

EVALUACIÓN DEL DESEMPEÑO DE LOS MODELOS DE TRÁFICO BASADOS EN
PROCESOS DE MARKOV Y MOVIMIENTO BROWNIANO EN EL NIVEL DE TRANSITO
NGN PARA VoIP



ROBINSON EDWIN AGREDO FAJARDO
RONY ALEJANDRO ARANA TAQUINAS

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
POPAYÁN
2008

EVALUACIÓN DEL DESEMPEÑO DE LOS MODELOS DE TRÁFICO BASADOS EN
PROCESOS DE MARKOV Y MOVIMIENTO BROWNIANO EN EL NIVEL DE TRANSITO
NGN PARA VoIP

ROBINSON EDWIN AGREDO FAJARDO
RONY ALEJANDRO ARANA TAQUINAS

Monografía para optar al título de
INGENIERO EN ELECTRÓNICA Y TELECOMUNICACIONES.

Director
EVELIO ASTAIZA HOYOS
Ingeniero en Electrónica y Telecomunicaciones.

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
POPAYÁN
2008

*A mis padres,
a mis hermanos,
a mis queridos sobrinos Sebastián y Andrés
por su paciencia, animo y apoyo incondicional.*

Rony Alejandro Arana Taquinas.

*A Dios por ser el dador de la vida.
A mis padres por su apoyo en todo momento.
A Arlen, Fabio, Sonia por su animo y ayuda constante.
Y todas las personas que con sus consejos me ayudaron y
han creído en mí.*

Robinson Edwin Agredo Fajardo

AGRADECIMIENTOS

Los autores expresan sus agradecimientos a todos los que de una u otra manera hicieron posible la realización de este trabajo de grado. Primero que todo, a Dios por ser la guía en el transcurso de nuestras vidas. Al director del proyecto, ingeniero Evelio Astaiza Hoyos, por su paciencia, apoyo y compromiso para la elaboración de este trabajo. A todos los profesores que estuvieron presentes durante nuestros estudios de pregrado. A nuestras familias por su apoyo incondicional, y a las demás personas que con sus aportes colaboraron en el cumplimiento de este objetivo.

A Todos Muchas Gracias...!!!

CONTENIDO

	pág.
1 INTRODUCCIÓN	11
1.1 MODELOS DE TRÁFICO Y SUS APLICACIONES	11
1.2 MOTIVACIÓN DE LA TESIS	13
1.3 OBJETIVOS	15
1.4 ESTRUCTURA DEL TRABAJO DE GRADO	16
1.5 PRINCIPALES CONTRIBUCIONES	17
2 MARCO TEÓRICO	19
2.1 INTRODUCCIÓN	19
2.2 CONCEPTOS DE VoIP	20
2.2.1 <i>Introducción</i>	20
2.2.2 <i>Componentes VoIP</i>	21
2.2.3 <i>Codecs</i>	22
2.2.4 <i>Calidad de servicio en el marco de voz en redes NGN</i>	23
2.2.5 <i>Problemas técnicos relacionados con las redes de paquetes</i>	25
2.3 TELETRÁFICO Y TEORÍA DE COLAS	26
2.3.1 <i>Teletráfico</i>	26
2.3.2 <i>Conceptos de tráfico.</i>	27
2.3.3 <i>Teoría de colas</i>	29
2.3.3.1 <i>Definición</i>	29
2.3.3.2 <i>Conceptos básicos de los sistemas de colas</i>	30
2.4 PROCESOS DE MARKOV.	39
2.4.1 <i>Cadenas de Markov.</i>	39
2.4.2 <i>Procesos de Poisson</i>	44
2.5 MOVIMIENTO BROWNIANO	45
2.6 CARACTERÍSTICAS DEL TRÁFICO EN REDES	46

