

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY**

CAMPUS MONTERREY

**PROGRAMA DE GRADUADOS DE LA DIVISION DE
COMPUTACION, INFORMACION Y COMUNICACIONES**



**ESQUEMA DE CONTROL DE ADMISION DE LLAMADAS
PARA REDES DE BANDA ANCHA USANDO UNA
APROXIMACION GAMMA**

**TESIS
PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO ACADEMICO DE
MAESTRO EN CIENCIAS
EN TECNOLOGIA INFORMATICA**

Catalina Pico Ruiz

MONTERREY, N. L.

MAYO DEL 2000

**INSTITUTO TECNOLOGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY**

CAMPUS MONTERREY

**PROGRAMA DE GRADUADOS DE LA DIVISION DE
COMPUTACION, INFORMACION Y COMUNICACIONES**



**ESQUEMA DE CONTROL DE ADMISION DE LLAMADAS
PARA REDES DE BANDA ANCHA USANDO UNA
APROXIMACIÓN GAMMA**

TESIS

Presentada como requisito parcial para
obtener el grado académico de

**MAESTRO EN CIENCIAS EN
TECNOLOGIA INFORMATICA**

Catalina Pico Ruiz

Mayo 2000

Agradecimientos

Agradezco al Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), por brindarme la oportunidad de llevar a cabo mis estudios de posgrado. Así como al Centro de Electrónica y Telecomunicaciones (CET) y todos sus maestros por su apoyo incondicional.

A mi asesor, Ph.D. Ramón Rodríguez Dagnino, le expreso mi agradecimiento por su incondicional apoyo y orientación que me brindó durante el curso de mi maestría.

Agradezco al Ph.D. Juan A. Nolazco Flores y M. Sc. Artemio Aguilar Coutiño, por aceptar ser mis sinodales y por sus valiosas observaciones.

También agradezco a mis compañeros y amigos del CET, Arturo, Francisco, Calix, Samuel, Elsa, Salvador, Fermín, Isaura, Alejandro, Anabel, Patty, Manuel, Betty, Alejandro Gibran y Ricardo, por todo el apoyo, amistad y cariño que me brindaron.

Además, deseo expresar un agradecimiento muy especial a mis mejores amigos Alejandro Molina, Edgar Seañez y Enrique Arcos. Gracias por todo su cariño, apoyo incondicional, por haber estado siempre a mi lado y por haberme hecho sentir que cuento con una verdadera familia aquí en México. Gracias mis Ñañitos por su cariño y amistad.

Finalmente, quiero agradecer a las personas más maravillosas del mundo: mis Padres, que gracias a su apoyo incondicional moral y económico he podido culminar esta etapa de mi vida. Gracias Papitos adorados por todo su cariño y amor.

A las personas que más quiero en este mundo, que son mi guía, estímulo y fortaleza para seguir adelante,
A mi Papito que siempre fue mi ejemplo de rectitud, tenacidad y lucha constante
A mi Mamita que siempre ha sido fuente de cariño, ternura y amor incondicional
A ellos con todo mi cariño y amor les dedico esta Tesis

Indice

Capítulo I: Introducción	1
1.1 Objetivo	2
1.2 Justificación	2
1.3 Descripción de la tesis	3
Capítulo II: Modelamiento y Características de tráfico	
para redes de banda ancha	5
2.1 Introducción	5
2.2 Modelo General de las Fuentes de Tráfico	6
2.2.1 Arribos Poisson y Procesos de Markov	8
2.3 Calidad de Servicio	11
2.3.1 Clases de QoS	13
2.3.2 Descriptores de Tráfico	15
2.4 Modelamiento de Paquete de Voz	16
2.5 Caracterización del Tráfico de Vídeo	19
2.6 Modelo de Tráfico Bursty	20
2.7. Control de Admisión	21
2.8 Distribución Estadística de Recursos	24
Capítulo III: Control de Admisión usando aproximación Gamma	26
3.1 Introducción	26

3.2	Caracterización del Tráfico	27
3.3	Multiplexaje Estadístico	29
3.4	Aproximación Gamma	34
<i>Capítulo IV: Análisis de Resultados</i>		47
4.1	Introducción	47
4.2	Aproximación Gamma a la Función Binomial	48
4.3	Probabilidad de Pérdida P_L	50
4.4	Probabilidad de Saturación ϵ	52
4.5	Control de Admisión	53
<i>Capítulo V: Conclusiones y Recomendaciones</i>		56
<i>Apéndice A: Aproximación Gaussiana</i>		58
<i>Apéndice B: Variables Aleatorias</i>		62
B.2	Distribución Gamma	64
B.3	Distribución Binomial	67
<i>Bibliografía</i>		69

Lista de Figuras

FIGURA 2.1: TIPOS DE TRÁFICO , B-ISDN _____	7
FIGURA 2.2: MODELO DEL PROCESO DE MARKOV PARA TIEMPO DISCRETO _____	9
FIGURA 2.3: MODELO DE PROCESO DE MARKOV PARA TIEMPO CONTINUO _____	10
FIGURA 2.4: MODELO DE ACCESO AL BUFFER _____	16
FIGURA 2.5: MODELO DE N FUENTES DE VOZ _____	17
FIGURA 2.6: PROCESO GENERAL DE NACIMIENTO Y MUERTE _____	19
FIGURA 3.1: FUENTE DE TRÁFICO ON-OFF _____	28
FIGURA 3.2: MULTIPLEXACIÓN ESTADÍSTICA DE FUENTES HOMOGÉNEAS _____	30
FIGURA 3.3: CONTROL DE ADMISIÓN, FUENTES HOMOGÉNEAS ON-OFF _____	31
FIGURA 3.4: MODELO DE N FUENTES DE VOZ _____	32
FIGURA 3.5: DISTRIBUCIÓN BINOMIAL Y GAUSSIANA, PARA N=150 Y P=0.02 ____	35
FIGURA 3.6: REPRESENTACIÓN DE LAS DISTRIBUCIONES BINOMIAL, GAUSSIANA Y GAMMA _____	36
FIGURA 3.7: COLAS DE LAS DISTRIBUCIONES GAUSSIANA, GAMMA Y BINOMIAL _	37
FIGURA 3.8: APROXIMACIÓN GAMMA A LA DISTRIBUCIÓN BINOMIAL _____	38
FIGURA 3.9: REPRESENTACIÓN DE LA PROBABILIDAD DE PÉRDIDA P _L vs. CAPACIDAD, PARA N = 500 Y P = 0.02 _____	39
FIGURA 3.10: REPRESENTACIÓN DE LA PROBABILIDAD DE SATURACIÓN vs. CAPACIDAD, PARA N = 500 Y P = 0.02 _____	40
FIGURA 3.11: CONTROL DE ADMISIÓN, P = 0.02. _____	43
FIGURA 3.12: CONTROL DE ADMISIÓN PARA DIFERENTES VALORES _____	44
FIGURA 4.1: COLAS DE LA DISTRIBUCIÓN BINOMIAL, GAUSSIANA Y GAMMA ____	49
FIGURA 4.2: PROBABILIDAD DE PÉRDIDA DE CELDAS _____	51
FIGURA 4.3: PROBABILIDAD DE SATURACIÓN ϵ _____	52
FIGURA 4.4: CONTROL DE ADMISIÓN _____	54

FIGURA 4.5: REGIÓN DE ADMISIÓN, $\varepsilon = 0.00007$	55
FIGURA B.1: FUNCIÓN DE DENSIDAD DE PROBABILIDAD DE LA VARIABLE ALEATORIA GAUSSIANA	63
FIGURA B.2: FUNCIÓN DE DENSIDAD DE PROBABILIDAD DE LA VARIABLE ALEATORIA GAMMA	65
FIGURA B.3: FUNCIÓN DE DENSIDAD DE PROBABILIDAD DE LA VARIABLE ALEATORIA BINOMIAL. (A) $P=0.2$; (B) $P=0.5$	68

Resumen

El control de admisión es uno de los mecanismos que tiene que ser ejecutado en el momento que se establece una solicitud de llamada en una red de banda ancha, con la finalidad de evitar que se produzca congestión en la red. El nodo de acceso es el que decide aceptar o rechazar el pedido de conexión, basado en la descripción de la carga de tráfico encontrada por la conexión a lo largo de la ruta seleccionada.

Esta tesis propone un esquema de control de admisión de llamadas para redes de banda ancha basado en un método de estimación de los descriptores de tráfico para cada uno de los tipos de fuentes de tráfico. La estimación se basa en una aproximación gamma al modelo binomial usado para caracterizar las diferentes fuentes de tráfico y mediante el cual se calcula los parámetros de calidad de servicio .

Capítulo I

Introducción

El modo de transferencia asíncrona (ATM) es la clave para integrar los diferentes tipos de servicios de comunicación en las redes B-ISDN (Broadband Services Integrated Digital Networks). Tomando en cuenta que ATM involucra un esquema de multiplexación estadística, el retardo de colas, pérdida de celdas o cualquier otro tipo de degradación de calidad de servicio (QoS) pueden fácilmente ocurrir cuando la red se encuentre congestionada. El tráfico tipo ráfaga, tal como transferencia de imágenes, tiene un serio impacto en la calidad de la comunicación debido a las fluctuaciones de la carga estadística.

De esta manera, el control de tráfico es necesario para evitar la congestión. Los métodos de control de tráfico pueden ser divididos en dos categorías (control reactivo y control preventivo) y el uso más eficiente de los recursos de la red puede ser alcanzado cuando se combina los dos tipos de controles. El control de admisión de llamadas, es un tipo de control de forma preventiva; el cual es por su naturaleza más efectivo que el control reactivo para el uso en redes de alta velocidad, [1].

Por lo tanto, el principal objetivo del control de admisión de llamadas es el determinar cuando aceptar o no una nueva llamada, con la finalidad de mantener la carga de la red en un nivel adecuado y cumplir con los requerimientos de calidad de servicio

(QoS). En otras palabras, cuando se establece una llamada, el nodo de acceso decide el aceptar o rechazar el pedido de conexión, basado en la descripción de la carga del tráfico encontrada por una conexión a lo largo de la ruta seleccionada, [2].

En el presente trabajo se propone un esquema de control de admisión de llamadas basado en un método para el cálculo de los descriptores de tráfico para cada una de las fuentes de tráfico. Esta estimación se basa en aproximar a una distribución gamma los parámetros de calidad de servicio; que para este caso se considerará la probabilidad de pérdida de celdas y la probabilidad de estar en la región de saturación.

1.1 Objetivo

El objetivo de este trabajo es el determinar si en el momento que llega una llamada que requiere una conexión virtual con una calidad de servicio específica (ancho de banda, probabilidad de pérdida, retardo, etc.) esta es o no aceptada.

Para esto, se plantea un nuevo esquema de control de admisión de llamadas; el mismo que se basa en el cálculo de los parámetros de calidad de servicio, QoS, como son la probabilidad de pérdida de celdas y la probabilidad de estar en la región de saturación, mediante la aproximación del modelo de nacimiento y muerte ON – OFF de las fuentes de tráfico a una distribución Gamma.

1.2 Justificación

Como se mencionó anteriormente, las redes de banda ancha están diseñadas para soportar todo tipo de tráfico, tales como imágenes de alta resolución, vídeo de alta resolución, multimedia y en general tráfico que consume bastante recurso de la red, además de los datos interactivos, tráfico de datos y voz. Este tipo de tráfico formado por la mezcla de una variedad de servicios es bastante compleja su caracterización ya que cada tipo de tráfico tiene su propia calidad de servicio.

De esta manera, es importante el poder determinar cuanto tráfico puede manejar una red si una determinada calidad de servicio es establecida para cada tipo de tráfico y a la vez se debe mantener una utilización óptima de los recursos de la red. Adicionalmente,

que requerimientos de recursos, en términos de capacidad de enlace y buffer, número de enlaces, número de nodos, etc. son necesarios considerar en el diseño de una red para que esta sea capaz de manejar una carga de tráfico establecido con la integración de todos los diferentes tipos de servicios.

Por lo expuesto anteriormente, es importante enfocar el estudio al control de admisión de llamadas en el contexto de la redes de B-ISDN. El esquema de control de admisión de llamadas debe ser tal que permita que un máximo número de llamadas sean aceptadas en la red, pero a la vez haciendo un uso óptimo de los recursos de la red.

Por lo tanto, el estudio se centrará en establecer el esquema más óptimo que permita definir la capacidad mínima necesaria para poder manejar un máximo número de llamadas provenientes de diferentes tipos de tráfico y que cumpla con las calidades de servicio de cada una de ellas.

A lo largo de este trabajo se tratará de establecer que el esquema de control de admisión de llamadas basado en una aproximación Gamma hace un mejor uso de los recursos de la red que el que se obtiene con el esquema basado en una aproximación Gaussiana.

1.3 Descripción de la tesis

El presente trabajo se encuentra dividido en 6 capítulos, los cuales tienen la siguiente organización. En el capítulo 2, se revisa brevemente los diferentes tipos de variables y distribuciones existentes tanto discretas como continuas, las cuales será necesario usarlas para el análisis de las fuentes de tráfico .

En el capítulo 3, se presenta la caracterización de los diferentes tipos de tráfico, así como los conceptos de calidad de servicio, clases de servicio y control de acceso de llamada. El capítulo 4, se basa en el análisis del esquema de control de admisión de llamadas basado en una aproximación gamma que se hace al modelo de fuentes ON – OFF.

En el capítulo 5, se hace un análisis comparativo de los resultados que se obtiene con los dos tipos de esquema para el control de admisión de llamadas; es decir, el primero usando la aproximación Gaussiana al modelo de fuentes ON – OFF, y el segundo esquema que es el que se plantea en el presente trabajo y es la aproximación a la distribución Gamma. Por último, el capítulo 6 se presentan las conclusiones y recomendaciones de este trabajo así como los trabajos futuros que podrían realizarse.

Capítulo II

Modelamiento y Características de tráfico para redes de banda ancha

2.1 Introducción

Los modelos de tráfico han jugado un papel muy significativo en el diseño e implementación de las redes de banda ancha. Los modelos cuyas llamadas arriban con distribución Poisson y los tiempos de espera con una distribución exponencial han sido óptimos para el análisis y evaluación del rendimiento para redes de conmutación de paquetes. Sin embargo, la integración de paquetes de voz, vídeo, imágenes, y tráfico

generado por microcomputadores (ya sean ráfagas o transferencia de archivos más grandes), cada uno con sus respectivas calidades de servicio (QoS), requieren de un desarrollo de modelos de tráfico más sofisticados para lograr un diseño más preciso y un mejor rendimiento de la red, [6].

Los procesos de arribo tipo Poisson, que se usan muy comúnmente en el análisis del rendimiento de la red, no siempre proporcionan el modelo adecuado, en especial cuando se trabaja con tráfico de tipo flujo o con tasas de bit continuo. (Por la ley de los Números Grandes, la multiplexación de muchos flujos de tráfico independientes puede modelarse como un proceso con arribos tipo Poisson. El problema es poder determinar cuantos flujos multiplexados son necesarios para poder aproximar los arribos a procesos Poisson). Las fuentes tipo ráfaga han sido usadas como modelos de tráfico para transferencias de archivos y transmisión de imágenes. Este tipo de tráfico está caracterizado por períodos aleatorios de inactividad y actividad. Una fuente ráfaga es aquella en la que el período de inactividad es mucho más largo que el tiempo que esta activa y transmitiendo celdas.

Por lo tanto, es importante caracterizar cuidadosamente el tráfico para asegurar que los modelos que se usan tendrán resultados óptimos. De esta manera, el tráfico que es particularmente importante analizar es cuando se trata de fuentes de paquetes de vídeo, imágenes y tráfico multimedia. La caracterización de tráfico no es solo importante cuando se diseña los buffers para los multiplexores, sino también cuando se estudia el control de admisión, acceso y flujo.

Los tres tipos básicos de tráfico para redes B-ISDN son: voz con una tasa de bit continuo (CBR) (Clase A), vídeo para representar el tráfico de flujo con tasa de bit variable (Clase B) (Clase B), y datos tipo ráfaga (Clases C y D). Estos tipos de tráfico se encuentran esquematizados en la figura 2.1.

2.2 Modelo General de las Fuentes de Tráfico

El tráfico de una fuente puede ser caracterizado por diferentes parámetros, siendo los más importantes los siguientes:

- Tasa de bit promedio: es la media aritmética de la tasa de bit a la cual la fuente transmite.
- Tasa de bit pico: es la tasa de bit más alta a la que la fuente es capaz de enviar.
- Tasa de bit físico: es la tasa de bit real proporcionado por el enlace al usuario de la red local.
- Duración pico: es una medida del tiempo en el que la fuente puede mantener la tasa de bit pico.

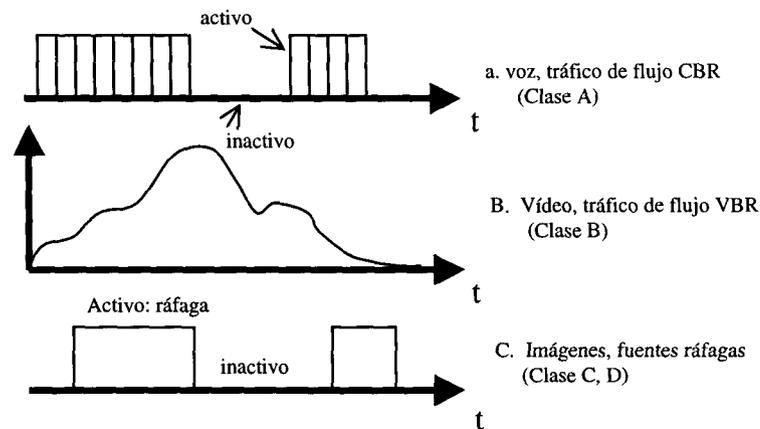


Figura 2.1: Tipos de tráfico, B-ISDN

La relación entre la tasa de bit pico con la tasa de bit promedio se le conoce como *burstiness* (relación de ráfaga), parámetro que proporciona una información vital en el momento de establecer una conexión. Este parámetro es comúnmente usado para medir la frecuencia con la que una fuente envía tráfico. Una fuente que envía tráfico con muy poca frecuencia se dice que es de tipo ráfaga, mientras que una fuente que siempre envía a la misma tasa se dice que no es ráfaga. La fórmula que define la intermitencia en términos de la tasa pico y la tasa promedio de celdas se define como:

$$\text{"Burstiness"} = \text{tasa pico} / \text{tasa promedio}$$

basados en los parámetros de QoS establecidos por el usuario el control de admisión de llamadas puede determinar el tipo de servicio requerido.

