancho de banda y tener una comunicación robusta al enfrentar ruido y reflejos de señales.

La tecnología OFDM parte una señal de alta velocidad en decenas o centenas de señales de menor velocidad, que son transmitidas en paralelo. Esto crea un sistema altamente tolerante al ruido, al mismo tiempo es muy eficiente en el uso del ancho de banda y por lo tanto permite una amplia cobertura de área punto a punto y multipunto.

Actualmente existen equipos con la capacidad de transmitir desde 1.5Mbps hasta 30Mbps en 25MHz de ancho de banda y pronto se estarán produciendo equipos que superaran velocidades de 100Mbps. Adicionalmente a la velocidad, se cuenta con opciones de seguridad que hacen virtualmente imposible descifrar la señal que se transmite.

Los equipos con tecnología OFDM ayudan a las empresas a evitar los altos costos de instalación de cable, y a eliminar rentas mensuales o cargos por licenciamiento.
Pueden ser una solución ideal en distancias moderadas para redes de información punto a punto, multipunto, acceso de alta velocidad a Internet, extensiones de LAN/WAN, Videoconferencia, Telefonía, Telemetría, Control, Etc.

W-OFDM es la base del estándar IEEE 802.11a que a su vez es la base para el estándar propuesto IEEE 802.16.
Sus principales características son:

- Ancho de Banda: 30Mbps
- Altamente inmune a interferencias
- Punto-a-Punto, 8 a 10Km
- Multi-Punto, 3 a 5Km

- Próximamente: 45Mbps, 90Mbps, 155Mbps

Esta patentada por: Wi-Lan, USA 5,282,222, CANADA 2,064,975

MC-DSSS (MultiCode Direct Sequence Spread Spectrum)

MC-DSSS (esta patentado también por Wi-Lan) Es la tecnología de radio de espectro disperso muy eficiente. Multiplica hasta por 10 la capacidad de los sistemas DSSS tradicionales.

Es la base para la tercera generación (3G) de redes móviles, incluyendo teléfonos celulares.

Sus principales características son:

- Ancho de Banda: Hasta 12Mbps
- Fácil Sincronización
- Altamente inmune a interferencias
- Punto-a-Punto, 40Km+
- Multi-Punto, 10Km

C3.10 PRACTICAS Y ESTÁNDARES EN EL AMBIENTE MÓVIL.

El uso de Internet y servicios accesibles a través de este, por dispositivos tales como teléfonos móviles están siendo realidad en los últimos años. Los modelos de la autentificación que han sido utilizados se han heredado en gran parte de las contrapartes alámbricas. Pero existen, de todos modos, algunos métodos para hacer autentificación en el Internet móvil que son distintos a los utilizados en redes alámbricas, e inclusive muchas veces no es posible utilizarlos para otra parte.

Métodos simples.

Los mismos medios de autentificación que se han utilizado en el pasado también se están utilizando en los sistemas móviles. En los servicios basados en WAP la manera más comúnmente utilizada para autentificarse es por medio del "username" y la palabra de paso. En algunos servicios la autentificación se puede también basar en el número de teléfono de GSM, es decir el número de MSISDN, del usuario que tiene acceso al servicio de WAP. Este método de autentificación es comparable a utilizar el IP-address de una máquina asociada al Internet como los argumentos de la autentificación - las mismas deficiencias se aplican a ambos casos

Productos en el mercado.

Hay varios vendedores de productos que proporcionan facilidades para la autentificación y la autorización en el ambiente móvil. Muchos de ellos están proporcionando soluciones propietarias. Esto es en parte debido al hecho de que muchos de los estándares que hay, siguen siendo muy jóvenes y la industria tiene dificultad en continuar con el paso.

En un futuro cercano la iniciativa móvil de las transacciones electrónicas, formada por los fabricantes de equipo móviles, Nokia, Ericsson y Motorola, será probablemente un cuerpo que estandarice significativamente en este sector. Aunque todavía, los resultados de su trabajo no están disponibles. El MeT de todos modos ha declarado que trabajará junto con las iniciativas de

estandardización existentes e iniciativas de especificación, "cuando sea apropiado", e intentará lo mayormente posible mantener la misma dirección del trabajo que se ha tomado ya en el foro de WAP, con las especificaciones de WTLS y de WIM [ver referencia 34 del capitulo 2].