2.6.1	<i>Fractalidad</i>	47
2.6.2	<i>Fractalidad en las Redes de Telecomunicaciones</i>	50
2.6.3	<i>Autosimilitud</i>	51
2.6.4	<i>Dependencia de Largo Rango</i>	59
2.6.5	<i>Distribución de Cola Pesada (Heavy-Tailed)</i>	61
2.6.6	<i>Autosimilitud en el Tráfico de Redes de Telecomunicaciones</i>	62
2.6.7	<i>Teorema Límite Para tráfico Agregado</i>	63
2.7	MODELOS DE TRÁFICO	64
2.7.1	<i>Procesos modulados de Markov (MMPP)</i>	65
2.7.2	<i>Movimiento Browniano Fraccional</i>	68
2.7.2.1	<i>Estacionaridad y Ergodicidad</i>	68
2.7.2.2	<i>Modelos estocásticos auto-similares</i>	70
3	MODELAMIENTO E IMPLEMENTACIÓN DE LA SIMULACIÓN	74
3.1	INTRODUCCIÓN	74
3.2	MOVIMIENTO BROWNIANO FRACCIONAL	75
3.2.1	<i>Método Cholesky (Exacto)</i>	75
3.2.1	<i>Método de Transformada de Fourier (Aproximado)</i>	76
3.3	PROCESOS DE POISSON MODULADOS POR MARKOV (MMPP).	78
4	PRUEBAS Y RESULTADOS	87
5	CONCLUSIONES	109
	GLOSARIO	113
	REFERENCIAS	115

LISTA DE TABLAS

	pág.
Tabla 1. Estándares de codificación	23
Tabla 2. Problemas típicos en teoría de colas.	29
Tabla 3. Parámetros de escala y de forma, arrojados para la distribución de Pareto	96
Tabla 4. Valor del parámetro de Hurst frente al número de muestras.	98
Tabla 5. Valor estimado del parámetro Hurst por medio del Periodograma	99
Tabla 6. Muestras generadas con el método de Cholesky y estimación de H	101
Tabla 7. Muestras generadas con el método de Paxson y estimación de H	103
Tabla 8. Muestras generadas con el método MMPP y la estimación de H.	106

LISTA DE FIGURAS

	pág.
Figura 1. Componentes VoIP [7].	21
Figura 2. Encapsulación de VoIP en una trama PPP y Ethernet.	22
Figura 3. Sistema de colas	30
Figura 4. Conjunto de Mandelbrot [36].	48
Figura 5. Dimensión fractal	50
Figura 6. Ilustración gráfica de la autosimilitud [2].	53
Figura 7. Diagrama de estados de un 2-MMPP [18].	68
Figura 8. Diagrama de estados de un 2-MMPP [28].	79
Figura 9. Representación gráfica de los arribos en un MMPP de 2 estados [16].	79
Figura 10. Método de la transformada inversa.	81
Figura 11. Pdf de una distribución exponencial e histogramas de una variable aleatoria.	82
Figura 12. Diagrama en Bloques: Generar los tiempos entre llegadas.	85

Figura 13. Diagrama en Bloques: Generar número de arribos.	86
Figura 14. Función de distribución para diferentes tamaños de muestras.	89
Figura 15. Ampliación de figura 14.	90
Figura 16. Las distribuciones: Exponencial, Gumbel, Weibull, y Pareto frente a la distribución acumulativa Empírica.	91
Figura 17. Ampliación de un sector de la figura 16.	92
Figura 18. Las distribuciones: Exponencial, Gumbel, Weibull, y Pareto frente a la distribución acumulativa Empírica para 10.000 muestras	93
Figura 19. Ampliación de un sector de la figura 18.	94
Figura 20. Las distribuciones: Exponencial, Gumbel, Weibull, y Pareto frente a la distribución acumulativa Empírica para 30.000 muestras.	94
Figura 21. Ampliación de un sector de la figura 20	95
Figura 22. Quantiles de voz frente a Quantiles de Pareto	97
Figura 23. Gráfica Varianza-Tiempo	99
Figura 24. Gráfica de Periodograma de voz IP	100
Figura 25. Gráfico generado por el método de Cholesky	101
Figura 26. Método de Varianza tiempo para las muestras generadas con el método de Cholesky	102
Figura 27. Método de Periodograma para las muestras generadas con el método de Cholesky	102
Figura 28. Gráfico generado por el método de Paxson	103
Figura 29. Método de Varianza tiempo para las muestras generadas con el método de Paxson.	104
Figura 30. Método de Periodograma para las muestras generadas con el método de Paxson.	104
Figura 31. Gráfica de 14.000 muestras de tráfico real.	106
Figura 32. Muestras sintéticas generadas con el modelo mmpp.	107
Figura 33. Método de Varianza tiempo para las 36579 muestras generadas con el método MMPP	107
Figura 34. Método de Periodograma para las 36579 muestras generadas con el método MMPP	108