Es también muy importante el conocer la duración pico antes de que la conexión sea establecida. Esto permite al nodo de la red ATM el determinar el número máximo de conexiones de celdas que pueden entrar en la red para una conexión dada durante un intervalo de tiempo.

Otro parámetro usado generalmente es la utilización, que representa la fracción de la capacidad de transmisión del enlace que es usada por la fuente, teóricamente medida sobre un periodo de tiempo infinito; sin embargo, en la práctica se mide sobre un intervalo de tiempo grande. La definición de la utilización está dada en términos de la tasa pico de celdas y de la tasa de transmisión del enlace

$$\text{Utilización} = \text{tasa pico} / \text{tasa de enlace}$$

2.2.1 Arribos Poisson y Procesos de Markov

Los arribos Poisson ocurren de tal manera que por cada incremento de tiempo (t) la probabilidad de los arribos es independiente de cualquier evento anterior. Estos eventos, en los modelos pueden ser celdas individuales, ráfagas, pedidos de servicio de celdas o paquetes o cualquier otro evento arbitrario.

La probabilidad de que el tiempo de interarribo t entre los eventos, tenga un cierto valor se conoce como la probabilidad de densidad del tiempo de interarribo. La siguiente ecuación da la probabilidad resultante de que el tiempo t de interarribo sea igual a algún valor x cuando la tasa promedio de arribo es λ eventos por segundo.

$$P(t=x) = \lambda e^{-\lambda x} \quad (2.1)$$

En este proceso la probabilidad de que el tiempo de interarribo sea de x segundos es independiente de la memoria de cuanto tiempo a transcurrido. Este hecho simplifica el análisis de los procesos aleatorios ya que no hay historia, o memoria, que sea guardada. Este tipo de procesos son comúnmente conocidos como procesos de Markov, [7].

La probabilidad de que n arribos independientes ocurran en t segundos esta dado por la distribución de Poisson:

$$P(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (2.2)$$

Cuando se combina los dos procesos se obtiene un modelo muy comúnmente usado conocido como proceso de Poisson con modulación de Markov (MMPP). Existen dos esquemas para este proceso: el discreto (que corresponde a las celdas ATM) y el continuo.

En la Figura 2.2 se muestra la probabilidad de las transiciones de la fuente entre los estados activo a inactivo, así como la probabilidad de que permanezca en el mismo estado para cada tiempo de la celda. En otras palabras, durante cada tiempo de celda la fuente hace la transición de estado, ya sea para el otro estado, hacia atrás o se queda en el mismo estado, con una probabilidad para cada una de las acciones indicadas por las flechas en el diagrama.

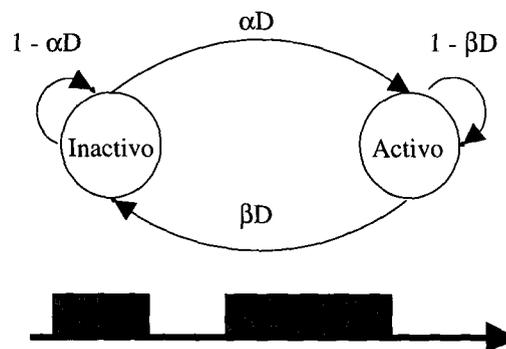


Figura 2.2: Modelo del Proceso de Markov para tiempo discreto

El burstiness, o la relación del pico promedio, del modelo de la fuente discreta esta dada por la siguiente formula:

$$b = \frac{\alpha + \beta}{\beta} \quad (2.3)$$

donde α es el número promedio de ráfagas que llegan por segundo y β es la tasa promedio de las ráfagas que se completan. Se define a D como el tiempo cuantificado por celda en unidades de celdas por segundos. Por lo tanto, αD define la probabilidad de que una ráfaga inicie en un tiempo de celda determinado, y βD se define como la probabilidad de que una ráfaga se termine en un tiempo determinado de celda, [7]. La duración promedio de una ráfaga d (en celdas) es calculada por la serie geométrica como sigue:

$$d = \frac{1}{\beta D} \quad (2.4)$$

Para el modelo de tiempo continuo que se ilustra en la figura 2.3, el tiempo que transcurre desde el inicio de la ráfaga hasta el final de la ráfaga es modelado en vez de modelar las celdas individualmente. En este diagrama de transición de estados las variables asociadas con las flechas se refieren al exponente de la tasa en la distribución exponencial negativa. Tanto los modelos de Markov discreto y continuo se basan en resultados equivalentes con excepción del factor de cuantificación de las celdas D .

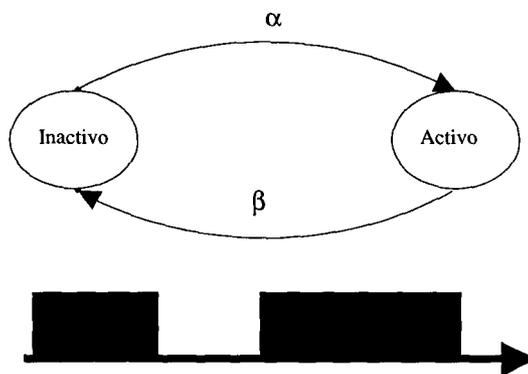


Figura 2.3: Modelo de proceso de Markov para tiempo continuo

El factor burstiness b correspondiente para el proceso continuo es :

$$b = \frac{\alpha + \beta}{\beta} \quad (2.5)$$

y la duración de la ráfaga promedio (en segundos) esta dada por la siguiente ecuación

$$d = \frac{1}{\beta} \quad (2.6)$$

Hay que observar que estas fórmulas son idénticas al caso discreto con la excepción de la ausencia del tiempo de la celda discreta D en el denominador de la ecuación de la duración promedio de la ráfaga del modelo continuo.

2.3 Calidad de Servicio

El servicio básico que proporciona las redes ATM es el transporte secuencial de celdas punto a punto. El servicio se inicia por el pedido de un usuario para una conexión virtual, la cual puede ser admitida o rechazada por la red. El grado de servicio (GOS) corresponde al tráfico ofrecido en términos de la probabilidad de bloqueo en el nivel de aceptar o rechazar conexiones. Una vez que una conexión ha sido aceptada, las celdas que son transportadas pueden experimentar dos tipos básicos de problemas dentro de la red: retardo y pérdida.

La calidad de servicio (QoS) se refiere a un conjunto de parámetros que caracterizan el tráfico tales como el retardo de las celdas, la variación de retardo y la tasa de pérdida de celdas; todos estos parámetros corresponden a las variaciones experimentadas por el tráfico transportado, [8]. Estos parámetros están relacionados con la percepción del usuario del servicio de red. La red es responsable de mantener el nivel de calidad de servicio que esperan los usuarios.

Para las diferentes mezclas de servicios en B-ISDN, las redes ATM soportaran un número específico de clases de QoS:

- Clase A: para emulación de circuitos y servicios de CBR
- Clase B: para vídeo con VBR y conversación
- Clase C: para conexión orientada (frame relay)
- Clase D: para datos sin conexión

Cada clase de QoS esta caracterizada por requerimientos específicos de QoS. Por ejemplo, las conexiones de clase A requieren de un límite de retardo de celdas y variaciones de retardo. Otro tipo de QoS pueden ser definidos posteriormente; como por ejemplo, un nuevo servicio ABR se ha definido con una probabilidad de pérdida de celdas

bastante baja o cero. Más aún, una clase de QoS sin un retardo de celda específico o requerimientos de pérdida de celdas puede ser soportada por una red ATM. Sin hacer la diferenciación del tráfico en sus clases de QoS, la red deberá de soportar requerimientos más estrictos de QoS para todo el tráfico. En principio, la diferenciación del tráfico beneficia al proveedor de la red ya que permite flexibilidad en los requerimientos de QoS.

Con respecto a la QoS, el rendimiento de la red se refiere al conjunto de parámetros que miden la habilidad de la red para proporcionar estos servicios entre los usuarios. Mientras la QoS es importante desde la perspectiva de los usuarios en el punto del acceso del servicio; en cambio, el rendimiento de la red es fundamental desde el punto de vista del que suministra la red en los diferentes puntos de la red. A nivel de llamada, los parámetros de rendimiento de la red pueden incluir el retardo de establecer una conexión, el retardo de liberar una conexión y la probabilidad de bloqueo. A nivel de celda, los parámetros de rendimiento de la red pueden incluir tasa de error de celda, tasa de pérdida de celdas, tasa de error en inserción de celdas, retardo punto a punto de las celdas y variación del retardo de las celdas. En general, el rendimiento de la red puede ser mayor que el necesario para satisfacer los requerimientos de QoS de un servicio.

En ATM, el rendimiento de la red puede estar afectada por la congestión causada por la limitación de recursos (por ejemplo: buffers, ancho de banda y procesadores). Debido a que ATM está orientado a conexiones y a paquetes, este tipo de redes pueden presentar congestión tanto a nivel de conexiones como de celdas. A nivel de conexiones, los procesadores de llamadas estarán pendientes de llamadas no realizadas. A nivel de celdas, los enlaces de transmisión se saturarán con el tráfico y los buffers se llenarán con celdas. De esta manera, una congestión no controlada puede manifestarse por un incremento en el bloqueo de llamadas, retardo de celdas, y tasa de pérdida de celdas. El control de congestión trata de detectar y de controlar la congestión para minimizar su intensidad, alcance y duración, [8].

De esta manera, la QoS está definida en términos de las medidas que se detallan a continuación y que se basan en las celdas enviadas de un usuario origen a un usuario destino, y son:

- Celdas transmitidas por el usuario origen
-

-
- Celdas transferidas exitosamente al usuario destino
 - Celdas perdidas que no llegan al usuario destino
 - Celdas que llegan al destino pero con error
 - Celda que llega al destino pero que no ha sido enviada por el origen. Esto puede ocurrir debido a un error de cabecera de celda no detectado o a un error en la configuración.

Los parámetros de QoS son definidos en términos de las medidas anteriores con las siguientes definiciones:

- Tasa de Pérdida de Celdas = Celdas perdidas / celdas transmitidas
- Tasa de Error de Celdas = Celdas con error / (celdas transferidas exitosamente + celdas con error)
- Tasa de Bloques de celda con error = bloques de celdas con error / bloques de celdas totales transmitidas
- Tasa de error de inserción = Celdas mal insertadas / intervalo de tiempo

2.3.1 Clases de QoS

Se han establecido un pequeño número de clases de calidad de servicio QoS predefinidas, con valores particulares de parámetros pre-especificados por la red para cada una de las clases de QoS. El Forum de ATM define cinco clases de QoS y ejemplo de aplicaciones se encuentran resumidas en la tabla 2.1, [7].

Clase de QoS	Parámetros de QoS	Aplicación
0	No especificado	Mejor esfuerzo
1	Específicos	Simulación de circuitos, CBR
2	Específicos	VBR Vídeo/Audio
3	Específicos	Datos conexión orientada
4	Específicos	Datos sin conexión

Tabla 2.1: Clases de QoS

Una clase de calidad de servicio proporciona rendimiento a una conexión virtual ATM (VCC o VPC) como se encuentra especificado por un subconjunto de parámetros de

rendimiento de ATM. Como se dijo anteriormente, para cada clase de calidad de servicio QoS existe un valor específico para cada parámetro de rendimiento.

Las clases de servicio definidas son:

- Clase de Servicio A: simulación de circuitos, vídeo con tasa de bit constante.
- Clase de Servicio B: audio y vídeo con tasa de bit variable.
- Clase de Servicio C: transferencia de datos con conexión orientada.
- Clase de Servicio D: transferencia de datos sin conexión.

Las clases de calidad de servicio que están actualmente definidas son:

- Clase 1 de calidad de servicio QoS: soporta una calidad de servicio que cumpla con los requerimientos de rendimiento de la clase A. El rendimiento de esta clase es comparable con el rendimiento de una línea digital privada.
- Clase 2 de calidad de servicio QoS: soporta una QoS que cumpla con los requerimientos de rendimiento de la clase B. Esta clase esta caracterizada por aplicaciones de empaquetamiento de vídeo, audio en teleconferencia y multimedia.
- Clase 3 de calidad de servicio QoS: soporta una QoS que cumpla con los requerimientos de rendimiento de la clase C. Esta clase esta caracterizada por los protocolos de interoperación de conexiones orientadas, tales como Frame Relay.
- Clase 4 de calidad de servicio QoS: soporta una calidad de servicio que cumple con los requerimientos de rendimientos de la clase D. Esta clase esta caracterizada por los protocolos de interoperación tales como IP o SMDS.

Varias aplicaciones tienen diferentes requerimientos de QoS. Por ejemplo, la voz empaquetada puede aceptar al menos el 1% de la tasa de pérdida de celdas sin ser objetada. Las nuevas aplicaciones no tienen estas bases, o requerimientos bien definidos; sin embargo, hay algunos requerimientos generales.

Los requerimientos para aplicaciones de vídeo dependen de varios factores, incluyendo los algoritmos de codificación de vídeo, el grado de movimiento requerido en la secuencia de las imágenes y la resolución requerida en la imagen. Las pérdidas generalmente causan alguna degradación de la imagen, que va desde una imagen con partes distorsionadas hasta la pérdida de un frame completo, dependiendo de la pérdida y de la sensibilidad del algoritmo de codificación del vídeo.

La combinación de vídeo y audio es bastante sensitivo a los retardos diferenciales. La percepción humana es altamente atenuado a la correlación correcta de audio y vídeo, la cual es aparentemente doblada en algunos idiomas. Las aplicaciones de transferencia de archivos son también sensibles a las pérdidas y variaciones en retardo, el cual es el resultado de las retransmisiones y de la reducción consecuente.

Los usuarios de aplicaciones interactivas también son sensibles a las pérdidas y variaciones en el retardo debido a las retransmisiones y el tiempo de respuesta inconsistente, el cual puede disminuir en productividad. El tiempo de respuesta consistente puede afectar en como los usuarios perciben la calidad de servicio de los datos.

2.3.2 Descriptores de Tráfico

Los descriptores de tráfico son una lista de parámetros los cuales capturan las características intrínsecas de las fuentes de tráfico.

Esta sección describe los parámetros de tráfico definidos por el Forum de ATM UNI versión 3.0, [7]:

- Tasa pico de celda: $1/T$ en unidades de celdas/segundo, donde T es el mínimo espaciamiento entre celdas en segundos (por ejemplo el intervalo de tiempo para el primer bit de una celda hasta el primer bit de la siguiente celda).
- Tolerancia de la variación del retardo de celdas = τ en segundos. Este parámetro de tráfico normalmente no puede ser especificado por el usuario, pero es establecido por la red. El número de celdas que pueden ser enviadas a la línea de acceso con tasa $\tau/T+1$.

- Tasa de celda aceptable es la tasa promedio máxima que una ráfaga, de una fuente de tráfico ON - OFF puede ser enviada a una tasa pico.
- Tamaño máximo de ráfaga, es el número máximo de celdas que pueden ser enviados.

2.4 Modelamiento de Paquete de Voz

Una fuente de voz puede ser representada por un proceso de dos estados: activo e inactivo. Para realizar una aproximación relativamente razonable, se asume que los estados tienen una distribución exponencial, teniendo de esta manera un proceso de dos estados de nacimiento y muerte. El parámetro λ representa la tasa de salida del estado de silencio, y α se define como la tasa de transición de salir del estado activo. De esta manera, el promedio de estar en estado activo es $1/\alpha$, y el intervalo de silencio es de $1/\lambda$.

El modelo compuesto por N fuentes de voz independientes multiplexadas pueden encontrarse en una forma bastante sencilla. En la Figura 2.4 se muestra N fuentes multiplexadas en el acceso del buffer de la red. Cada fuente genera V celdas/seg en el ciclo activo. Debido al multiplexaje estadístico, la capacidad del enlace en la red puede ser menor a NV que representa el número máximo de celdas/seg. que pueden ser generadas. Además, se tendrá que $C < N$. Tomando en cuenta que la tasa promedio de generación de celdas es $VN\lambda/(\alpha+\lambda)$, el parámetro C debe satisfacer el límite inferior de la desigualdad dada por

$$\left(\frac{\lambda}{\alpha + \lambda} \right) N < C \quad (2.7)$$

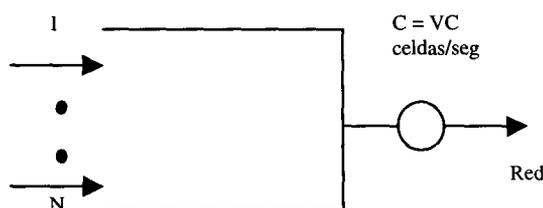


Figura 2.4: Modelo de acceso al buffer, N fuentes de voz, V celdas/seg cada fuente

Para un buffer infinito, la ecuación anterior representa la condición de estabilidad. Si se divide la ecuación anterior con respecto a C se puede definir a la utilización como

$$\rho \equiv \left(\frac{\lambda}{\alpha + \lambda} \right) \frac{N}{C} < 1 \quad (2.8)$$

Con este modelo de tráfico de paquete de voz multiplexado, se puede encontrar los parámetros de rendimiento y diseño así como las estadísticas de ocupación de la cola, retardo de las celdas, y el tamaño del buffer necesario para asegurar una probabilidad de pérdida de celdas aceptable.

Estas N fuentes de voz independientes y multiplexadas, representadas cada una por un modelo de dos estados, llevan a un modelo de tráfico que se muestra en la Figura 2.5, [6]. Este modelo puede ser visto, como un modelo de nacimiento y muerte de $(N + 1)$ estados. El estado representa el número de fuentes activas. De esta manera, el estado i representa que i fuentes están activas, y el promedio de tasa de entrega de celdas al buffer es iV celdas/seg. (A pesar de que cada fuente genera V celdas/seg periódicamente mientras se esta en activo, la suma de fuentes independientes y no sincronizadas, da como resultado un proceso de entrada aleatorio, con un valor promedio de iV celdas/seg).