HST

HST es un PKI en Finlandia. La idea es que publican todos una tarjeta inteligente que contenga un certificado y la llave privada asociada. La tarjeta se piensa utilizar para autentificar a los ciudadanos cuando se ocupan de autoridades del gobierno. Un requisito para usar la tarjeta, es que tiene que haber un lector de tarjetas inteligente y una parte del software que sabe como utilizar los servicios proporcionados por la tarjeta. Utilizar esta tarjeta por ejemplo, en un teléfono móvil no es actualmente posible. Hay también un mecanismo de "escrow" de llave[ver referencia 30 del capitulo] puesto en lugar en el PKI, que hace la seguridad del sistema cuestionable.

Seguridad de RSA - SecurID.

La seguridad de RSA SecurID (antes un producto dinámico de seguridad), es un sistema para los usuarios de autenticidad. Se basa en el uso de un código NIP (algo que uno sabe) y de un dispositivo, el autentificador que produce en intervalos regulares un nuevo código "token" (algo que el usuario sabe) que el usuario puede utilizar como la contraseña. En el extremo del servicio, hay por supuesto un servidor, que puede validar las contraseñas y los NIPs recibidos de los usuarios [ver referencia 38 del capitulo 2].

Este sistema proporciona un nivel absolutamente alto de seguridad, puesto que se basa en dos factores de la autentificación, el NIP y los "tokens" del autentificador. Las desventajas de este son que utiliza una tecnología algo cerrada - los autentificadores tienen que ser comprados de seguridad de RSA, así como el software del servidor. El usuario también tiene que llevar con él una cierta pieza adicional de equipo, el autentificador, que no responde a ningún propósito con excepción de producir los valores de la contraseña. La tecnología realmente se piensa solamente para el uso dentro de una empresa como la solución de autentificación.

Sonera SmartTrust - Cliente de Seguridad del Sim.

Sonera SmartTrust está ofreciendo una solución de seguridad para WAP. Su producto de cliente de seguridad SIM es un producto basado en la caja de herramientas de SIM, que pone básicamente las funciones en ejecución, por ejemplo, lo tratado en la sección 2.4.2 se especifica para ser proporcionado, pero de una manera propietaria. Parte del paquete del producto es el software del servidor, llamado " El servidor de la seguridad SmartTrust " que implementa las facilidades de PKI que los abastecedores de servicio necesitan [ver referencias 39, 33 del capitulo 2].

La idea detrás del ofrecimiento de SmartTrust es buena. Esta es ideal para ambientes cerrados, tipo Intranet de ambientes de uso WAP, pero no es satisfactoria a menudo para ambientes de operación amplia, puesto que un prerrequisito para utilizar esto es el cambio de la tarjeta GSM SIM, a medida que el producto en parte resida en este. Para las corporaciones que están dispuestas a pasar con este apuro, la solución debe trabajar agradablemente.

Lo malo sobre el producto, es que es propietario. El WIM es básicamente la manera estándar de poner las mismas funciones en ejecución y quizás no se tiene ninguna duda que sea una solución dominante en futuro.

Caso: Aplicación en un Banco

Definamos a SIABI como el Servicio de Internet de las Actividades Bancarias y de Inversión ofrecido por un Banco. Dicho banco cuenta con Internet, voz e interfaces de WAP para los servicios. En el servicio usuarios son autentificados por su nombre de usuario, que es asignado aleatoriamente a cada usuario y una palabra de paso de una sola vez de cuatro dígitos. La lista de las palabras de paso de una sola vez se entrega a los usuarios vía postal en un sobre sellado. En la misma hoja con las palabras de paso de una sola vez también un grupo de códigos de autentificación. Se Solicita al usuario entrar en ciertos de ellos siempre que él solicite el pago, la inversión u otra transacción del dinero ha ser realizada en el servicio.

El mismo modo de autentificación se utiliza para todos los canales del acceso del SIABI: Internet, voz y WAP. La comunicación en la versión Internet es protegida por el SSL de 40 dígitos binarios. El modo de acceso de WAP no es asegurado por WTLS - el uso del Banco de los usuarios de WAP poseen "dial-in" módem para acceder al servicio, así que la conexión es asumida segura. El banco actúa también como parte de confianza, en certificar transacciones del pago de eCommerce a e-Stores en el Internet. Proporciona un recurso de pago para las compras en el servicio SIABI, después de lo cual SIABI notifica al e-Store de la salida del pago. La información sobre la recepción del pago es firmada por el banco, que permite que el e-Store confíe en que el dinero se ha transferido a su cuenta.