LISTA DE ANEXOS

ANEXO A: Definiciones estadísticas.

ANEXO B: Simulador NS-2.

ANEXO C: Calidad de servicio en redes IP.

ANEXO D: Métodos de estimación del parámetro de Autosimilitud.

1 INTRODUCCIÓN

En el entorno de las redes de telecomunicaciones actuales, los modelos de tráfico cobran gran importancia debido a las ventajas que proporcionan en la gestión y dimensionamiento, estos pueden ser utilizados, por ejemplo, en la planificación de la capacidad de los enlaces, en la escogencia del tamaño de los buffers teniendo en cuenta los efectos de la multiplexación estadística y también en el análisis del desempeño. En la Internet, la complejidad asociada a los mecanismos de generación y control del tráfico, así como la diversidad de aplicaciones, han ocasionado la aparición de un conjunto distintivo de características tales como: autosimilitud, dependencia de largo rango y multifractalidad, las cuales hacen que el estudio del tráfico en las redes de comunicaciones sea mas complejo, razón por la cual requiere que sean debidamente modeladas.

1.1 MODELOS DE TRÁFICO Y SUS APLICACIONES

Los modelos de tráfico basados en parámetros entre los cuales se pueden citar: MMPP (Procesos de Poisson Modulados por Markov), FBM (Movimiento Fraccional Browniano), FARIMA (Fractional Autoregressive Integrated Moving Average) etc. permiten una descripción detallada del tráfico. Un buen modelo de tráfico debe, por un lado, permitir en su estructura matemática representar las características de tráfico con mayor impacto en el desempeño de las redes, y por otro, poseer un procedimiento de inferencia de parámetros que consiga capturar de forma eficiente las mismas. En general, estas características incluyen: (i) las estadísticas de primer y segundo orden, (ii) las propiedades de similaridad escalar y (iii) una correlación entre los procesos de llegada y

los tamaños de paquetes.

Los modelos de tráfico son una herramienta precisa en la concepción, gestión y en el dimensionamiento de una red de telecomunicaciones. Las ventajas que resultan del uso eficiente de los mismos son muy variadas, ejemplo de ello es que permiten hacer una previsión relativamente rigurosa del crecimiento y comportamiento del tráfico a mediano y largo plazo. Un operador de red puede planear en el tiempo el crecimiento de una red de forma organizada, garantizando en cada momento una utilización eficiente de los recursos. Esta posibilidad de crecimiento distribuido, permite garantizar, por un lado, la satisfacción de los clientes a nivel de calidad de servicio y por otro, disminuir en un buen porcentaje las inversiones a largo plazo. Esta ventaja económica para el operador puede resultar en un menor costo de servicio para los clientes. En este contexto los modelos de tráfico pueden también ser utilizados para definir políticas de tarificación que, por ejemplo, promuevan la utilización de las redes en los periodos con menor congestión.