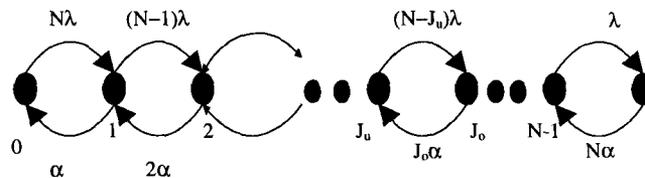


Figura 2.5: Modelo de N fuentes de voz

Dos estados específicos, J_u y J_o , se indican en la Figura 2.5. J_u , se conoce como el estado de no saturación, y está definido como la parte entera o el piso de la capacidad C

$$J_u = \lfloor C \rfloor \quad (2.9)$$

De igual manera, J_o se le define como el estado de saturación, y se define como el techo de C

$$J_o = [C] \quad (2.10)$$

Como se puede analizar en la figura anterior cuando el estado $i > C$ la cola tenderá a estar llena y cuando $i < C$, la cola tenderá a estar vacía. J_u representa el estado por debajo del cual la cola esta vacía; J_o representa el estado sobre el cual la cola esta llena. Por otro lado, cuando el sistema está en estados $i > C$, la cola se llena y el sistema se dice que esta en saturación; en cambio, cuando $i < C$, con la cola vacía, el sistema se dice que no está en saturación. La tasa de cambio de la cola de la Figura 2.4 cuando el sistema está en el estado i , es $V(C - i)$ celdas/seg.

La probabilidad de estado estable π_i , de que la fuente este en el estado i , es la probabilidad (binomial) de que i de las N fuentes de dos estados estén activas (en la parte activa), mientras que las restantes $(N - i)$ están inactivas (silencio). Tomando en cuenta que la probabilidad de que cada fuente esté activa es $\lambda/(\lambda + \alpha)$, la probabilidad π_i puede ser escrita como,

$$\pi_i = \binom{N}{i} \left(\frac{\lambda}{\lambda + \alpha} \right)^i \left(\frac{\alpha}{\lambda + \alpha} \right)^{N-i} \quad (2.11)$$

Esta expresión también puede ser escrita en su manera equivalente;

$$\pi_i = \binom{N}{i} \left(\frac{\lambda}{\alpha} \right)^i \left(1 + \frac{\lambda}{\alpha} \right)^{-N} \quad (2.12)$$

Considerando el proceso general, dependiente de estados de nacimiento y muerte de la Figura 2.6, el parámetro λ_i representa la tasa de transición con la que va del estado i al estado $i + 1$; μ_i es la tasa de transición que va del estado i al estado $i - 1$. La probabilidad de estar en el estado estable π_i de estar en el estado i está dado por

$$\pi_i = \frac{\lambda_0 \lambda_1 \lambda_2 \dots \lambda_{i-1}}{\mu_1 \mu_2 \dots \mu_i} \pi_0 \quad (2.13)$$

con π_0 que se lo encuentra sumando sobre todos los estados y estableciendo $\sum_{i=0}^N \pi_i = 1$.

Un segundo modelo que permite analizar aproximadamente la generación de celdas (paquetes) por una fuente en el estado activo es el que se conoce como proceso de Poisson con modulación de Markov (MMPP Markov-modulated Poisson process), el cual se basa en que el sistema genera paquetes a una tasa Poisson y la tasa de transferencia cambia de estado a estado. Las transiciones entre los estados están gobernados en base a una cadena de Markov en tiempo continuo.

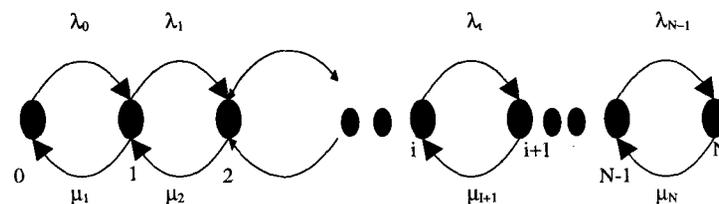


Figura 2.6: Proceso General de nacimiento y muerte

El proceso de Poisson con modulación de Markov es definido en forma general por n estados, con un proceso Poisson y una tasa de transferencia λ_i , mientras que el estado i esté entre $1 \leq i \leq n$. En este caso, en lugar de proceder con el análisis general MMPP, se realizará la caracterización, modelamiento y análisis de las N fuentes de voz multiplexadas en el buffer. Para esto, se modela cada fuente de voz como si estuviera en dos estados, el estado activo y el de silencio, con tasa λ , representando la tasa de transición del estado de silencio al estado activo, y α como la tasa en la dirección contrario.

2.5 Caracterización del Tráfico de Vídeo

El tráfico de voz proporciona un primer ejemplo de tráfico de flujo de clase A con una tasa de bit constante; en cambio, el tráfico de vídeo, en forma comprimida, es el primer ejemplo de tráfico de flujo de clase B con una tasa de bit variable.

MPEG fue desarrollado inicialmente para el almacenamiento y recuperación de vídeo comprimido digitalmente, las características de codificación no tomaron en cuenta las características de transmisión para redes de banda ancha. De esta manera, los estudios

posteriores se encaminaron para adaptar el standard MPEG para la transmisión sobre redes de alta velocidad. Una posibilidad sobre estas consideraciones es la de transmitir la codificación MPEG como un flujo de tasa de bit continuo (CBR) a una tasa de bit menor que la del flujo original no comprimido CBR, usando buffers estables. El problema con esta aproximación, es que la salida del tráfico resultante varia en calidad. Sin embargo y tomando en cuenta que es la transmisión sobre redes de alta velocidad la que se esta estudiando, se considerará únicamente las fuentes de vídeo con tasa de bit variable VBR usando el estándar JPEG.

Un gran número de modelos de tráfico que consideran estas características de tasa de bit variable VBR han sido propuestos. Algunos modelos que consideran los modelos de cadenas de Markov han sido considerados para modelar fuentes de vídeo multiplexados, mientras que el modelo autoregresivo (AR) se ha planteado para aproximar una sola fuente de vídeo. El modelo AR es útil para simular la tasa de bit de salida de una fuente de vídeo con tasa de bit variable VBR para un rango determinado. Específicamente, el modelo propuesto en [9] muestra características estadísticas muy cercanas a las que se obtiene con un codificador de vídeo de VBR. Algunos de los modelos de AR y cadenas de Markov no describen claramente los efectos de cambios de escena por lo que no les hacen muy útiles para las aplicaciones reales. Por tal motivo en [9] se propone un modelo AR con dos procesos complementarios que permitirá capturar la función de autocorrelación de la tasa de bit más exactamente y el poder determinar los bits adicionales que se generan en los cambios de escena.

2.6 Modelo de Tráfico Bursty

Anteriormente se establecieron algunas maneras de caracterizar la voz y el vídeo. Se completa la discusión de la caracterización del tráfico introduciendo un modelo simple de dos estados para tráfico de ráfagas. Hay que tomar en cuenta que este modelo de dos estados fue usado ya anteriormente para caracterizar la voz.

Una fuente ráfaga puede ser caracterizada mediante periodos alternativos de inactividad y de actividad. Esto es muy similar a los estados de silencio/conversación de la caracterización de la voz. La diferencia básica es que en el caso de las fuentes ráfagas, el periodo de inactividad es mucho mas grande que el periodo de actividad, durante el cual la

fuerza transmite. En otras palabras, una fuente ráfaga es en general considerada como una en la cual, después del período de inactividad transmite celdas en forma de ráfagas y luego regresa al período de inactividad. Una variedad de modelos han sido propuestos para tal proceso. Uno de los más comunes es el proceso de arribo Poisson en el cual el número de celdas o paquetes son generados en cada momento del arribo de Poisson.

2.7. Control de Admisión

El modo de transferencia asíncrona (ATM) es un método para redes de Banda ancha con la finalidad de poder distribuir un amplio rango de servicios de comunicaciones, tales como voz, datos y servicios de vídeo. En redes ATM, la información para cada terminal es segmentada en paquetes pequeños de longitud fija, llamadas celdas, las cuales son asíncronamente multiplexadas en los enlaces de transmisión y conmutadas sin control de flujo de ventana como se hace en X.25. Cuando se incrementa el número de llamadas en un enlace de capacidad fija se degrada la calidad de servicio en términos de pérdida de celdas y retardo. De esta manera, con la finalidad de asegurar que los requerimientos de calidad de servicio se cumplan, las redes deben decidir cuando o no aceptar una nueva llamada. Esta decisión es la que se conoce como control de admisión de llamadas, [10].

El control de admisión de llamadas define un conjunto de reglas o de algoritmos para decidir cuando una nueva llamada puede o no ser aceptada. La decisión es hecha en base a las características de tráfico de la nueva llamada y la disponibilidad de los recursos requeridos por la red para manejar el tráfico adicional, sin afectar los requerimientos de recursos de las llamadas existentes. En otras palabras, el tráfico potencial adicional ha ser generado por la nueva llamada puede no tener ningún efecto en los requerimientos del rendimiento de tráfico de las llamadas existentes.

De esta manera, una nueva conexión es admitida por la red únicamente si se estima que la nueva conexión puede establecerse con la QoS requerida mientras se mantenga la QoS de las conexiones existentes.

El pedido de conexión se pasa a lo largo de la ruta. Cada nodo decide donde se puede conseguir los recursos necesarios. El nuevo pedido es aceptado únicamente si es aceptado en cada nodo. La aceptación de nuevas conexiones implica un acuerdo en el

contrato de tráfico especificando las obligaciones entre los usuarios y la red. Una vez que se establece los parámetros de una conexión virtual estos pueden ser cambiados únicamente a través de una renegociación a través del usuario y la red, [11].

El control de admisión es uno de los aspectos más importantes de discusión, para lo que es necesario el conocimiento tanto de las condiciones de tráfico actuales como del impacto que se tendría al añadir una nueva conexión. Esta información es proporcionada tomando en cuenta la cantidad del ancho de banda actualmente distribuido para las conexiones existentes y mediante la identificación de la cantidad de ancho de banda que no está siendo utilizado en el circuito virtual sobre el cual una nueva conexión trata de ser ruteada, así como cuanto ancho de banda adicional es necesario para aceptar una nueva conexión.

De esta manera, el control de admisión es uno de los mecanismos de control de congestión que deben ser ejecutados cuando se establece una llamada virtual. Los nodos de acceso deciden aceptar o rechazar un pedido de conexión basada en la descripción del tráfico anticipada por la fuente o por la carga de tráfico encontrada por la conexión a lo largo de la ruta de la red, [12]. Para cada nueva conexión un ancho de banda equivalente es distribuido con la finalidad de asegurar que los objetivos de transporte de celdas (pérdida de celda y retardo) permanezcan a un nivel satisfactorio para todos los usuarios.

El control de admisión es necesario analizarlo bajo ciertas consideraciones de los descriptores de tráfico adoptados para anticipar las características de las fuentes de tráfico. El control de admisión debe por lo tanto decidir cuando aceptar una nueva conexión basado en el conocimiento de la carga actual de la red, los descriptores de tráfico de la nueva conexión y de los requerimientos de retardo/pérdida de celda.

En el momento de establecer una llamada virtual, la decisión de aceptar o no una nueva conexión se la realiza tomando en cuenta los objetivos de rendimiento de transporte de celda bajo la carga de la red presente a lo largo de la conexión de la ruta y durante la duración de toda la conexión.

El mecanismo de control más sencillo que se puede aplicar es el distribuir la tasa de bit pico requerida para todas las conexiones. De esta manera, cuando la suma de las

tasas pico de las conexiones en cualquier enlace que componen la ruta no excede la tasa de bit de ese enlace la nueva llamada es aceptada. Esto corresponde a la selección de un modo de operación no estadístico. Este modo de operación es muy usado en el ambiente ATM para manejar fuentes de tráfico de tasa constante y uniforme. Los modos de operación no estadísticos pueden ser también aplicados a fuentes tipo ráfagas, caracterizadas por descriptores de tráfico que no están disponibles para lograr ganancias de ancho de banda significantes, cuando se trabaja en un modo de operación estadístico.

Cuando se multiplexa estadísticamente la tasa de transmisión de celdas, la pérdida de celdas y el rendimiento del retardo dependen de la utilización del enlace y del tamaño del buffer, así como de las características de tráfico de las conexiones multiplexadas en el enlace. Las características de tráfico de conexiones ráfagas deben incluir descriptores tales como: P tasa pico (Mbit/s), A tasa promedio (Mbits/s), P/A (burstiness) rafagas.

En un modo de operación estadístico para admitir a una nueva conexión es necesario definir un ancho de banda equivalente W asociado con la conexión donde $A < W < P$. El ancho de banda equivalente W debe ser derivado para el conjunto de descriptores de tráfico declarado para las fuentes en el momento de establecer una llamada virtual. Si esta medida de W es independiente de la carga particular de tráfico en cada enlace que compone la ruta a través de la red, entonces el criterio de aceptación puede va a ser tal que la suma de los anchos de banda equivalente de las conexiones en cualquier enlace dado no deberá exceder a la tasa del enlace o deberá ser una fracción del 90%.

Por lo tanto, es necesario caracterizar apropiadamente la fuente de tráfico con la finalidad de establecer adecuadamente sus descriptores. Por otro lado, para determinar el ancho de banda equivalente es necesario considerar ciertas limitaciones en las cuales las fuentes tipo ráfaga pueden ser estadísticamente manejadas y poder lograr una significativa ganancia en la eficiencia del ancho de banda.

El control de admisión de conexión en ATM es realizado cuando un pedido de conexión ha sido hecho. Tal conexión puede establecerse ya sea en el momento de la provisión de un servicio o por medio de un protocolo de señalización en el momento en que se origina la llamada en una base dinámica. El mecanismo de control de admisión debe determinar si se puede o no dar la cantidad necesaria de recursos, tales como los

recursos necesarios para proporcionar un ancho de banda adecuado o limitar el retardo de transmisión, para servir la conexión. Si los mecanismos de control de admisión determinan que los recursos necesarios no pueden ser disponibles, la red no aceptará el pedido de la conexión.

En las redes ATM, el control de admisión de llamadas decide cuando aceptar o no una nueva llamada, con la finalidad de asegurar la calidad de servicio tanto para las llamadas existentes como para la nueva llamada.

La medida de calidad de servicio usada para el control de admisión de llamadas que se considera es la probabilidad de pérdida de celdas.

Un esquema de control de admisión de llamadas, permite que cada llamada use un ancho de banda que mantendrá una probabilidad de pérdida de celdas dentro de los límites permitidos. De esta manera, una llamada es aceptada únicamente cuando la suma de los anchos de banda de las llamadas existentes y de la nueva llamada no excedan la capacidad del enlace.

2.8 Distribución Estadística de Recursos

El control de tráfico en ATM es difícil debido al problema de lograr simultáneamente los principales objetivos: la distribución estadística de los recursos, con la finalidad de tener una mejor eficiencia y, la separación entre los flujos de tráfico para lograr tener la calidad de servicio deseada. Una ganancia en la eficiencia puede ser lograda mediante la multiplexación estadística de las conexiones VBR cuya tasa de pico total excede la tasa de transmisión física del enlace (mientras que la tasa total promedio es menor que la tasa del enlace). Si los flujos de tráfico son numerosos e independientes, entonces la probabilidad de que la tasa total instantánea pueda exceder la tasa del enlace (por ejemplo, cuando muchas fuentes están activas a sus tasa picos simultáneamente) va a ser pequeña. Este comportamiento es descrito por la Ley de los números grandes, la cual establece que la tasa total se aproximará a la tasa total promedio con probabilidad de uno cuando el número de flujos independientes se hace bastante grande. Para lograr una ganancia de multiplexación estadística (la relación de la tasa pico total con respecto a la tasa del

enlace), es necesario mantener un factor de utilización alto y maximizar el grado de distribución estadística de los recursos de la red.

Una consecuencia no deseable del multiplexaje estadístico es la posibilidad de que la QoS de una conexión pueda ser afectada adversamente por el tráfico en otra conexión. Por ejemplo, una ráfaga en un flujo puede llenar el buffer del multiplexor y por lo tanto incrementar el retardo de las celdas para todos los flujos. Muchos flujos de tráfico pueden estar en ráfagas simultáneamente y causar por lo tanto una saturación del buffer. La probabilidad de que el buffer se sature y se presenten retardos excesivos en las colas (síntomas de una congestión) es alto con cargas bastante grandes. Por lo tanto, es deseable mantener un factor de utilización bajo (el cual puede ser económicamente no factible) o por otro lado proporcionar una separación entre flujos de tráfico para reducir el efecto de las ráfagas en uno de los flujos en la QoS del otro flujo. Desafortunadamente, estas aproximaciones representan un conflicto con el objetivo de multiplexaje estadístico.

Las prioridades son un mecanismo útil para separar o modificar los efectos mutuos entre los flujos de tráfico. Las prioridades de retardo prescriben el orden en el cual las celdas que están en cola son ordenadas para la transmisión en enlaces compartidos; las prioridades de pérdidas especifican la ocupación preferencial del espacio del buffer compartido. El efecto del tráfico de baja prioridad en los retardos y del tráfico de alta prioridad es por lo tanto reducido. Sin embargo, el tráfico de alta prioridad tiene obviamente un gran efecto en los retardos del tráfico de baja prioridad.

Capítulo III

Control de Admisión usando aproximación Gamma

3.1 Introducción

El objetivo es determinar si una nueva llamada, que requiere de una conexión virtual con una calidad de servicio (QoS) específica (ancho de banda, probabilidad de pérdida, retardo, etc.), puede o no ser admitida. Para un control real de la red, los mensajes deberán ser enviados a lo largo de todo el camino virtual que proporciona la conexión, con la finalidad de determinar donde se cumple con los objetivos de calidad de servicio (QoS) sin afectar a las otras llamadas que se encuentran ya en progreso. En general, esto implica el verificar no solo el camino virtual específico dentro del cual el circuito virtual pudo ser establecido, sino todos los otros caminos virtuales que comparten parcial o totalmente la ruta que es usada en la red, punto a punto.