C3.11 INTERFACES AÉREAS Y MODELOS DE SEGURIDAD EN PCS.

En los últimos años, las PCS han estado siendo ampliamente desplegadas en Europa y Estados Unidos. La amplitud de servicios y ubicación de los PCN requiere enriquecer la seguridad para satisfacer los requerimientos de privacidad, autentificación y transacciones, porque las interfaces aéreas exponen el contenido de las comunicaciones. Tal exposición hace difícil mantener la confidencialidad y el control de fraudes. Por lo que los requerimientos mínimos de seguridad para PCS, son dados como:

1. Autentificación de subscriptores para impedir fraudes y garantizar la disponibilidad de los servicios contratados: En los sistemas de telefonía tradicional la línea alámbrica automáticamente identifica el subscriptor. En los ambientes inalámbricos, el subscriptor tiene que suministrar su identidad al servidor de red para verificaciones necesarias, aun así, espera que su identidad sea protegida. Existen varios caminos para implementar esta protección. A la unidad móvil se le puede asignar un ID temporal para proteger la identidad valida del subscriptor o la identidad puede ser enviada en forma de texto cifrado.

2. Autentificación mutua para impedir ataques de "playback": El requerimiento es eliminar el intercambio que usuarios ilegales hacen en las llamadas telefónicas fraudulentas. Un problema similar es la impersonación de los servidores de red por un intruso. Las técnicas de criptografía modernas pueden ser empleadas para implementar autentificación mutua entre el subscriptor y el servidor de red.

3. No-repudiación del servicio para asegurar al subscriptor que no pueden ser negados servicios actualmente recibidos: este problema inclusive engloba transacciones precisas. Firmas digitales pueden ser empleadas para permitir esta meta, pero su complejidad computacional es un factor limitante.

4. Transmisión de mensajes en texto cifrado par impedir ataques de "eavesdropping": ya que las señales de radio transmitidas por el aire pueden ser fácilmente interceptadas, los mensajes deben ser transmitidos en forma de texto cifrado. En consecuencia, una sesión de llave común debe ser acordada durante la autentificación.

En principio, estos requerimientos de seguridad pueden ser permitidos por protocolos de autorización empleando técnicas de llave publica. Las llaves publicas pueden ser empleadas para verificar las identidades de las entidades en ambos fines de enlaces inalámbricos y para establecer una llave de sesión secreta entre ellas para comunicaciones seguras subsecuentes. Sin embargo, los sistemas de llave publica no han sido ampliamente adoptados porque estos son computacionalmente intensos.

Protocolos con rasgos de seguridad para redes alámbricas no pueden ser directamente aplicados a los propios de los sistemas de comunicación personal con ciertas características de las redes inalámbricas, incluyendo la carencia de asociación física entre subscriptor y la red, limitaciones propias de las baterías, y fuerzas económicas en la complejidad del hardaware de las unidades móviles. Demandas de complejidad computacional y gran ancho de banda pueden en primer instancia ser requeridos para un aceptable verificación de retardos. Muchos protocolos de autentificación para PCS han sido propuestos empleando diferentes técnicas y enfatizando diferentes requerimientos de las PCS anteriormente mencionados. Sin embargo, estos protocolos comparten características comunes. La mayoría de los sistemas asumen que la unidad móvil y su red local comparten una llave común para autentificación mutua cuando una red externa es visitada, algunos parámetros de seguridad son derivados desde la llave común y enviados desde la red local del subscriptor hacia la red visitada. Esta ventaja reduce el riesgo de comprometer la llave de autentificación, y si los parámetros de seguridad son comprometidos, nuevos parámetros pueden ser calculados para remplazar los viejos.

Resumen de interfaces aéreas y sus modelos de seguridad

PCS y celulares Norte Americanos y Europeos soportan variedad de protocolos de interfaces aéreas.