Un operador para administrar de forma más eficiente los recursos de red puede hacer uso de los modelos de tráfico, estos permiten llevar a cabo estudios de desempeño de la red, con base en la simulación de eventos discretos con el fin de optimizar las diferentes funciones de control de la red (tales como el enrutamiento, el control de admisiones y control de flujos) lo cual ayuda al mejoramiento del servicio que se le presta a los clientes. Otra posible utilización de los modelos de tráfico en este contexto es la previsión de tráfico generado por grupos de clientes con un determinado perfil. Esta característica ayuda a la escogencia de un punto ideal de acceso a una red para un nuevo cliente con un perfil de utilización específico, también puede servir para agrupar, en un mismo punto de acceso a la red, a los clientes con solicitudes de grandes ancho de banda y en otros clientes cuyas

demandas de ancho de banda no sean tan altas, de esta forma es posible minimizar el número de segmentos de acceso a la red congestionados.

La utilización de los modelos de tráfico en la gestión o administración de las redes da como resultado un aumento en la eficiencia y eficacia en el servicio que se les ofrece a los clientes. Una previsión del tráfico generado por un grupo de clientes con un determinado comportamiento puede ser hecha, recogiendo muestras de tráfico que permitan el cálculo de parámetros que correspondan con los modelos previamente inferidos. Sería posible de esta manera, a partir del perfil de utilización de la red, deducir la cantidad de tráfico total generado por un cliente (o grupo de clientes).

Las capacidades de previsión de los modelos de tráfico a corto plazo pueden ser explotadas en el control de la red, concretamente en algoritmos de control de admisión, en métodos de asignación de recurso y en estrategias de enrutamiento adaptativas.

En resumen los modelos de tráfico pueden ser usados para estudiar el desempeño de una red en operación o para efectuar la prevención de tráfico para efectos de control y de planeamiento.

1.2 MOTIVACIÓN DE LA TESIS

Recientes estudios [1] [2] [3] realizados sobre el tráfico en redes reales, en las cuales la tendencia es integrar todo tipo de servicios en una única infraestructura de red basada en IP (*Internet Protocol*), muestran propiedades de similaridad escalar sobre un amplio rango de escalas, con un impacto significativo en la calidad de servicio ofrecido por la red, dichas propiedades pusieron a prueba las metodologías de gestión y dimensionamiento de tráfico utilizadas. Son ejemplos de comportamientos de tráfico que tienen en común

una propiedad de similaridad escalar la autosimilitud, dependencia de largo tiempo y multifractalidad. Estos comportamientos son inducidos por diversos factores: (i) un gran número de aplicaciones generando tráfico en ráfagas; (ii) algunos mecanismos de control en la red, como por ejemplo el mecanismo de control de congestión usado en TCP (Protocolo de Control de Flujo), introduce correlación entre las llegadas de paquetes, (iii) los archivos descargados por los usuarios poseen, en algunos casos, tamaños cuya distribución es de cola grande y (iv) la propia interacción humano-computador (por ejemplo, el intervalo entre click's, periodo de utilización y número de aplicaciones activas simultáneamente). Debido a la influencia que estos comportamientos del tráfico pueden tener en el desempeño de la red se torna necesario desarrollar modelos adecuados para su descripción.

La evidencia de Dependencia de Largo Rango (Long Range Dependence - LRD) y de las propiedades de autosimilitud en el tráfico de paquetes condujo a muchos investigadores a abandonar la usual presunción Markoviana a favor de modelos de tráfico más nuevos y más complejos.

En medio de los numerosos modelos genéricos LRD propuestos en la literatura, el Movimiento Browniano Fraccional (Fractional Brownian Motion - FBM) recibió mucha atención, gracias a que su naturaleza Gaussiana ayuda en el estudio del comportamiento de las colas. Sin embargo, en recientes estudios [4], [5], [6] fue expuesto que modelos simples donde la correlación es limitada, como por ejemplo los procesos de Poisson modulados por Markov (Markov Modulated Poisson Process - MMPP), pueden ser empleados exitosamente en las redes de paquetes.

De acuerdo a lo anterior, el presente trabajo de grado plantea estudiar estas características del tráfico en este tipo de ambientes, particularmente en el núcleo de la NGN (Red de Nueva Generación), con miras a evaluar el desempeño de dos modelos de tráfico particulares como lo son el modelo basado en Procesos de Markov y el modelo basado en el Movimiento Browniano, que permitan a los diseñadores de redes realizar un análisis de tráfico y dimensionamiento adecuados a este tipo de ambientes.