Una solución muy simple para este problema es el de asignar un ancho de banda específico para cada uno de los caminos virtuales a lo largo de todo el camino punto a punto y posteriormente asignar este ancho de banda a cada circuito virtual que esta usando el camino virtual, tomando como base el requerimiento de la tasa pico de transferencia. El camino seleccionado y la asignación del buffer a cada nodo del switch a lo largo del camino puede ser usado luego para garantizar la calidad de servicio QoS deseada. La desventaja en este método es que la asignación de tasa pico excluye el uso de multiplexación estadística, y requiere posiblemente un ancho de banda bastante grande del que puede ser usado en otro caso.

Otros algoritmos de control, más complejos, pueden proporcionar una multiplexación estadística permitiendo utilizar los recursos de la red más eficientemente. Sin embargo, mientras más complejo es el algoritmo más costosa y difícil será su implementación.

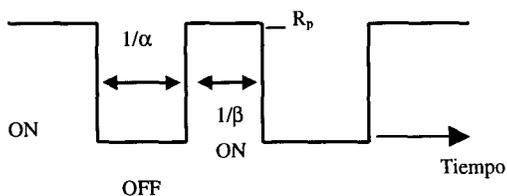
De esta manera, el objetivo principal es encontrar o establecer algoritmos de control más sencillos que puedan ser implementados para cumplir con una calidad de servicio dada y hacer un uso eficiente de la red; lo cual es en general, un problema bastante difícil de solucionar. Por lo tanto, se plantean soluciones sencillas, centradas exclusivamente a un solo punto de acceso en vez de considerar todo el camino virtual.

3.2 Caracterización del Tráfico

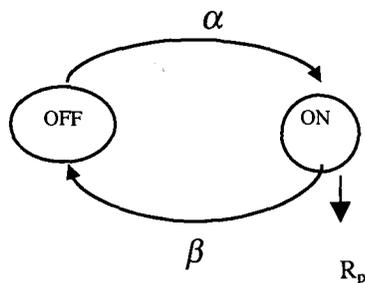
Para este estudio, se considerará los modelos de tráfico ON-OFF, los mismos que pueden ser usados para voz, fuentes tipo ráfagas, incluyendo imágenes y vídeo entre otras fuentes. Una fuente de tráfico ON-OFF requiere de al menos tres parámetros para representarla: la tasa pico de transmisión de una celda R_p , en unidades de celdas/segundos (o bps), la longitud promedio de una ráfaga $1/\beta$, en segundos, y el promedio del tiempo en OFF, $1/\alpha$, [6]. Estos parámetros se muestran en la Figura 3.1. Para un tráfico tipo ráfaga se tiene que $1/\beta \ll 1/\alpha$, o que $\alpha/\beta \ll 1$.

Estos tres parámetros son conocidos como “descriptores de tráfico”, y son proporcionados por el usuario junto con los requerimientos de calidad de servicio (QoS). Se considera que el comportamiento de los intervalos “ON” y “OFF” son exponenciales.

Cada fuente esta representada por una cadena de Markov de dos estados de tiempo continuo, [13]. Las fuentes heterogéneas tendrán cada una valores diferentes para los tres parámetros. Para fuentes homogéneas en cambio los tres valores serán los mismos. Se asume implícitamente que mientras dura el período de ráfaga, las celdas son transmitidas ya sea continua o periódicamente, así como en el caso de voz con una tasa pico R_p .



a. Modelo de Tráfico



b. Modelo de Markov

Figura 3.1: Fuente de tráfico ON-OFF

Además, se asumirá que la política usada para el ingreso a los buffers es FIFO. En este caso todos los buffers de las fuentes son combinados para formar un solo buffer que almacena todas las celdas provenientes de todas las fuentes en orden FIFO.

Cuando la estrategia FIFO es usada, debe haber un máximo número de conexiones o llamadas de cada clase para que pueda ser manejada por el enlace de acceso. Este número máximo depende de la estrategia de calendarización, la capacidad de acceso C_L , el número de clases, y los requerimientos de calidad de servicio QoS para cada uno. Este número máximo, por su parte, representan los vértices en una región k-dimensional que representa la región admisible para este sistema. La estrategia “óptima” puede por lo tanto

proporcionar la región más grande; mientras que la estrategia FIFO da la más pequeña. Por lo tanto, dada una región admisible una llamada será admitida si el sistema, con la llamada presente, todavía trabaja dentro de la región admisible. Una llamada es bloqueada si su aceptación puede llevar al sistema fuera de esta región permisible.

3.3 Multiplexaje Estadístico

El estudio se centralizará en el control de admisión de una sola clase de tráfico. Esto significa que cada conexión de usuario, con diferentes descriptores de tráfico, debe recibir el mismo grado de servicio. En términos de ATM, significa que se centralizará el estudio en un camino virtual (VP) en particular ya establecido y diseñado para proporcionar una calidad de servicio (QoS) dada. A este camino virtual VP se le asigna una capacidad de enlace C_L . El punto aquí es determinar cuantas llamadas o conexiones virtuales para cada tipo pueden ser manejadas. Para conexiones de usuarios homogéneas, el objetivo es determinar el número total de circuitos virtuales que pueden ser manejados. Alternativamente, para un número dado de conexiones que deben ser manejadas, uno desearía saber la capacidad C_L requerida para cumplir con el objetivo de un rendimiento dado.

Una manera obvia para admitir nuevas llamadas es determinando el número máximo de tal manera que la tasa promedio de transmisión (ancho de banda) no exceda a la capacidad C_L . En otras palabras esto proporciona el máximo multiplexaje estadístico o ganancia. En particular, para muchas fuentes tipo ráfagas ($\alpha/\beta \ll 1$) se puede esperar una alta ganancia. Este procedimiento claramente permite un número máximo de llamadas a ser admitidas si la transmisión FIFO es usada.

El problema con esta técnica es que no toma en cuenta las fluctuaciones estadísticas, y puede de hecho resultar en una probabilidad de pérdida de celdas bastante alta. El diseño en base a la asignación promedio de ancho de banda es equivalente a una utilización del enlace igual a uno. Esto es fácilmente demostrable para los modelos de fuentes de tráfico ON-OFF, como el que se muestra en la Figura 3.1. Se considerará un solo acceso de tipo FIFO al buffer que alimenta a un enlace con capacidad C_L , como se muestra en la Figura 3.2, que representa el modelo del multiplexor estadístico. Para simplicidad de análisis se asume fuentes homogéneas. De esta manera, el objetivo es

encontrar el número máximo de fuentes N que pueden ser multiplexadas. Para el caso contrario, dado N cual es la capacidad C_L necesaria?

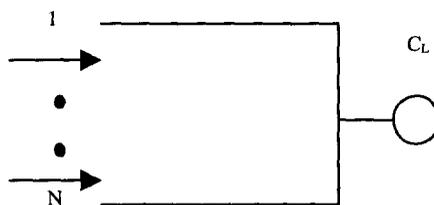


Figura 3.2: Multiplexación estadística de fuentes homogéneas

Para el modelo de fuente de tráfico como el de la Figura 3.1, la probabilidad de que la fuente se encuentre en el estado “ON” es $p = \alpha/(\alpha+\beta)$. La tasa promedio de transmisión es por lo tanto pR_p . Para N fuentes, la tasa promedio de transmisión es de NpR_p y la utilización esta dada por:

$$\rho = NpR_p/C_L \quad (3.1)$$

El establecer la tasa promedio a la capacidad del enlace C_L , para determinar el número máximo de fuentes que pueden ser multiplexadas es equivalente a tener $\rho = 1$. El mismo resultado se obtendrá cuando se tenga k clases de fuentes, cada una con sus propios descriptores de tráfico, las mismas que serán multiplexadas con el criterio de ocupación del ancho de banda promedio. De esta manera, la utilización máxima de 1 se obtiene cuando

$$\sum_{i=1}^k n_i p_i R_{pi} = C_L \quad (3.2)$$

Para el caso de fuentes homogéneas, la “mejor” estrategia de multiplexación en términos de utilización, o de ganancia de multiplexación, es el método de asignación del ancho de banda promedio. Con este método se obtiene un número máximo de llamadas N permitidas para una capacidad C_L dada; sin embargo, este método puede ser inaceptable en términos de pérdida de celdas. En otras palabras, este método proporciona un límite

superior en el número de llamadas que pueden ser aceptadas, o inversamente, proporciona un límite inferior en la capacidad C_L requerida para manejar un número de llamadas dadas.

Por otro lado, la estrategia de asignación pico garantiza que no haya pérdidas de celdas (esto es, el mejor rendimiento de pérdida de celdas) y proporciona un límite inferior en el número de llamadas que pueden ser aceptadas, dada una capacidad del enlace C_L .

Una representación gráfica de lo expuesto anteriormente se muestra en la Figura 3.3. En esta se muestra la relación entre N (el número de fuentes) y C_L (capacidad del enlace); así, la región admisible con una capacidad del enlace C_L dado y una QoS específica. Mientras los parámetros de QoS sean más estrictos, la curva de control de asignación se acerca más a la curva de asignación pico; y mientras estos son más flexibles, la curva de asignación se acerca a la línea de asignación promedio.

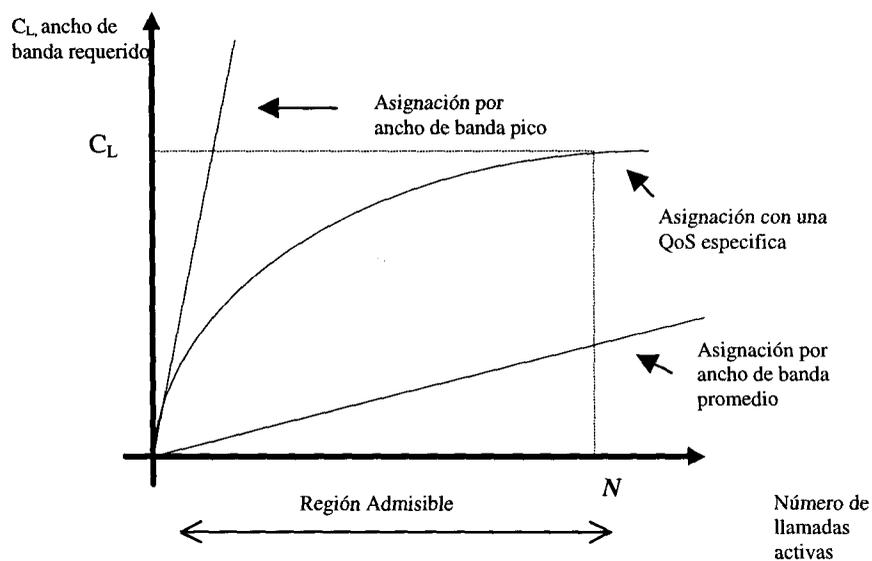


Figura 3.3: Control de Admisión, fuentes homogéneas ON-OFF

Una primera aproximación para un valor aceptable de la capacidad C_L es que este valor difiera del valor medio mR_p por algún múltiplo de la desviación estándar σR_p . De esta manera, una primera aproximación para el valor de C_L , puede ser:

$$C_L = (m + \mathbf{K}\sigma) R_p \quad (3.3)$$

Con una constante K a ser determinada, la cual varía con la calidad de servicio. De esta manera, la constante K se incrementará cuando la calidad de servicio (QoS) sea más estricta, y disminuirá acercándose a 0 cuando la calidad de servicio sea más flexible.

Hay que notar, que la distribución de la capacidad descrita por la ecuación (3.3) no está restringida únicamente a fuentes ON-OFF multiplexadas. Esta ecuación puede ser una aproximación correcta para fuentes VBR multiplexadas, con mR_p y σR_p como la media y desviación estándar, respectivamente, de las tasas de bit multiplexadas.

Como se mencionó anteriormente, para este tipo de fuentes se considerará que el valor medio está dado por $m = Np$ y la desviación estándar como $\sigma^2 = Np(1-p) = m(1-p)$. Con la finalidad de simplificar el análisis y la notación, se puede expresar la ecuación (3.3) en su forma normalizada equivalente,

$$C = m + K\sigma = Np + K\sqrt{Np(1-p)} \quad (3.4)$$

con $C = C_1/R_p$ el parámetro de capacidad $C \leq N$.

Por otro lado, las fuentes ON-OFF multiplexadas pueden ser representadas por un proceso de nacimiento y muerte de $(N + 1)$ estados, como se muestra en la Figura 3.4. Siendo los estados de interés dentro de este proceso, el de saturación J_o , el estado anterior a la saturación $J_u = J_o - 1$ y el parámetro de capacidad C establecido entre estos dos estados.

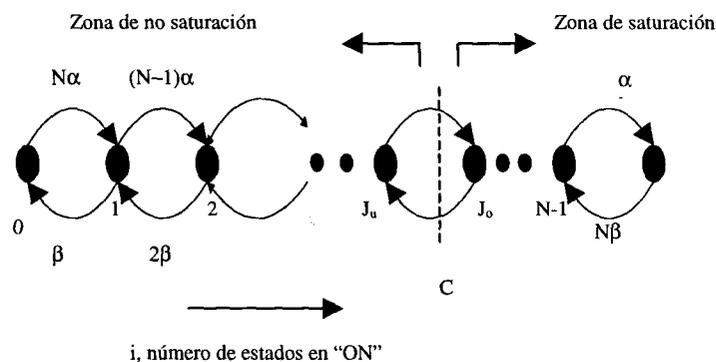


Figura 3.4: Modelo de n fuentes de voz

Es decir, el estado J_o representa el techo del valor de C y J_u en cambio es el piso de C . De esta manera, todos los estados $i \geq J_o$ representan la región de saturación en la cual el buffer tiende a estar lleno; mientras que, $i \leq J_u$ corresponde a la región de no saturación, en la cual el buffer tiende a estar vacío. El número promedio de fuentes en "ON", m , se encuentra en la región de no saturación, como se puede observar de la ecuación (3.4) o de la expresión $\rho = NpR_p/C_L = m/C < 1$.

Por lo tanto para poder determinar la capacidad necesaria se debe obtener el valor de la constante K de la ecuación (3.4) y por medio de este valor, establecer un procedimiento para el control de admisión de las fuentes ON – OFF. Si se considera que la calidad de servicio QoS deseada esta dada ya sea por la probabilidad de pérdida de celdas P_L o por la probabilidad de estar en la zona de saturación ϵ , estos dos parámetros serán las medidas de calidad de servicio que se usará generalmente para el control de admisión.

Una aproximación conservadora de la probabilidad de pérdida de celdas es expresarle como la relación que existe entre la tasa de entrada de las celdas cuando el estado $(N + 1)$ del sistema se encuentra en la región de sobrecarga con respecto a la tasa promedio de entrada. De esta manera se tiene que:

$$P_L = \sum_{i=J_o}^N (i - C)\pi_i / m \quad (3.5)$$

donde π_i es la probabilidad de que el sistema se encuentra en el estado i . Esta es una aproximación bastante conservadora ya que asume que cualquier tasa de entrada por encima de la capacidad del enlace, va a resultar en una pérdida de celda. Estas aproximaciones ignoran totalmente al buffer. De esta manera, dada una P_L , y el descriptor de tráfico p de que una fuente este en "ON", se puede encontrar la relación deseada entre la capacidad C y el número de llamadas N conectadas.

Otra medida relacionada con pérdidas es la probabilidad de estar en la región de saturación, esto es, la probabilidad de que la tasa de entrada exceda la capacidad C . Si se le denomina a este parámetro como ϵ , se tiene que

$$\varepsilon = \sum_{i=J_0}^N \pi_i \quad (3.6)$$

Podremos observar posteriormente que para valores bien pequeños de P_L y ε , como se desea tener en la práctica, cualquiera de los dos parámetros proporcionan una relación bastante similar de $C - N$; esto es, el mismo valor de \mathbf{K} .

3.4 Aproximación Gamma

Para poder tener valores pequeños de P_L y/o ε , se necesita un número relativamente grande de fuentes a ser multiplexadas. De esta manera, se asume $N \gg 1$ y $p \ll 1$; esto es, que las fuentes sean tipo ráfagas. Estas suposiciones facilitan el cálculo de la constante \mathbf{K} . Una vez calculada la constante \mathbf{K} , se puede asumir heurísticamente, que la región admisible o la capacidad equivalente C dado un valor de N obedece a la ecuación (3.4).

Cuando se tiene un valor de N bastante grande y p pequeño, la probabilidad de que el sistema se encuentre en el estado i puede ser calculada usando la aproximación de la distribución binomial. De esta manera, la distribución binomial con parámetros N , p , media $m = Np$ y varianza $\sigma^2 = Np(1-p)$, está dada por

$$\pi_i = \binom{N}{i} p^i (1-p)^{N-i} = \frac{N!}{i!(N-i)!} p^i (1-p)^{N-i} \quad (3.7)$$

Usando la fórmula de Stirling, $n! \approx \sqrt{2\pi n} (n/e)^n$, para realizar las aproximaciones de los factoriales de la ecuación (3.7) y considerando que $\sigma^2 = Np(1-p) \gg 1$, se tiene que

$$\pi_i \approx \frac{1}{\sigma} \phi\left(\frac{i-m}{\sigma}\right) \quad (3.8)$$

Si se observa, la ecuación (3.8) corresponde a la función de densidad de probabilidad de función Gaussiana. De esta manera, se logra expresar la distribución binomial en términos de una distribución Gaussiana con media $m = Np$ y varianza $\sigma^2 = Np(1-p)$. Sin embargo, la cola de la distribución Gaussiana, que corresponde a la región

de saturación, no representa una buena aproximación, como se muestra en la Figura 3.5, razón por la cual es necesario plantear una mejor aproximación.