- Sistemas Telefónicos Móviles Avanzados (Advanced Mobile Phone Systems AMPS)
- IS-95 Accesos Múltiple por División de Códigos (Code División Múltiple Access CDMA)
- Sistema Global para Comunicaciones Móviles (Global System for Mobile communications (GSM)
- Sistemas de Comunicación de Acceso Personal (Personal Access Communications Systems PACS)
- Protocolos PCS-2000
- IS-136 Accesos Múltiples por División de Tiempo (Time División Múltiple access TDMA)
- Sistemas CDMA de banda ancha (W-CDMA)

Existen 4 modelos de seguridad asociados con estas interfaces aéreas.

1. MIN/ESN (Mobile Identification Number / Electronic Serial Number). Este es el modelo de seguridad original. Cuando un usuario hace o recibe una llamada, el teléfono transmite su MIN y ESN hacia la red. La red Checa una lista de unidades robadas. Si el teléfono no esta en la lista, entonces la llamada es procesada. Este método ha sido actualizado (updated) en el protocolo IS-41 para emplear la red SS7.
2. Datos Secretos Compartidos (Shared Secret Data SSD). Este esquema emplea una llave de autentificación común en las unidades telefónicas móviles y en la red. En el momento en que la unidad telefónica es puesta en servicio, la llave es registrada dentro de la unidad y también dentro del registro de ubicación local de red HLR. De la llave, un dato secretos compartido es derivado y empleado para autentificar el aparato telefónico y establecer privacía (vía una llave de sesión).
3. "Triplets" de seguridad (basado en "tokens"). Los diseñadores de PCS buscaban sistemas de seguridad que estuvieran bajo control del proveedor de servicios y que no fueran demandantes computacionalmente o en ancho de banda. El método basado en "token" satisfizo este requerimiento. Cuando una unidad móvil vaga dentro de la red, un mensaje es enviado hacia el sistema local preguntando por una serie de "triplets" (pregunta única, respuesta a la pregunta y una llave privada de voz derivada de la pregunta) cada llamada que es hecha o recibida emplea un "triplet". Este es el método empleado en GSM.
4. Llave Publica. Cuando una PCN emplea criptografía de llave publica, es publicada la llave publica de la red. La llave publica de la unidad móvil es también conocida para la red. Las llaves publicas son empleadas para autentificar la unidad móvil y para comunicarse con la red. Un anonimato completo es garantizado ya que solo la red puede desencriptar los mensajes. La privacidad de datos y voz son posibles por el empleo de sistemas de llave publica en un ambiente de transmisión segura. Este método es empleado en los sistemas de comunicación de acceso personal (PACS). La tabla 1 resume la relación entre interfaces aéreas y modelos de seguridad.

	MIN/ESN	SSD	Basado en Token	Llave publica
AMPS	X	X		
CDMA	X			
GSM		X	X	
PACS		X		X
PCS-2000		X	X	
TDMA		X		
W-CDMA		X		

Tabla C3.3 Relación entre interfaces aéreas y modelos de seguridad.

REFERENCIAS DE ANEXOS

[1] D.E. Denning and G. M. Sacco. "Timestamps in key distribution protocols". Communications of the ACM, 24(8): pp. 198-208, 1981.

[2] G.J. Simmons. "How to (selectively) broadcast a secret." In Proceedings of the IEEE CS Symposium on Research in security and Privacy, pp. 108-113, 1985

[3] M.Burrows, M. Abadi, and R, Needham. "A logic of authentication". ACM Transactions on Computer Systems, 8(1): pp. 18-36, Febrero 1990.

[4] M. Debbadi, M. Mejri, N. Tawbi, and I. Yahmadi. "A new algorithm for the automatic verification of authentication protocols: From specifications to flaws and attack scenarios. In DIMACS Wokshop on Design and Formal Verification of Security Protocols, 1997.

[5] C. A. Meadows. "Formal verification of ceyptographic protocols", A survey. In Advances in Cryptology - ASICRYPT'94, pp. 135-150. Springer-Verlang, 19995.

[6] S. Gritzalis, N. Nikitakos, and P. Georgiadis. "Formal methods for the analisys and design of cryptographyc protocols: A state-of-the –art review". In proceedings of the IFIP Working Conference on Communications and Multimedia Security, Vol. 3. pp. 119-132, 1997.