En este trabajo de grado son abordados dos modelos: uno el MMPP debido a su importancia entre los modelos clásicos y el otro desarrollado recientemente el Movimiento Browniano Fraccional.

1.3 OBJETIVOS

La continua evolución de las redes de comunicaciones y la aparición de nuevos servicios o aplicaciones en las mismas, ha ocasionado un incremento en los flujos y tipos de tráfico circulantes por la red, en consecuencia, fenómenos como la autosimilitud y la dependencia de largo rango se tienen en cuenta para explicar el comportamiento del tráfico en las redes actuales. No obstante, es importante el estudio de los modelos de tráfico que dentro de su concepción tengan en cuenta estos fenómenos.

El objetivo de este trabajo de grado es evaluar el desempeño de los modelos de tráfico basados en Procesos de Markov y movimiento Browniano en el nivel de tránsito en redes NGN para VoIP.

Ligado a este objetivo general se encuentran los siguientes objetivos específicos:

Caracterizar el tráfico proyectado en las redes de próxima generación, para usuarios de Voz sobre IP.

Simular el comportamiento del tráfico de VoIP en las redes de próxima generación de acuerdo a los modelos de tráfico basados en Procesos de Markov y Movimiento Browniano.

Analizar los resultados de la simulación y establecer el modelo que más se adapta para el tráfico en las redes de próxima generación de acuerdo a la caracterización realizada.

1.4 ESTRUCTURA DEL TRABAJO DE GRADO

Este trabajo de grado describe y analiza dos de los modelos de tráfico usados para capturar las características del tráfico en las redes. El trabajo se encuentra organizado en cinco capítulos.

Capítulo 1. Introducción: Consigna la motivación del trabajo de grado, describe el entorno donde se aplica, los objetivos planteados y la estructura de la monografía.

Capítulo 2. Marco Teórico: Describe brevemente los aspectos esenciales del estado del arte de la temática tratada en el trabajo de grado, fundamentos de teoría de colas, conceptos básicos de tráfico, conceptos de VoIP y en especial de los dos modelos de tráfico. Para el modelo basado en procesos de Markov se presentará la teoría relacionada con las cadenas de Markov, los procesos de Poisson y los procesos de Poisson modulados por Markov. Para el modelo Browniano se tratara la matemática relacionada con el movimiento Browniano Fraccional. Además se presentaran las características del tráfico que los modelos deben caracterizar o capturar.

Capítulo 3 .Modelamiento e Implementación de la Simulación: En este se presentan los aspectos más relevantes de la implementación del modelo de simulación para cada uno de los modelos de tráfico.

Capítulo 4. Pruebas y Resultados: En este se realiza el análisis de las trazas reales de voz sobre IP y de las muestras generadas sintéticamente mediante modelos MMPP y Movimiento Browniano Fraccional, los resultados de este análisis también se encuentran dentro de esta capítulo.

Capítulo 5. Conclusiones: Contiene la síntesis de los resultados obtenidos a partir del estudio del comportamiento del tráfico de voz sobre IP y su aproximación por medio MMPP y FBM, aquí también se habla de la experiencia adquirida y las recomendaciones para trabajos futuros.

1.5 PRINCIPALES CONTRIBUCIONES

- Análisis de un modelo de tráfico basado en los procesos de Poisson modulados por Markov y evaluación de su capacidad para capturar de forma precisa las estadísticas de primer y segundo orden, los cuales conducen a la presencia de autosimilitud y dependencia de largo rango en el proceso de llegada de paquetes en las redes actuales, el número de parámetros es el equivalente a un proceso de dos estados, con el cual se hace la adaptación a las características particulares del tráfico de VoIP.
- Análisis de un modelo de tráfico basado en movimiento Browniano, más específicamente el movimiento fraccional browniano, el cual presenta condiciones de

autosimilitud y dependencia de largo rango. Estos conceptos se consideran hoy de vital importancia para comprender la naturaleza dinámica del tráfico y de este modo lograr un mayor control del tráfico cursado. Al ser este modelo uno de los más utilizados es de gran importancia realizar el análisis del tráfico de voz por medio de este modelo.