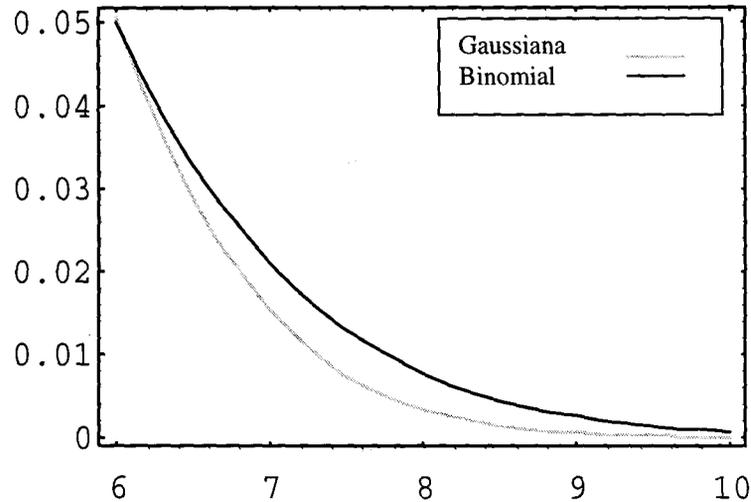


Figura 3.5: Distribución Binomial y Gaussiana, para $N=150$ y $p=0.02$

Expresando de otra manera la ecuación (3.8) se tiene que

$$\pi_i \approx \frac{1}{\sigma} \phi\left(\frac{i-m}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(i-m)^2/2\sigma^2} \quad (3.9)$$

Usando nuevamente una de las aproximaciones de Stirling para la función gamma dada por $\Gamma(\alpha+1) \approx \sqrt{2\pi\alpha} (\alpha/e)^\alpha$, la ecuación (3.9) puede aproximarse a la siguiente ecuación

$$\pi_i \approx \frac{\lambda(\lambda i)^{\alpha-1} e^{-\lambda i}}{\Gamma(\alpha)} \quad (3.10)$$

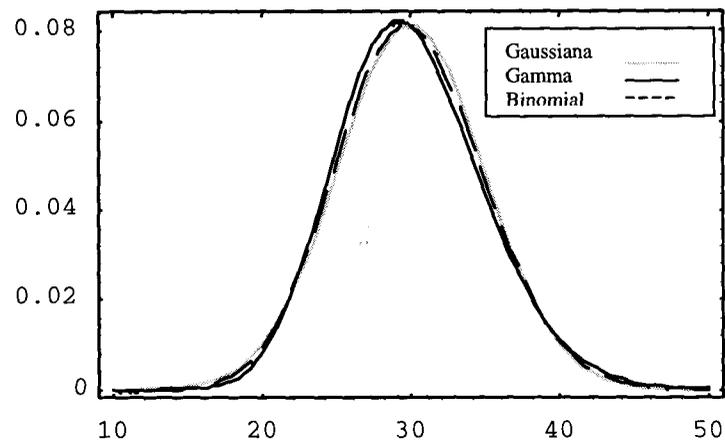
La ecuación (3.10) corresponde a la función de densidad de probabilidad de la variable gamma. De esta manera, la probabilidad de que el sistema se encuentre en el estado i puede ser aproximada a una distribución gamma con media $m = \alpha/\lambda$ y varianza $\sigma^2 = \alpha/\lambda^2$. Siendo α y λ los parámetros de la función gamma. Al aproximar la distribución binomial a una distribución gamma se tiene que

$$m = Np = \frac{\alpha}{\lambda} \qquad \sigma^2 = Np(1-p) = \frac{\alpha}{\lambda^2} \qquad (3.11)$$

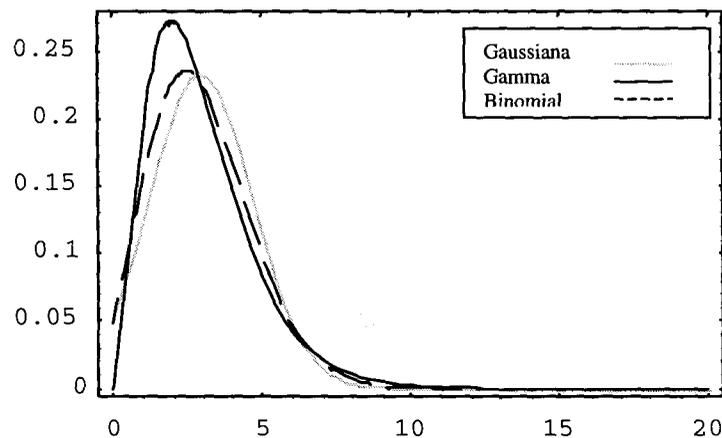
trabajando con las ecuaciones anteriores se puede encontrar valores para α y λ en función del número de llamadas N y la probabilidad de que la fuente este en "ON", de esta manera

$$\lambda = \frac{1}{1-p} \qquad \alpha = \frac{Np}{1-p} \qquad (3.12)$$

A continuación en la Figura 3.6 se presentan gráficas con las aproximaciones tanto Gaussiana y Gamma a la distribución binomial.



(a) $N=150, p=0.2$



(b) $N=150, p=0.02$

Figura 3.6: Representación de las Distribuciones Binomial, Gaussiana y Gamma

En la Figura 3.6.a se presentan los tres tipos de distribuciones para un número de fuentes $N = 150$ y una probabilidad $p = 0.2$. La Figura 3.6.b representa los tres tipos de distribuciones para igual número de fuentes y una probabilidad $p = 0.02$. Como puede observarse en la Figura 3.6 mientras más pequeña es p , es decir mientras más ráfagas son las fuentes, la distribución Gamma representa una mejor aproximación de la distribución binomial en la zona de saturación que está representada por la cola de la distribución binomial.

En la Figura 3.7 se puede observar más a detalle las colas de los tres tipos de distribuciones para cuando $N = 150$ y $p = 0.02$ y como se puede apreciar la distribución que presenta una mejor aproximación a la binomial en esta zona es la distribución gamma.

La región de saturación corresponde a la cola de la distribución, y es en esta zona la que se analizará el parámetro ϵ . Las sumas de las ecuaciones (3.5) y (3.6) son aproximadas por las integrales de la distribución Gamma sobre la misma región. Esto es, de hecho, otra razón para necesitar que $N \gg 1$.

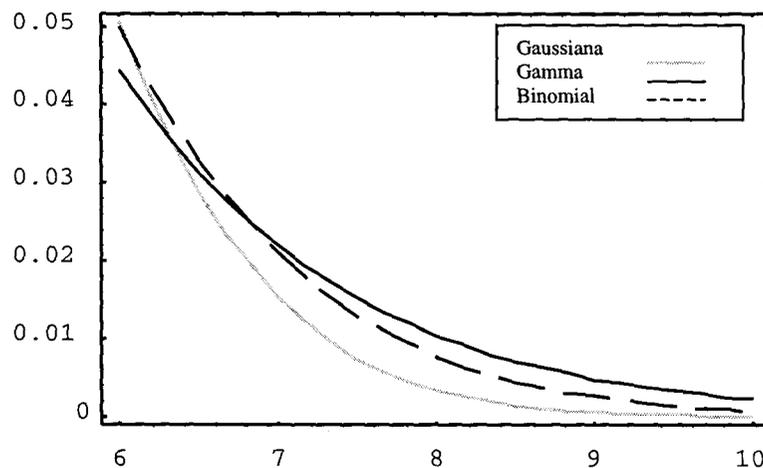


Figura 3. 7: Colas de las distribuciones Gaussiana, Gamma y Binomial

El convertir las sumas a integrales implica el convertir la variable aleatoria discreta a su variable aleatoria continua equivalente, como se muestra en la Figura 3.8.

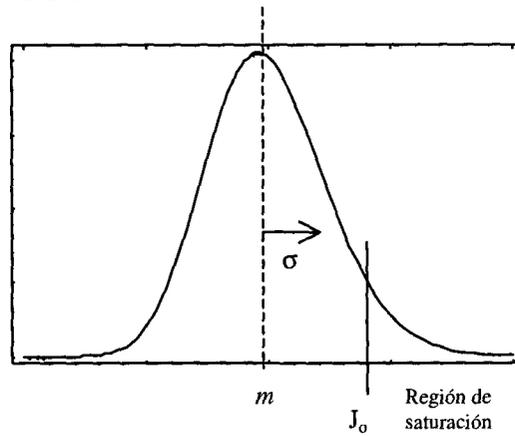


Figura 3.8: Aproximación Gamma a la distribución Binomial,
 $m=Np$ y $\sigma^2 = m(1-p)$

Las dos probabilidades P_L y ε dadas en las ecuaciones (3.5) y (3.6) son aproximadas a

$$P_L = \frac{\lambda}{\alpha} \int_{J_0}^{\infty} \frac{(x-c)\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \quad (3.13)$$

y

$$\varepsilon = \int_{J_0}^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \quad (3.14)$$

Tomando en cuenta que se realizó una aproximación a una distribución continua, se puede considerar $J_0 = C$, y realizar los cálculos como se indica a continuación.

La ecuación (3.13) puede ser escrita como

$$P_L = \frac{\lambda}{\alpha} \int_c^{\infty} \frac{(x-c)\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \quad (3.15)$$

$$P_L = \frac{\lambda * \lambda^\alpha}{\alpha * \Gamma(\alpha)} \left[\int_c^{\infty} x^\alpha e^{-\lambda x} dx - C \int_c^{\infty} x^{\alpha-1} e^{-\lambda x} dx \right] \quad (3.16)$$

$$P_L = \frac{1}{\Gamma(\alpha)\alpha} \left\{ \frac{\alpha \Gamma(\alpha, \lambda C)}{\lambda} + \frac{(\lambda C)^\alpha e^{-\lambda C}}{\lambda} - \Gamma(\alpha, \lambda C) \right\} \quad (3.17)$$

$$P_L = \frac{\alpha - \lambda C}{\alpha} \frac{\Gamma(\alpha, \lambda C)}{\Gamma(\alpha)} + \frac{(\lambda C)^\alpha e^{-\lambda C}}{\Gamma(\alpha+1)} \quad (3.18)$$

En la Figura 3.9 se puede observar la gráfica de la probabilidad de pérdida P_L con respecto a la capacidad, por motivo de análisis y de comparación se presentan los dos tipos de aproximaciones, la correspondiente a la distribución Gamma y la distribución Gaussiana.

Por otro lado, para evaluar el valor de ε se considera la ecuación (3.14) y la consideración de que $J_0 = C$, de esta manera se tiene que:

$$\varepsilon = \int_c^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \quad (3.19)$$

$$\varepsilon = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_c^{\infty} x^{\alpha-1} e^{-\lambda x} dx \quad (3.20)$$

$$\varepsilon = \frac{\Gamma(\alpha, \lambda C)}{\Gamma(\alpha)} \quad (3.21)$$

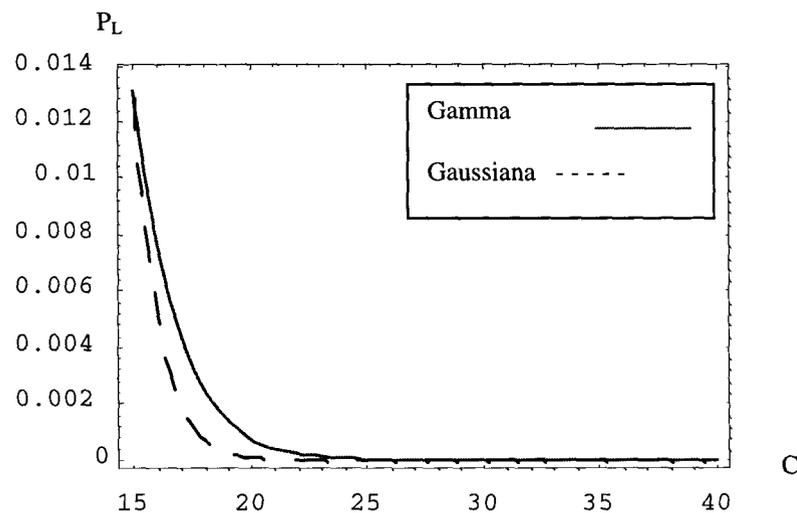


Figura No. 3.9: Representación de la Probabilidad de Pérdida P_L vs. Capacidad, Para $N = 500$ y $p = 0.02$

En la Figura 3.10 se puede observar la gráfica correspondiente a la probabilidad de saturación ε con respecto a la capacidad, tanto para la aproximación por la distribución Gamma como para la distribución Gaussiana.

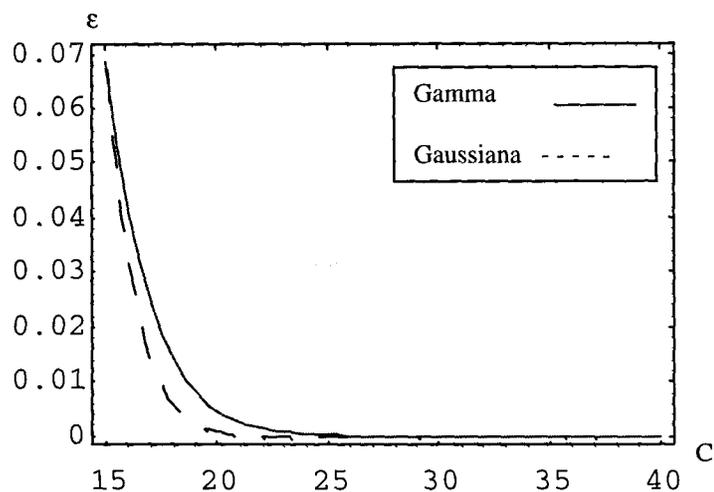


Figura 3.10: Representación de la probabilidad de saturación ε vs. Capacidad, Para $N = 500$ y $p = 0.02$

Trabajando con las ecuaciones (3.18) y (3.21) la probabilidad de pérdida de celdas P_L puede expresarse en función de ε , como se indica a continuación:

$$P_L = \frac{\alpha - \lambda C}{\alpha} \varepsilon + \frac{(\lambda C)^\alpha e^{-\lambda C}}{\Gamma(\alpha + 1)} \quad (3.22)$$

Además, a ε puede expresarse de la siguiente manera:

$$\varepsilon = \frac{\Gamma(\alpha, \lambda C)}{\Gamma(\alpha)} = \frac{(\alpha - 1)! e^{-\lambda C} \sum_{i=0}^{\alpha-1} \frac{(\lambda C)^i}{i!}}{(\alpha - 1)!} = e^{-\lambda C} \sum_{i=0}^{\alpha-1} \frac{(\lambda C)^i}{i!} \quad (3.23)$$

Reemplazando esta expresión de ε (3.23) en la ecuación de probabilidad de pérdidas de celdas P_L (3.22) se tiene que:

$$P_L = \frac{\alpha - \lambda C}{\alpha} e^{-\lambda C} \sum_{i=0}^{\alpha-1} \frac{(\lambda C)^i}{i!} + \frac{(\lambda C)^\alpha e^{-\lambda C}}{\alpha \Gamma(\alpha)} \quad (3.24)$$

$$P_L = \frac{\alpha - \lambda C}{\alpha} e^{-\lambda C} \left[1 + (\lambda C) + \frac{(\lambda C)^2}{2!} + \dots + \frac{(\lambda C)^{\alpha-1}}{(\alpha-1)!} \right] + \frac{(\lambda C)^\alpha e^{-\lambda C}}{\alpha(\alpha-1)!} \quad (3.25)$$

$$P_L = \frac{e^{-\lambda C}}{\alpha} \sum_{i=0}^{\alpha-1} (\alpha-i) \frac{(\lambda C)^i}{i!} = \varepsilon - \frac{e^{-\lambda C}}{\alpha} \sum_{i=0}^{\alpha-1} i \frac{(\lambda C)^i}{i!} \quad (3.26)$$

$$P_L = \varepsilon - \frac{\lambda C}{\alpha} e^{-\lambda C} \sum_{i=0}^{\alpha-2} \frac{(\lambda C)^i}{i!} \quad (3.27)$$

Como se puede observar la probabilidad de pérdida difiere de la probabilidad de saturación ε por el factor

$$\frac{\lambda C}{\alpha} e^{-\lambda C} \sum_{i=0}^{\alpha-2} \frac{(\lambda C)^i}{i!}$$

en general se puede observar que la aproximación de ε es mucho más conservativa que la de pérdida de celdas P_L que resulta en una capacidad más grande.

Como se expuso anteriormente, es necesario encontrar un valor de K que permita determinar o calcular el valor de la capacidad. Si se toma en cuenta la ecuación (3.4) planteada anteriormente:

$$C = m + K\sigma = Np + K\sqrt{Np(1-p)}$$

y si en esta ecuación reemplazamos los valores del valor medio y varianza por los correspondientes a la distribución Gamma se tiene que

$$C = \frac{\alpha}{\lambda} + K \frac{\sqrt{\alpha}}{\lambda} \quad (3.28)$$

despejando de la ecuación (3.28) el valor de K se tiene que:

$$K = \frac{\lambda C - \alpha}{\sqrt{\alpha}} \quad (3.29)$$

de esta manera, el valor de la constante K se encuentra en función de los parámetros λ y α de la distribución Gamma así como de un valor de capacidad dada.

Para poder determinar el valor de K usando la ecuación (3.29) es necesario establecer antes un valor para α . De esta manera, y tomando en cuenta que el objetivo es encontrar la capacidad o el número máximo de llamadas que pueden ser aceptadas para que cumplan con una calidad de servicio dada, se puede establecer como dato o como parámetro conocido el valor de la probabilidad de estar en la zona de saturación ε . Asumiendo lo anteriormente expuesto el parámetro α para la distribución Gamma se calculará usando la ecuación (3.21) la cual puede expresarse de la siguiente manera:

$$\Gamma(\alpha, \lambda C) - \varepsilon \Gamma(\alpha) = 0 \quad (3.30)$$

Para entender más claramente lo expuesto, consideremos el siguiente ejemplo para el que $p = 0.02$ (este valor representa una fuente de ráfagas la cual esta en "OFF" el 98% de las veces y únicamente el 2% se encuentra en "ON"). Para encontrar el valor de la constante K es necesario realizar ciertos cálculos previos y consideraciones. Para este ejemplo vamos a considerar una Capacidad igual a 26.5 con una calidad de servicio QoS, $\varepsilon = 0.0007$, estos valores se reemplazan en la ecuación (3.30) y se calcula la raíz de la ecuación, teniendo como resultado, para este ejemplo, un valor de $\alpha = 10.2013$.