[7] L. Gong, R. Needham, and R. Yahalom. " Reasoning about belief in cryptographic protocols". In proceedings of the IEEE CS Symposium on Research in Security and Privacy, pp. 234-248. 1990.

[8] P. Bieber. "A logic of communication in a hostile environment." In Proceedings of the Computer Security Foundations Workshop III, pp. 14-22, Junio 1990.

[9] P.F. Syverson. " Formal semantics for logics of cryptographic protocols". In proceedings of the Computer Security Foundations Workshop III, pp. 32-41, Junio 1990.

[10] P. V. Rangan. " An axiomatic basis of trust in distributed systems". In Proceedings of the IEEE CS Symposium on Research in Security and Privacy, pp. 204-211, Abril 1988.

[11] P. F. Syverson and P. C. van Oorschot. "On unifying some cryptographic protocol logics. In Proceedings of the IEEEE CS symposium on research in Security and Privacy, pp. 14-28, 1994.

[12] A. Huima and T, Aura. "Using a multimodal logic to express conflicting interests in security protocols. In DIMACS Wokshop on Design and Formal Verification of Security Protocols, 1997.

[13] R. Kemmerer, C. Meadows, and J. Millan. "Three Systems for cryptographic protocol analysis". Journal of Cryptology, 7(2): 79-130, 1994.

[14] A. Abadi, M. Burrows, and R. Needham, "A Logic of Authentication", Report 39, Digital Equipment Corporation Systems Research Center, Palo Alto, California, February 1990.

[15] A. Abadi, M. Burrows, and R.Needham, "A Logic of Authentication", ACM Transactions on Computer Systems, Vol. 8, No. 1, February 1990, 18-36.

[16] C. Boyd and W. Mao, "On a Limitation of BAN Logic", Advances in Cryptology EUROCRYPT '93, Lofthus, Norway, May 1993.

[17] Paul C. van Oorschot, "An Alternate Explanation of two BAN-logic 'failures'", Advances in Cryptology - EUROCRYPT '93, Lofthus, Norway, May 1993, 443-447.

[18] Needham, R. M., and Schroeder, M. D. "Using encryption for authentication in large networks of computers". Commun. ACM 21, 12 (Dec. 1978) pp. 993-999.

[19] CCITT. CCITT draft recommendation X.509. "The directory-authentication framework, version 7". CCITT, Gloucester, Noviembre 1987.

[20] Halpern, J. Y., And Moses, Y. O. "Knowledge and common knowlwdge in a distributed environment". In Proceedings of the 3rd ACM Conference on the Principles of Distributed Computing (Vancouver, British Columbia, Agosto 1984), ACM, New York, 1984, pp. 480-490.

[21] DeMillo, R. A., Lynch, N. A, and Merrit, M. J. "Cryptographic protocols". In Proceedings of the 14 th ACM Symposium on the theory of computing. ACM, New York, 1982, pp. 383-400.

[22] Halpern, J. Y., Moses Y. O. and Tuttle, M. R. " A knowledge-based analysis of zero knowledge" (preliminary report). In proceedings of the 20th ACM Symposium on theory of Computing. New York, 1988, pp. 132-147.

[23] Merrit, M. J. And Wolper, P. L. "States of knowledge in cryptographic protocols". Draft.

[24] D. M. Nessett. " A critique of Burrows, Abadi and Nedham logic". Operating System Review 24 (2): pp. 35-38, Abril 1990.

[25] M. Abadi and M. R. Tuttle. " Asemantics for logic of authentication. In Proceedings of the Tenth Annual ACM Symposium on Principles of Distributed Computing, pp. 201-216, 1991.

[26] P.F. Syverson. " A new look at an old protocol." Operating Systems Review, 30(3): pp. 1-4, 1996.

[27] J. Alves-Foss and T. Soule. A weakest precondition calculus for analysis of cryptographic protocols. In DIMACS Workshop on Design and Formal Verification of Security Protocols, 1997.

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY
CAMPUS ESTADO DE MÉXICO

PROTOCOLOS DE AUTENTIFICACIÓN SEGURA PARA COMUNICACIÓNES EN AMBIENTES PCS

TESIS QUE PRESENTA

DAVID HIGUERA ROSALES

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

MCCR 95, ITESM-CEM

JULIO, 2002