2 MARCO TEÓRICO

2.1 INTRODUCCIÓN

Actualmente la Internet es una red heterogénea donde un número creciente de usuarios utiliza un gran número de aplicaciones, muchas de las cuales generan tráfico en ráfagas. La interacción de los individuos con la red acarrea unas propiedades estadísticas de grandes colas, por ejemplo, el intervalo entre click's, el periodo de utilización, el número de aplicaciones activas simultáneamente y el tamaño de los archivos descargados. Estas características, junto con algunos mecanismos de red, como por ejemplo el control de congestión usado por TCP, influyen en el comportamiento del tráfico, debido a esto la similaridad escalar o la invariancia de estadística en una escala temporal hace presencia.

Los comportamientos de similaridad escalar tienen un impacto significativo en la calidad de servicio ofrecida por la red y trae nuevos desafíos para modelar el tráfico de Internet. Una consecuencia es la incorporación de estas características en los modelos de tráfico, donde el movimiento fraccional Browniano es un claro ejemplo de ello. En lo referente a los modelos clásicos como por ejemplo los procesos de Poisson modulados por Markov, estos pueden dar una buena aproximación al comportamiento de similaridad escalar, la transición entre los estados no corresponde a la generación de paquetes, pero permite los procesos de llegadas, los cuales son descritos como una superposición de distribuciones exponenciales (procesos de Poisson), lo cual se asemeja a las características de LRD. En general, consiguen capturar las características estadísticas del tráfico el cual se asumía markoviano.

En este capítulo se presenta la base teórica fundamental para el desarrollo de este trabajo de grado. En la primera parte se presentan algunos conceptos básicos de VoIP, teoría de colas y algunos de teletráfico; en la segunda parte se habla de las características del tráfico a tener en cuenta en el análisis de los dos modelos, así como también la descripción matemática de los mismos.

2.2 CONCEPTOS DE VoIP

2.2.1 *Introducción*

Inicialmente las redes de datos no fueron pensadas para el transporte de aplicaciones que necesitan respuesta en tiempo real. En los últimos años la demanda que se tiene de estos servicios por parte de las empresas y clientes puntuales es bastante grande, lo cual sumado a la evolución que han tenido las redes IP, hicieron que los ISPs (Internet Service Providers) e ITSPs (Internet Telephony Service Providers) vieran viable el transporte de la voz sobre estas redes.

Para conseguir este objetivo el fortalecimiento y la consolidación de una única infraestructura de red en donde se utilicen equipos comunes para la transferencia de voz, video y datos se hace necesario, con esto, también se logra reducir los costos relacionados con el mantenimiento, operación y gestión de la red.

Con la convergencia de servicios, se pueden desarrollar e implementar servicios de valor agregado a través del acople de aplicaciones multimedia. Ejemplo: mensajes unificados, centrales telefónicas Web etc. Sin embargo, el diseño de una red de VoIP requiere una planificación cuidadosa para asegurar que la calidad de la voz puede ser debidamente conservada [7].