Con esta valor de α se sustituyen estos valores en la ecuación (3.29), obteniendo un valor de $K = 5.27232$, este valor se sustituye en la ecuación (3.4) de la capacidad, obteniendo lo siguiente:

$$C = 0.738125 \sqrt{N} + 0.02 N$$

Esta ecuación es graficada en la Figura 3.11, en la que se presenta la curva correspondiente a la relación entre la capacidad y el número de fuentes N , además y para motivo de análisis se presenta también la curva que se obtiene con la distribución Gaussiana. Para valores grandes de ε (o P_L), la constante K en la ecuación disminuirá, lo cual hace que la ecuación de la capacidad se acerque a la curva de asignación por ancho de banda promedio. Para valores pequeños de ε (o P_L) esta curva se acercará más a la curva correspondiente de asignación pico.

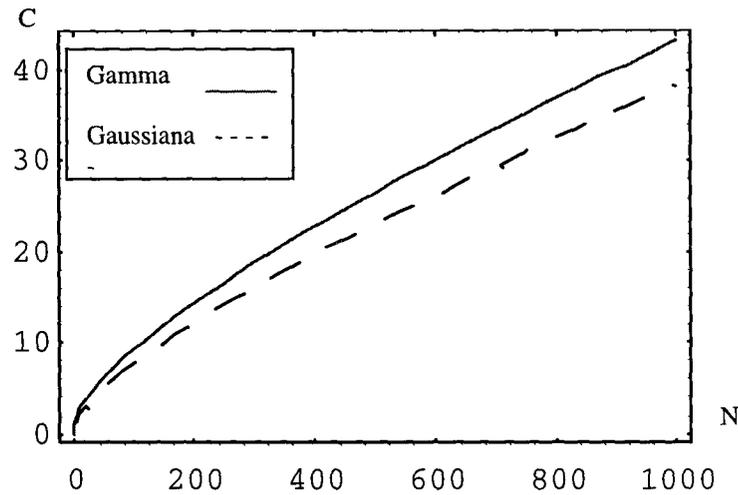


Figura 3.11: Control de Admisión, $p = 0.02$.

La expresión de la capacidad normalizada define la capacidad que se requiere para acceder al buffer con la finalidad de poder manejar una fuente de tráfico cuya tasa de bit promedio normalizada es m , con una desviación estándar sobre m de σ . (Hay que recordar que la normalización es con respecto a la tasa de bit pico R_p , de tal manera que la tasa actual del enlace es de $C_L = C \cdot R_p$). Este valor por lo general se le conoce como la capacidad “equivalente” o “efectiva” de la fuente. Para N fuentes ON-OFF homogéneas multiplexadas, se tiene $m = Np$ y $\sigma^2 = Np(1-p)$, con $p = \alpha/(\alpha + \beta)$ la probabilidad de que una fuente este en el estado “ON”. Se puede invertir esta expresión fácilmente en este caso para obtener un valor específico para N , el número de conexiones ON-OFF que son permitidas, dada la capacidad actual del enlace C_L (o en su forma normalizada para $C = C_L / R_p$) y el parámetro de calidad de servicio ϵ .

Específicamente, trabajando con la ecuación (3.4) y despejando de esta el valor de N , se tiene que:

$$N = \frac{C}{p} + \frac{K^2(1-p)}{2p} - \frac{1}{2p} \sqrt{K^2(1-p)[K^2(1-p) + 4C]} \quad (3.31)$$

La ecuación anterior define la región admisible. El control de admisión usando esta relación admitirá una nueva llamada si existe un número menor de N actualmente

conectadas. La Figura 3.12 muestra las regiones admisibles considerando diferentes valores de ϵ para una distribución Gamma.

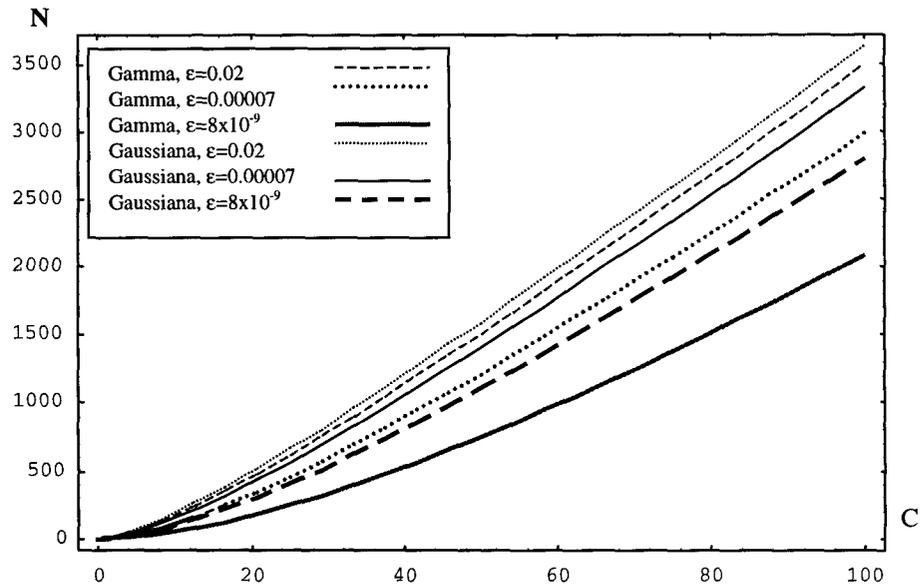


Figura 3.12: Control de Admisión para diferentes valores de ϵ , $p = 0.02$

Además de proporcionar una región admisible específica para este caso de fuentes homogéneas ON-OFF multiplexadas, la ecuación 3.31 permite el enfocarnos en la ventaja de obtener las fuentes por medio de un multiplexaje estadístico. En otras palabras, las N fuentes, una vez que han sido admitidas por el control de admisión, son combinadas estadísticamente en el acceso del buffer, con celdas transmitidas por cada una en un orden FIFO sobre el enlace. Con este análisis se puede definir una ganancia multiplexada estadísticamente que será calculada combinando uniformemente el arribo de las celdas de las fuentes en el buffer.

Esta ganancia G_ϵ , la cual es especificada por el parámetro de QoS ϵ y más generalmente por la probabilidad de pérdida de celdas P_L , es el incremento en conexiones permitidas para el caso en el cual la distribución de tasa de bit pico ha sido usada. Si se usa el control de admisión con la tasa de bit pico, el número de llamadas permitidas por el sistema será

$$C_L / R_p = C \quad (3.32)$$

Esto resultará en una pérdida de celdas nula, esto es $P_L(o \epsilon) = 0$. En términos del modelo de multiplexación, esto significa que el número de conexiones permitidas es tal que mantiene al sistema siempre en la región de no saturación. La ganancia de multiplexación se obtiene permitiendo que el número N se incremente por debajo de C , llegando de esta manera a pérdidas de celdas. De esta manera la ganancia multiplexada es definida como

$$G_\epsilon = N/C \quad (3.33)$$

con el parámetro de QoS incluido en el valor de C .

El máximo valor posible de la ganancia se encuentra fácilmente. Esta se obtiene cuando el control de admisión se basa en la asignación por ancho de banda promedio, esto es, cuando el número de llamadas permitidas se determina para $mR_p = C_L$ o $m = C$. Tomando en cuenta que $m = Np$ para el modelo de fuentes ON-OFF, la ganancia máxima multiplexada G está dada por

$$G = N/C = 1/p \quad (3.30)$$

Analizando la ecuación anterior se puede observar que para tráfico altamente de ráfagas, con $p \ll 1$, $G = 1/p$ puede ser bastante grande. Mientras que la probabilidad p de que una fuente este en el estado "ON" se incrementa, la ganancia posible por multiplexación de fuentes disminuye. Las fuentes que están en ON todo el tiempo, con $p=1$, no pueden ser multiplexadas estadísticamente, el control de tasa pico debe ser usado, y la ganancia $G = 1$.

La ganancia actual G_ϵ es posible, para una capacidad C normalizada, el parámetro p , y una probabilidad de pérdida ϵ , se obtiene directamente de la ecuación (3.31). De esta manera, la ganancia puede ser definida por

$$G_\epsilon = \frac{1}{p} + \frac{K^2(1-p)}{2pC} - \frac{1}{2pC} \sqrt{K^2(1-p)[K^2(1-p)+4C]} \quad (3.35)$$

El parámetro K depende de la probabilidad de pérdida ϵ . Hay que observar que mientras C se incrementa o K disminuye, se puede tener una mayor ganancia.

De esta manera, se puede concluir que una ganancia mayor es posible siempre y cuando la capacidad disponible C sea incrementada, lo cual correspondería a un mayor número de fuentes que pueden ser conectadas y por lo tanto la utilización se incrementaría.

Capítulo IV

Análisis de Resultados

4.1 Introducción

Como se mencionó anteriormente el objetivo del presente trabajo es el proponer un nuevo esquema de control de admisión basado en un método para estimar la probabilidad de pérdida de celdas y la probabilidad de estar en la zona de saturación de los modelos de fuentes de tráfico ON – OFF.

Para esto, se ha propuesto una nueva aproximación a la variable discreta de la probabilidad de que el sistema se encuentre en un estado determinado, la cual esta dada por la distribución binomial. Como se mencionó en el capítulo anterior el método que se propone es el aproximar esta distribución binomial a una distribución gamma con la finalidad de obtener mejores resultados que los que se obtienen con la aproximación

Gaussiana que es la aproximación que por lo general es más transparente realizar a una distribución binomial.

De esta manera, en este capítulo se presentará una serie de gráficas comparativas con la distribución binomial en orden de establecer la ventaja que conlleva el realizar este tipo de aproximaciones y que por lo tanto se puede llegar a tener valores más reales en el momento de determinar los parámetros de calidad de servicio para una fuente de tráfico dada.

4.2 Aproximación Gamma a la Función Binomial

La primera aproximación que se realiza en el análisis del modelo de fuente de tráfico ON – OFF, es el considerar que la probabilidad de que el sistema se encuentre en el estado i pueda ser expresada en función de una distribución gamma. Para un N grande y p pequeña, se sabe que la probabilidad de que el sistema este en el estado i esta dada por la solución de una distribución binomial, cuya expresión es:

$$\pi_i = \binom{N}{i} p^i (1-p)^{N-i} \quad (4.1)$$

de esta manera el esquema que se plantea en el presente estudio es el realizar una aproximación de la distribución binomial a una distribución gamma, de tal manera que la cola de la distribución binomial represente una mejor aproximación que la que se obtiene cuando se hace estas aproximaciones mediante la distribución Gaussiana. Para esto, la probabilidad de que el sistema se encuentre en el estado i es aproximada mediante una distribución gamma y está dada por

$$\pi_i \approx \frac{\lambda(\lambda_i)^{\alpha-1} e^{-\lambda_i}}{\Gamma(\alpha)} \quad (4.2)$$

Como se dijo anteriormente, la región de saturación de las fuentes de tráfico se encuentran representadas por la cola de la distribución binomial; es por este motivo que se plantea una nueva alternativa de aproximación a esta zona que resulte más exacta que la que se consigue con la aproximación Gaussiana.

De esta manera, se presenta a continuación una serie de gráficas en la que se encuentran representados los tres tipos de distribuciones. Debido que la zona de interés es la zona de saturación, se ha graficado únicamente la cola de estos tres tipos de distribuciones.

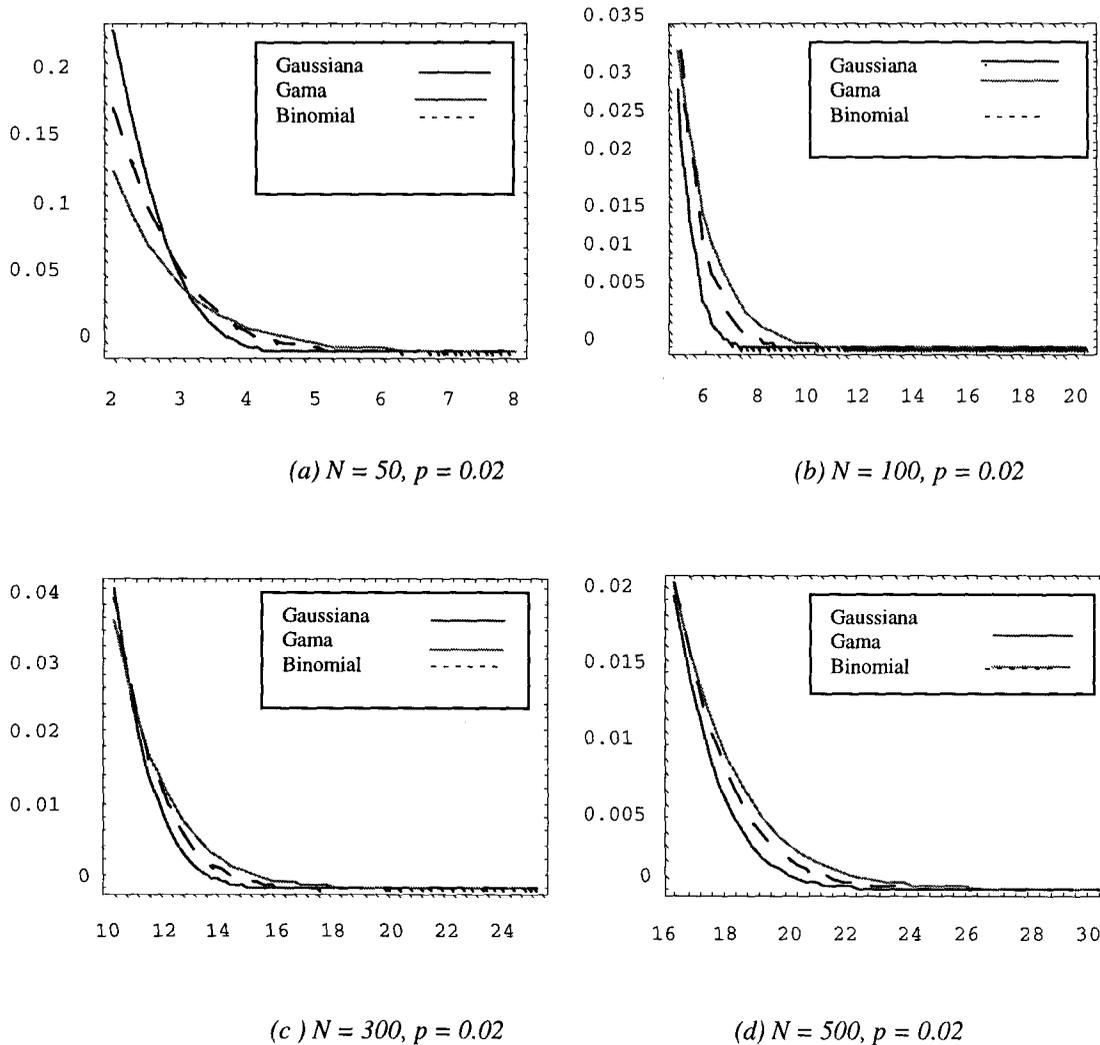


Figura 4.1: Colas de la distribución Binomial, Gaussiana y Gamma

En la figura 4.1 puede observarse que la zona de saturación representada por la cola de la distribución binomial tiene una mejor aproximación mediante la distribución Gamma, ya que el envolvente de la distribución Gaussiana es menor al envolvente de la distribución binomial. Como puede observarse mientras más grande es N , número de

llamadas, la curva de la distribución Gamma se acerca más a la distribución binomial. Adicionalmente, se puede observar que para valores de N grandes la aproximación Gamma de la distribución binomial es óptima en el área de la cola, que representa el área de interés para el análisis de los parámetros de la calidad de servicio, ya que para el resto de la curva la distribución Gamma tiende a alejarse de la distribución binomial.

Sin embargo, y a pesar de que la distribución Gamma tenga un comportamiento diferente a la distribución binomial en la mayoría de la curva, esta representa una solución válida ya que el análisis de los parámetros de calidad de servicio se los realiza únicamente en la zona de saturación y es precisamente en esta zona en que el comportamientos de las distribuciones binomial y gamma son bastante similares.

De esta manera, se puede concluir que un modelo de tráfico ON – OFF puede ser representado por un proceso de nacimiento muerte de $(N + 1)$ estados y que la probabilidad de que el sistema se encuentre en el estado i puede ser aproximada a una distribución Gamma con media $m = Np$ y varianza $\sigma^2 = Np(1-p)$, y caracterizada esta curva por sus parámetros que están definidos como $\lambda = 1/(1 - p)$ y $\alpha = Np/(1 - p)$.

4.3 Probabilidad de Pérdida P_L

Como se mencionó anteriormente una de las medidas de calidad de servicio que se usa para el control de admisión de llamadas es la probabilidad de pérdida de celdas. Este parámetro está definido como la relación de la tasa de entrada cuando el estado $(N + 1)$ del sistema se encuentra en la región de saturación con respecto a la tasa de entrada promedio. De esta manera, la expresión de la probabilidad de pérdida de celdas usando una aproximación gamma está dada por

$$P_L = \frac{\alpha - \lambda C}{\alpha} \frac{\Gamma(\alpha, \lambda C)}{\Gamma(\alpha)} + \frac{(\lambda C)^\alpha e^{-\lambda C}}{\Gamma(\alpha + 1)} \quad (4.3)$$

Con la finalidad de poder comparar el comportamiento de los dos tipos de aproximaciones se grafica también la aproximación Gaussiana, que está dada por la siguiente expresión

$$P_L = \frac{\sigma^3}{\sqrt{2\pi}(c-m)^2 m} e^{-(c-m)^2/2\sigma^2} \quad (4.4)$$

A continuación se presentan varias gráficas de la probabilidad de pérdida de celdas considerando los dos tipos de aproximaciones tanto la Gamma que es la que se propone en este estudio, así como la Gaussiana que es la aproximación comúnmente usada.

Como puede observarse de la figura (4.2), dado un valor determinado del parámetro de calidad de servicio, que para este caso es la P_L , puede observarse que para que N número de llamadas sean aceptadas la capacidad del canal tiene que ser mayor mediante la aproximación gamma a la que sería necesaria usando la aproximación Gaussiana. De esta manera, mientras menos pérdidas se desee tener en el enlace, el canal debe ser dimensionado de un tamaño ligeramente mayor al considerado convencionalmente para la solución Gaussiana. Las gráficas corresponden a fuentes de tráfico relativamente de tipo ráfaga.

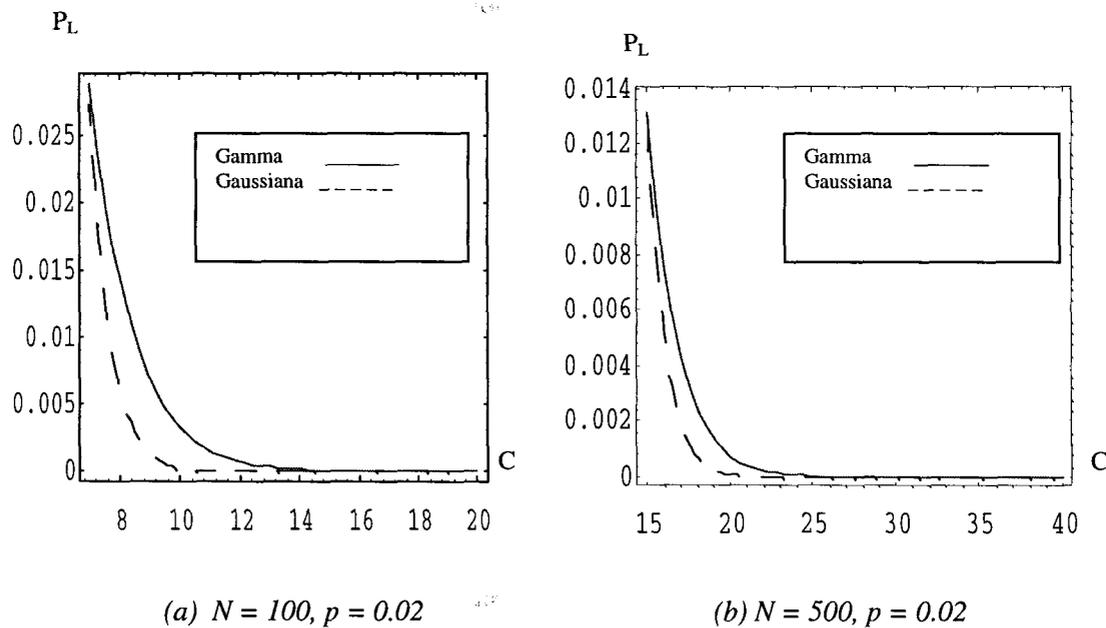


Figura 4.2: Probabilidad de Pérdida de celdas

4.4 Probabilidad de Saturación ε

Otro de los parámetros que se toman en cuenta para determinar la calidad de servicio y que ha sido analizada es la probabilidad de estar en la región de sobrecarga, o en otras palabras, la probabilidad de que la tasa de entrada exceda la capacidad del enlace.

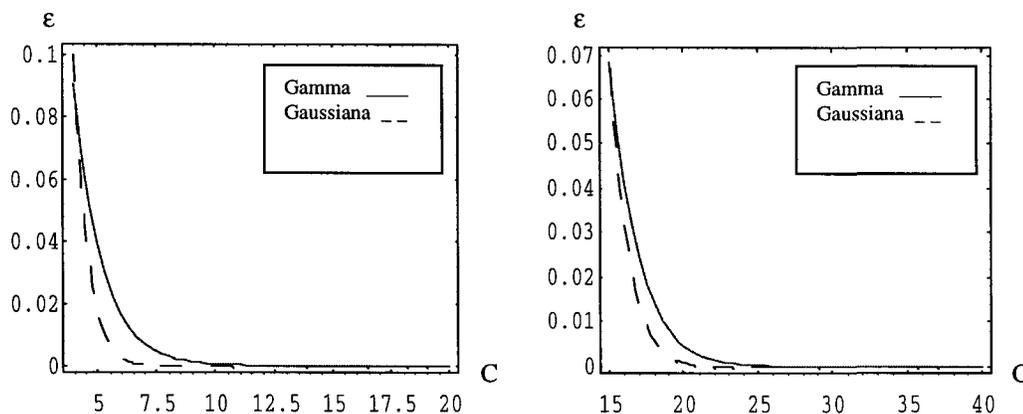
Al igual que en el caso anterior de la probabilidad de pérdidas, P_L , la probabilidad de estar en la región de sobrecarga puede obtenerse usando una aproximación gamma, y su expresión está dada por:

$$\varepsilon = \frac{\Gamma(\alpha, \lambda C)}{\Gamma(\alpha)} \quad (4.5)$$

Por otro lado, la expresión de esta probabilidad mediante el análisis de la aproximación Gaussiana es la que se muestra en la ecuación (4.6)

$$\varepsilon = \frac{\sigma}{\sqrt{2\pi}(c-m)} e^{-(c-m)^2 / 2\sigma^2} \quad (4.6)$$

Para poder analizar la alternativa planteada con la aproximación gamma se han graficado las dos alternativas



(a) $N = 100, p = 0.02$

(b) $N = 500, p = 0.02$

Figura 4.3: Probabilidad de Saturación ε

De las gráficas obtenidas puede concluirse que el esquema de control de admisión mediante la aproximación Gamma es más estricto que el que se obtiene con la aproximación Gaussiana. Como puede observarse en la Figura 4.3 para un valor determinado de ϵ y un número dado de llamadas que tienen que ser aceptadas el dimensionamiento de la capacidad del enlace tiene que ser mayor cuando se considera el criterio de aproximación Gamma que la que sería necesaria si la aproximación se la realiza con la distribución Gaussiana.

4.5 Control de Admisión

Como ya se ha mencionado a lo largo de este trabajo uno de los objetivos del control de admisión es el mantener la carga de la red a niveles tales que permitan conseguir los objetivos de rendimiento con una calidad de servicio dada.

De esta manera, es necesario encontrar una región admisible; es decir, dado un conjunto de fuentes con determinados descriptores de tráfico, es necesario determinar el número máximo de llamadas N que pueden ser admitidas para una calidad de servicio dada.

La expresión para el número de fuentes esta dada por la siguiente ecuación

$$N = \frac{C}{p} + \frac{K^2(1-p)}{2p} - \frac{1}{2p} \sqrt{K^2(1-p)[K^2(1-p)+4C]} \quad (4.7)$$

esta expresión se usa para los dos tipos de aproximaciones y lo que difiere es el valor de la constante K , cuyo cálculo varía para cada una de las aproximaciones.

Con respecto a la aproximación Gamma la constante K se la encuentra mediante la siguiente expresión:

$$K = \frac{\lambda C - \alpha}{\sqrt{\alpha}} \quad (4.8)$$

en lo que respecta a la aproximación Gaussiana la constante K puede calcularse a partir de la expresión de la ecuación (4.9)

$$K = \sqrt{-\ln(2\pi) - 2 \ln \varepsilon} \quad (4.9)$$

A continuación se presenta graficada la ecuación para encontrar el número máximo de llamadas que pueden ser aceptadas para una calidad de servicio determinada establecida por la probabilidad de estar en la zona de saturación ε y para un valor de capacidad fija.

En la Figura 4.4 se encuentra graficada la expresión para encontrar el máximo número de llamadas que pueden ser aceptadas en un canal con diferente calidad de servicio. Como puede observarse, la aproximación Gamma establece un criterio más estricto con respecto al número de llamadas; de esta manera, para una capacidad del enlace dada el número de llamadas que pueden ser aceptadas por la red es menor usando el esquema de la aproximación Gamma, mientras que la solución con la aproximación Gaussiana es más flexible y se puede aceptar un mayor número de llamadas. La diferencia entre los criterios de admisión para los dos tipos de soluciones se vuelve más marcado para cuando el parámetro de calidad de servicio ε es más pequeño, para valores relativamente altos de la probabilidad de estar en la región de saturación la diferencia entre los dos tipos de soluciones es menor.

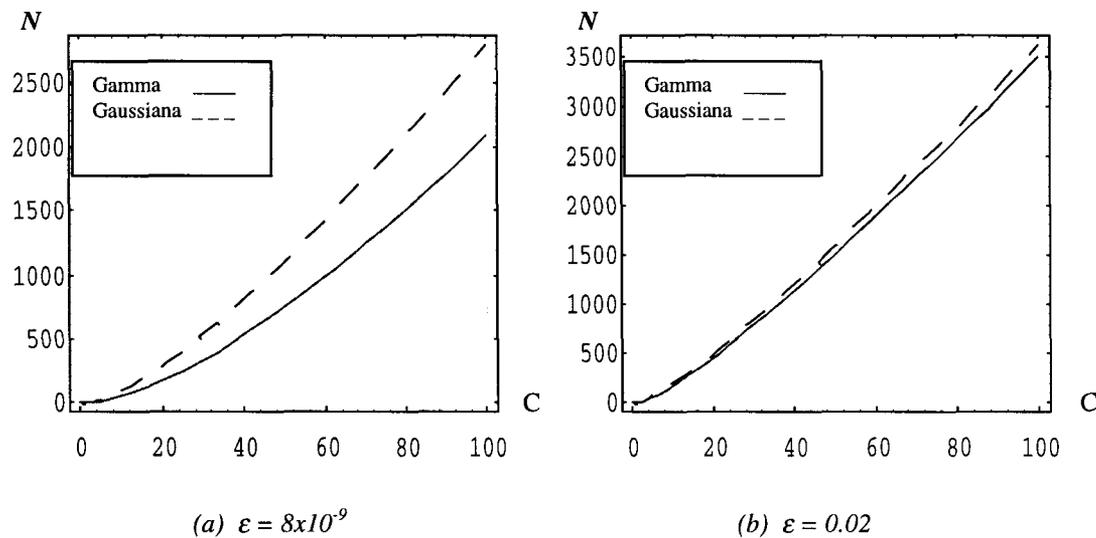


Figura 4.4: Control de Admisión

De esta manera, y siendo uno de los objetivos el tener valores pequeños para la probabilidad de estar en la región de saturación, el esquema de control de admisión con la aproximación Gamma representa un mejor control y garantiza un mejor rendimiento de la red.

Otro aspecto importante es el poder determinar la región admisible; la misma que esta establecida a través de la relación existente entre la capacidad del enlace y el número de llamadas que pueden ser aceptadas. De esta manera, es necesario considerar la expresión de la capacidad, la cual es la misma para los dos tipos de esquemas y la diferencia entre los dos tipos de soluciones radica al igual que en el caso anterior en el calculo de la constante K . La ecuación para el calculo de la capacidad en función del valor de una constante K es

$$C = m + K\sigma = Np + K\sqrt{Np(1-p)} \quad (4.10)$$

De la Figura 4.5 se puede concluir que para una capacidad del enlace dada, la región de admisión con el esquema de aproximación Gamma es menor que la que se obtiene con la aproximación Gaussiana. De igual manera, para un número máximo determinado de llamadas que pueden ser admitidas, la capacidad que se requiere para el enlace es menor al usar la aproximación Gaussiana, mientras que al usar el esquema de la aproximación Gamma es necesario considerar un mayor valor para la capacidad del enlace.

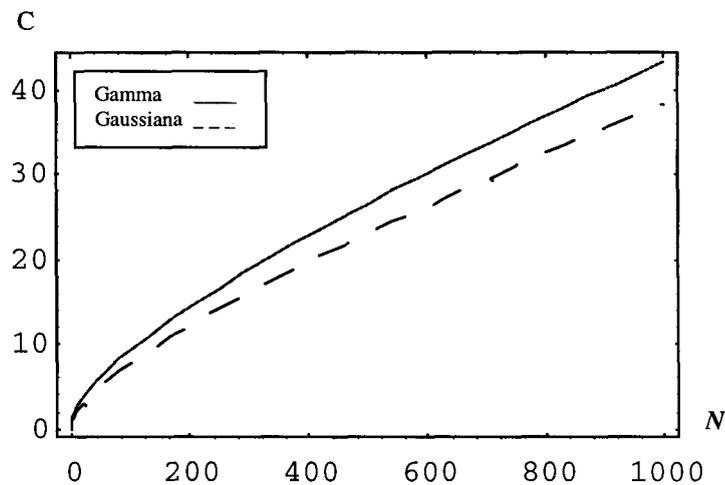


Figura 4.5: Región de admisión, $\epsilon = 0.00007$

Capítulo V

Conclusiones y Recomendaciones

En el presente trabajo se plantea un esquema de control de admisión de llamadas para redes de banda ancha, con la finalidad de mantener la probabilidad de pérdida de celdas, P_L , y la probabilidad de estar en la región de saturación, ϵ , por debajo de un valor específico y garantizar de esta manera un rendimiento óptimo de la red.

Para establecer la calidad de servicio, se ha usado como indicadores de la misma a los parámetros P_L y ϵ , los cuales son calculados cuando el sistema se encuentra en un estado de saturación (se considera la capacidad del buffer infinita).

El cálculo de estos parámetros se basó en considerar modelos de fuentes de tráfico ON – OFF, con los siguientes descriptores de tráfico: R_p , tasa pico de transmisión de celdas, $1/\beta$ longitud promedio de una ráfaga en seg, y $1/\alpha$ tiempo promedio de estar en OFF.

Para la implementación del control de admisión de llamadas, se ha planteado una expresión que permita determinar aproximadamente la capacidad del enlace necesaria, considerando las características de tráfico y la calidad de servicio deseada. El cálculo para establecer la capacidad óptima que se necesitaría para poder aceptar un número máximo de

llamadas se basa en considerar una aproximación de la cola de la distribución binomial, es decir el hacer una aproximación de la zona de saturación, considerando que esta puede ser modelada mediante una distribución gamma.

De esta manera, en este trabajo se plantea un esquema que permite calcular el número máximo de llamadas que pueden ser aceptadas dado un conjunto de fuentes con sus propios descriptores de tráfico para una calidad de servicio dada. Este esquema puede ser considerado para fuentes como de voz, tipo ráfagas incluyendo imágenes y vídeo, entre otras.

Adicionalmente, al modelar las expresiones obtenidas tanto para el número máximo de llamadas como para la capacidad del enlace se puede observar que el criterio de control de admisión con el esquema de la aproximación gamma es más estricto que el que se obtiene con la aproximación Gaussiana. Además, para probabilidades de pérdida bastante pequeñas la aproximación Gamma tiene un mejor comportamiento que la distribución Gaussiana la cual es más flexible en el cálculo del número máximo de llamadas permitidas o en el cálculo de la capacidad.

De esta manera, se puede concluir que tanto las aproximaciones Gamma como Gaussiana representan una cota superior e inferior respectivamente, con respecto a la distribución binomial.

Para investigaciones posteriores se podría considerar este esquema de control de admisión de llamadas basado en una aproximación Gamma, pero ahora tomando en cuenta buffers de tamaño finito y políticas de calendarización diferentes a FIFO. Estas consideraciones hacen que el algoritmo de control se haga más complejo debido a las diferentes variantes que hay que considerar.

Además, también se podría considerar este esquema de control de admisión para el modelamiento de las fuentes de vídeo de tasa de bit variable que consideran los modelos de cadenas de Markov de tres estados y los modelos de auto regresión. De esta manera, el modelamiento y algoritmo de control de admisión propuesto en [9] para fuentes de vídeo podría ser modificado de tal manera que para el análisis de los procesos de auto regresión ya no se considere los esquemas de distribución Gaussiana sino que al contrario se realicen los análisis pero considerando el esquema de control de admisión con distribución Gamma.

Apéndice A

Aproximación Gaussiana

Un método heurístico para obtener la constante \mathbf{K} y por lo tanto un método para el control de admisión de llamadas para las fuentes ON – OFF es la aproximación Gaussiana. Si se considera que el parámetro de calidad de servicio es la probabilidad de pérdida de celdas P_L .

Una estimación conservativa de la Probabilidad de Pérdida es considerarla como la relación entre la tasa de entrada cuando el estado $(N - 1)$ del sistema se encuentra en la región de sobrecarga con respecto a la tasa de entrada promedio.

$$P_L = \sum_{i=J_0}^N (i - C)\pi_i / m \quad (\text{A.1})$$

siendo π_i la probabilidad de que el sistema se encuentre en el estado i . Esta es una estimación conservativa ya que asume que cualquier tasa de entrada por encima de la capacidad resultará en una pérdida de celdas.

Una medida relacionada con la probabilidad de pérdida es la probabilidad de estar en la región de sobrecarga; esto es, la probabilidad de que la tasa de entrada exceda la capacidad del enlace C , esta medida se llama ε y esta expresada como

$$\varepsilon = \sum_{i=J_0}^N \pi_i \quad (\text{A.2})$$

Para tener una probabilidad pequeña de P_L o ε , se requiere un número relativamente grande de fuentes que deben ser multiplexadas. De esta manera, se asume que $N \gg 1$ y que $p \ll 1$; esto es, se asume que las fuentes son relativamente de tipo ráfagas. Estas suposiciones permiten encontrar la constante K en una manera bastante fácil. Con N grande y p pequeño, es muy conocido que la solución de la probabilidad de estar en el estado i dado por la distribución binomial

$$\pi_i = \binom{N}{i} p^i (1-p)^{N-i} \quad (\text{A.3})$$

puede ser aproximada a la distribución Gaussiana con los mismos valores de mediana $m = Np$ y la varianza $\sigma^2 = Np(1-p)$. La suma de las ecuaciones (A.2) y (A.3) son aproximadas por las integrales de la distribución Gaussiana sobre la misma área.

De esta manera, la probabilidad de pérdida de celdas P_L puede ser aproximada por

$$P_L = \frac{1}{m} \int_{J_0}^{\infty} \frac{e^{-(x-m)^2 / 2\sigma^2} (x-C)}{\sqrt{2\pi\sigma^2}} dx \quad (\text{A.4})$$

$$P_L = \frac{1}{m} \frac{e^{-(c-m)^2 / 2\sigma^2}}{\sqrt{2\pi}} - \frac{(c-m)}{m} Q(c-m) \quad (\text{A.5})$$

la ecuación anterior puede ser simplificada a la siguiente expresión

$$P_L = \frac{\sigma^3}{\sqrt{2\pi}(c-m)^2 m} e^{-(c-m)^2 / 2\sigma^2} \quad (\text{A.6})$$

con $(c-m)/\sqrt{2}\sigma > 3$

De igual manera, la probabilidad de encontrarse en la región de sobrecarga, ϵ puede ser aproximada por

$$\epsilon = \int_{J_0}^{\infty} \frac{e^{-(x-m)^2 / 2\sigma^2}}{\sqrt{2\pi\sigma^2}} dx \quad (\text{A.7})$$

$$\epsilon = \frac{\sigma}{\sqrt{2\pi}(c-m)} e^{-(c-m)^2 / 2\sigma^2} \quad (\text{A.8})$$

con $(c-m)/\sqrt{2}\sigma > 3$

Por lo tanto, con las expresiones de las probabilidades de pérdida y de estar en la región de saturación y además dado un número N de llamadas se puede encontrar la capacidad C , mediante la siguiente ecuación

$$C = m + \sigma \sqrt{-\ln(2\pi) - 2 \ln \epsilon} \quad (\text{A.9})$$

$$K = \sqrt{-\ln(2\pi) - 2 \ln \epsilon} \quad (\text{A.10})$$

Con este valor de K se puede obtener una expresión para N , donde N es el número de conexiones ON – OFF permitidas, dado un enlace con capacidad C_L y con parámetro de calidad de servicio ϵ .

$$C = Np + K \sqrt{Np(1-p)} \quad (\text{A.11})$$

esta expresión puede ser escrita como

$$Np + K \sqrt{N} \sqrt{p(1-p)} - C = 0 \quad (\text{A.12})$$

resolviendo esta ecuación se puede obtener una expresión para el número de conexiones permitidas

$$N = \frac{C}{p} + \frac{K^2(1-p)}{2p} - \frac{1}{2p} \sqrt{K^2(1-p)[K^2(1-p) + 4C]} \quad (\text{A.13})$$

Esta expresión representa el control de admisión de llamadas para fuentes homogéneas ON – OFF.

La Ganancia multiplexada se obtiene cuando se permite que el número de fuentes N se incremente más allá de C , incurriendo por lo tanto en una pérdida de celdas, $G_e = N/C$.

La Ganancia G_e máxima ocurre cuando $m = C$ (límite inferior en la capacidad), esto es cuando el control de admisión de llamadas se basa en la asignación del ancho de banda promedio, esto es, cuando el número de llamadas permitidas está determinado para $mR_p = C_L$ o $m = C$.

Para fuentes ON – OFF, que presentan una aproximación binomial $m = Np$, se tiene que

$$G_e \max = \frac{N}{m} = \frac{N}{Np} = \frac{1}{p} \quad (\text{A.14})$$

La ganancia actual está dada por,

$$G_e = \frac{N}{C} = \frac{1}{C} \left\{ \frac{C}{p} + \frac{2\alpha}{p} - \frac{1}{p} \sqrt{4\alpha|\alpha + C|} \right\} \quad (\text{A.15})$$

donde

$$\alpha = \frac{K^2(1-p)}{4}$$

Apéndice B

Variables Aleatorias

B.1 Distribución Gaussiana

La función de densidad de probabilidad de una variable aleatoria Gaussiana X está dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} \quad (\text{B.1})$$

donde m y $\sigma > 0$ son números reales y representan la media y la desviación standard de X . La figura B.1 muestra que la función de densidad de probabilidad es una curva en forma de campana, centrada y simétrica sobre m y cuyo ancho se incrementa con σ , [11].

La función de distribución acumulada de una variable aleatoria Gaussiana está dada por

$$P[X \leq x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(x'-m)^2/2\sigma^2} dx' \quad (\text{B.2})$$

El cambio de variable $t = (x' - m)/\sigma$ resulta en

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-m)/\sigma} e^{-t^2/2} dt = \Phi\left(\frac{x-m}{\sigma}\right) \quad (\text{B.3})$$

donde $\Phi(x)$ es la función de distribución acumulada de una variable aleatoria Gaussiana con $m = 0$ y $\sigma = 1$.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (\text{B.4})$$

Por lo tanto, cualquier probabilidad que involucra una variable aleatoria Gaussiana puede ser expresada en términos de $\Phi(x)$.

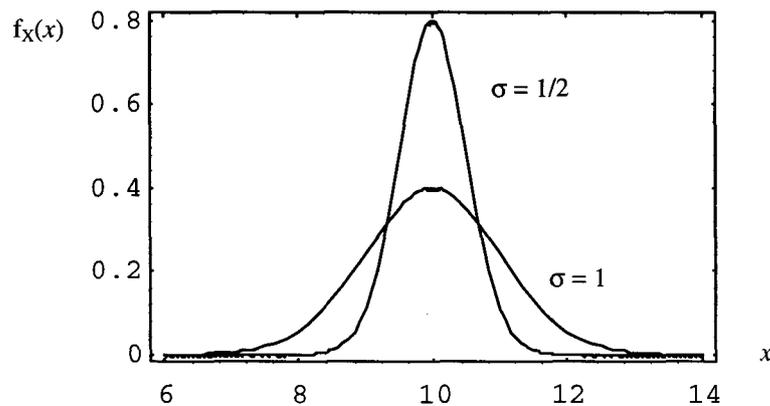


FIGURA B.1: Función de densidad de probabilidad de la variable aleatoria Gaussiana

B.2 Distribución Gamma

La función de densidad de probabilidad de la variable aleatoria gamma tiene dos parámetros, $\alpha > 0$ y $\lambda > 0$, y esta dada por:

$$f_X(x) = e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, \quad 0 \leq x < \infty \quad (\text{B.5})$$

donde $\Gamma(z)$ es la función gamma, la cual esta definida por la integral

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx, \quad z > 0 \quad (\text{B.6})$$

La media y la varianza de la distribución gamma está dada por

$$E[x] = \frac{1}{\lambda} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\alpha}{\lambda} \quad (\text{B.7})$$

$$\text{Var}[x] = \frac{1}{\lambda^2} \left[\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} - \left\{ \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \right\}^2 \right] = \frac{\alpha}{\lambda^2} \quad (\text{B.8})$$

La función gamma tiene las siguientes propiedades:

- $\Gamma(1/2) = \sqrt{\pi}$
- $\Gamma(z+1) = z\Gamma(z)$ para $z > 0$, y
- $\Gamma(m+1) = m!$ para m entera no negativa

Las funciones gamma incompletas se definen como, [12],

$$\gamma(z, x) = \int_0^x t^{z-1} e^{-t} dt \quad \Gamma(z, x) = \int_x^{\infty} t^{z-1} e^{-t} dt \quad (\text{B.9})$$

La versatilidad de la variable aleatoria gamma es debido a la riqueza de la función gamma $\Gamma(z)$. La función de densidad de probabilidad de la variable aleatoria gamma puede tomar una variedad de formas como se puede observar en la Figura B.2. Variando los parámetros α y λ es posible hacer coincidir la función de densidad de probabilidad con muchos tipos de datos experimentales. La variable aleatoria exponencial se obtiene con $\alpha = 1$. Al dar valores a $\lambda = 1/2$ y $\alpha = k/2$, donde k es un entero positivo, se obtiene la variable aleatoria chi-cuadrada. La variable aleatoria m-Erlang se obtiene cuando $\alpha = m$, un entero positivo. La variable aleatoria m-Erlang es usada en los modelos de disponibilidad del sistema y en los de colas.

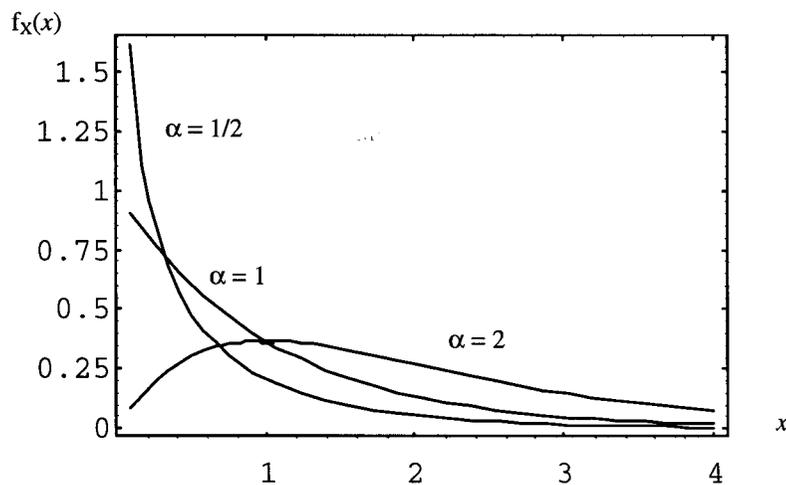


Figura B.2: Función de densidad de probabilidad de la variable aleatoria Gamma

Dada una distribución Poisson, con una tasa de cambio λ , la función de distribución acumulada $F_X(x)$ dado los tiempos de espera hasta el cambio h es

$$\begin{aligned}
 F_X(x) &= P(X \leq x) = 1 - P(X > x) \\
 &= 1 - \sum_{k=0}^{h-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!} = 1 - e^{-\lambda x} \sum_{k=0}^{h-1} \frac{(\lambda x)^k}{k!}
 \end{aligned} \tag{B.10}$$

para $x \geq 0$. La función de probabilidad $f_X(x)$ se obtiene diferenciando $F_X(x)$,

$$\begin{aligned}
 f_X(x) &= F_X'(x) \\
 &= \lambda e^{-\lambda x} \sum_{k=0}^{h-1} \frac{(\lambda x)^k}{k!} - e^{-\lambda x} \sum_{k=0}^{h-1} \frac{k(\lambda x)^{k-1} \lambda}{k!}
 \end{aligned}$$

$$\begin{aligned}
&= \lambda e^{-\lambda x} + \lambda e^{-\lambda x} \sum_{k=0}^{h-1} \frac{(\lambda x)^k}{k!} - e^{-\lambda x} \sum_{k=1}^{h-1} \frac{k(\lambda x)^{k-1} \lambda}{k!} \\
&= \lambda e^{-\lambda x} - \lambda e^{-\lambda x} \sum_{k=0}^{h-1} \left[\frac{k(\lambda x)^{k-1} \lambda}{k!} - \frac{(\lambda x)^k}{k!} \right] \\
&= \lambda e^{-\lambda x} \left\{ 1 - \left[1 - \frac{(\lambda x)^{h-1}}{(h-1)!} \right] \right\} = \frac{\lambda (\lambda x)^{h-1}}{(h-1)!} e^{-\lambda x} \tag{B.11}
\end{aligned}$$

ahora si se considera que $\alpha \equiv h$ y se define a λ como el tiempo entre los cambios, la ecuación B.11 puede ser escrita como

$$P(x) = \begin{cases} \frac{\lambda (\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases} \tag{B.12}$$

Una aproximación muy útil de la función gamma esta dada por

$$\Gamma(\alpha + 1) \approx \sqrt{2\pi\alpha} (\alpha/e)^\alpha \quad \text{para } \alpha \gg 1 \tag{B.13}$$

la cual es conocida como la Fórmula de Stirling. El error absoluto de esta aproximación tiende al infinito a medida que $\alpha \rightarrow \infty$, pero su error relativo tiende a cero; esto es ,

$$\frac{\sqrt{2\pi\alpha} (\alpha/e)^\alpha}{\Gamma(\alpha + 1)} \approx 1, \quad \alpha \gg 1 \tag{B.14}$$

De la ecuación anterior se tiene

$$\frac{\sqrt{2\pi n} (n/e)^n}{n!} \approx 1, \quad n \gg 1 \tag{B.15}$$

La aproximación $\sqrt{2\pi n} (n/e)^n$ a $n!$ es también conocida como la fórmula de Stirling. Los errores relativos correspondientes son pequeños incluso para valores moderados de n .

En resumen, la distribución gamma con un parámetro de forma α y un parámetro de escala λ tiene una media α/λ , varianza α/λ^2 , y una desviación standard de $\sqrt{\alpha}/\lambda$. Específicamente, la distribución exponencial con parámetro de escala λ , tiene media λ , varianza λ^2 , y desviación standard λ .

Sea $f_X(x)$ la función de densidad de probabilidad de la distribución gamma con un parámetro de forma α y un parámetro de escala λ . Supóngase que $0 < \alpha \leq 1$. De esta manera, $f_X(x)$ es una función decreciente en $[0, \infty)$. Si $0 < \alpha < 1$, entonces $f_X(x)$ se aproxima a ∞ a medida que x se aproxima a 0 por su derecha; si $\alpha = 1$, entonces $f_X(x)$ se aproxima a λ así como x se aproxima a 0 por su derecha. En cambio cuando $\alpha > 1$, entonces $f_X(x)$ es continua en su origen y tiene un valor de cero en este. Como una función en $[0, \infty)$ esta crece a su máximo en su modo único, el cual es igual a $(\alpha - 1)/\lambda$, y luego disminuye a 0, [13].

B.3 Distribución Binomial

Consideremos un experimento en el cual sus salidas pueden ser de dos tipos, es decir el tener un éxito o un fracaso. Sea p la probabilidad de éxito, donde $0 < p < 1$. Entonces $1 - p$ es la probabilidad de fracaso. Sea i el evento que indique la ocurrencia de la variable aleatoria, la cual tiene una distribución de Bernoulli con parámetro p . De esta manera i tiene la función de probabilidad f dada por $f(0) = 1 - p$ y $f(1) = p$, y tiene una media p , varianza $p(1 - p)$ y una desviación standard $\sqrt{p(1 - p)}$.

Sea X el número de veces que un evento A ocurre en n intentos. X es la variable aleatoria con un rango $S_X = \{0, 1, \dots, n\}$. Si se considera I_j sea la función indicadora para el evento A en el j intento, entonces

$$X = I_1 + I_2 + \dots + I_n \quad (\text{B.16})$$

Esto es, X es la suma de las variables aleatorias de Bernoulli asociadas con cada uno de los n eventos independientes. De esta manera, la función de probabilidad de masa de X es la siguiente:

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{para } k = 0, \dots, n \quad (\text{B.17})$$

X se conoce como variable aleatoria binomial. La Figura B.3 muestra la función de densidad de probabilidad de X para $n = 24$, $p = 0.2$ y $p = 0.5$. Hay que tener en cuenta, que $P[X = k]$ es máximo en $k_{\max} = [(n+1)p]$, donde $[x]$ representa el entero más grande que es más pequeño o igual a x . Cuando $(n+1)p$ es un entero, entonces el máximo se logra en k_{\max} y $k_{\max} - 1$.

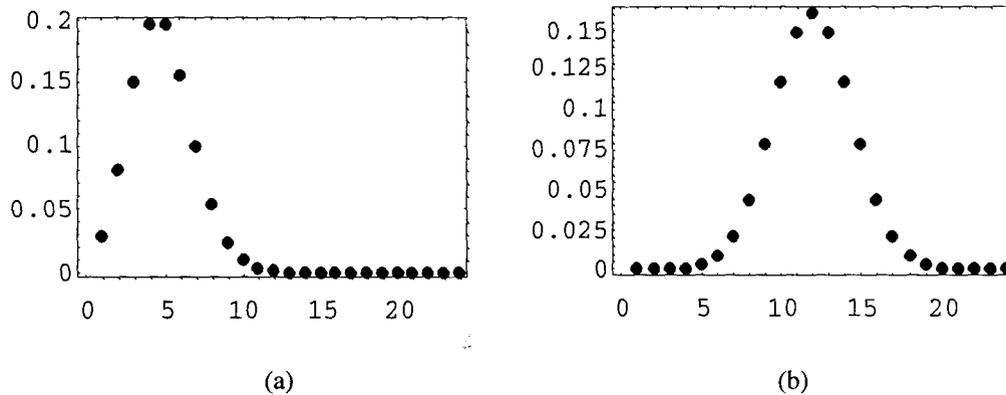


Figura B.3: Función de densidad de probabilidad de la variable aleatoria binomial. (a) $p=0.2$; (b) $p=0.5$

La variable aleatoria binomial se presenta en aplicaciones donde existen dos tipos de objetos (por ejemplo: bits correctos/erróneos, activos/silencio) y se está interesado en el número de objetos del tipo 1 en un conjunto de tamaño n y seleccionado aleatoriamente, donde el tipo de cada objeto es independiente de los tipos de los otros objetos en el lote.

Bibliografía

- [1] Tutomu Murase, Hiroshi Suzuki, “*A Call Admission Control Scheme for ATM Networks Using a Simple Quality Estimate*”, IEEE Journal on Selected areas in Communications, Vol. 9, pp 1461 – 1470, 1991.
 - [2] Maurizio Decina, Tiziana Toniatti, “*On Bandwidth Allocation to Bursty Virtual Connections in ATM Networks*”, IEEE Network, pp 844 – 851, 1990.
 - [3] Mischa Schwartz, “*Broadband Integrated Networks*”, Prentice Hall, 1996.
 - [4] David E. McDysan, Darren L. Spohn, “*ATM Theory and Application*”, McGraw – Hill, Inc., 1994.
 - [5] Thomas M. Chen, Stephen S. Liu, “*ATM Switching Systems*”, Artech House, 1995.
 - [6] Erwin Rathgeb, “*Modeling and Performance Comparison of Policing mechanisms for ATM Networks*”, IEEE journal on selected areas in Communications, Vol. 9, pp 325 – 334, 1991.
-

-
- [7] Yasuhiro Miyao, "A call admission control scheme in ATM Networks", ICC '91, pp 391 – 396, 1991.
- [8] Abhijit S. Pandya, Ercan Sen, " *ATM Technology for Broadband Telecommunications Networks*", CRC Press, 1999.
- [9] Hiroshi Esaki, "Call Admission Control Method in ATM Networks", ICC, pp 1628 – 1633, 1992.
- [10] Cheul Shim, Intae Ryoo, "Modeling and Call Admission Control Algorithm of Variable Bit Rate Video in ATM Networks", IEEE Journal on selected areas in Communications, Vol. 12, pp 332 – 344.
- [11] Alberto Leon Garcia, "Probability and Random Processes for Electrical Engineering", Segunda Edición, Addison – Wesley, 1994.
- [12] T. Strauss, S. Oren, "Probability in the Engineering and Informational Sciences", Cambridge University, 1994.
- [13] Charles J. Stone, "A Course in Probability and Statistics", Duxbury Press, 1996.
-

Centro de Información-Biblioteca



30002005826441