2.2.2 Componentes VoIP

Generalmente las conexiones en una red de voz IP que se pueden dar son: de teléfono a teléfono, de computador a computador, de teléfono a computador. En la figura 1 se muestran los principales componentes de una red de VoIP. A grandes rasgos la función que desempeña cada uno es la siguiente: la pasarela (gateway) permite el establecimiento de una llamada a través de una red IP convirtiendo las señales desde la interfaz del teléfono tradicional (POTS, T1/E1, ISDN) en paquetes de datos. Un teléfono IP es un dispositivo que se conecta directamente a la red, incluye codecs para la digitalización, codificación y decodificación de las conversaciones, también se encarga de la formación de paquetes de la conversación codificada. El servidor provee funciones de administración, gestión, registro y coordinación de señalización para la asignación de rutas a las llamadas. En sistemas basados en H.323, el servidor es conocido como un *gatekeeper* quien realiza control de admisión y funciones de traducción de direcciones. En SIP/SDP, el servidor es un *servidor SIP* que puede ser de diferentes tipos entre los cuales se encuentran: *SIP Proxy*, *Redirección*, *Registro* etc. En un sistema basado en MGCP o MEGACO, el servidor es un *CALL AGENT* quien maneja todo lo relacionado con la lógica de la conmutación y el control de las llamadas. Finalmente, la red IP es la encargada de proveer la conectividad entre todos los terminales. La red IP puede ser una red privada, una Intranet o Internet [7].

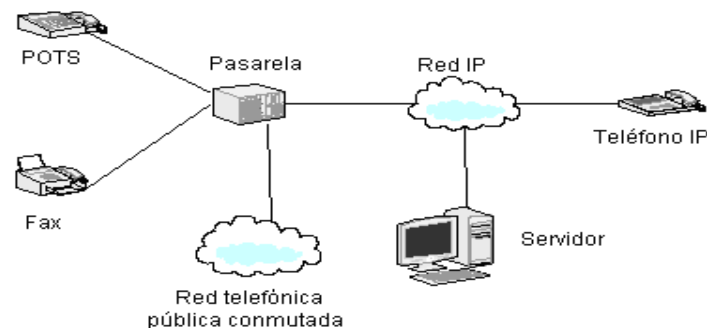


Figura 1. Componentes VoIP [7].

Una vez se establece la llamada, la conversación será digitalizada y transmitida a través de la red como paquetes IP. RTP (Real-time Transport Protocol) se utiliza conjuntamente con UDP (User Datagram Protocol) proporcionando funciones de transporte a aplicaciones de datos en tiempo real. En conclusión el modo en que se encapsula las tramas de VoIP es RTP/UDP/IP. En la figura 2 se puede observar una trama de VoIP de una LAN y WAN [7].

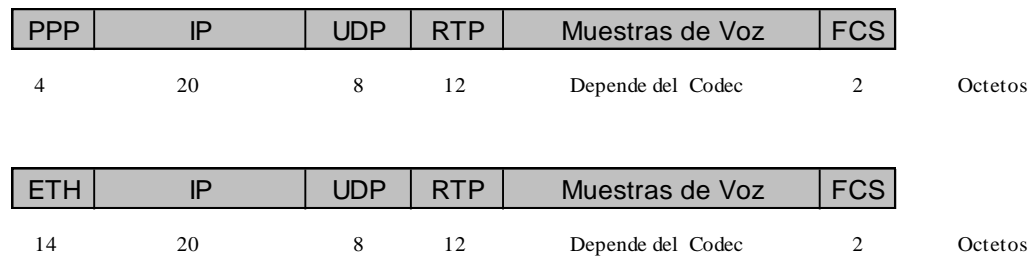


Figura 2. Encapsulación de VoIP en una trama PPP y Ethernet.

“Por ejemplo, si el CODEC usado es G.711 y el periodo de paquetización es 20 ms, la carga útil (payload) sería de 160 bytes. Esto resulta en un tamaño de trama total de 206 bytes en WAN y 218 bytes en LAN”.

2.2.3 Codecs

Antes que la voz analógica pueda ser transmitida sobre una red IP, esta debe ser primero digitalizada. Para la transmisión de la voz por las redes de paquetes los codecs son vitales, se busca con ellos la mayor eficiencia posible y disminución de la redundancia, existen muchos codecs, pero lo que se requiere para la escogencia de alguno es un balance entre complejidad del codec, la eficiencia de la carga útil y el retardo en la paquetización [8] [9]. Los estándares comunes de codificación son listados en la tabla 1 [7].

Estándar de Codificación	Algoritmo	Velocidad de datos
G.711	PCM (Modulación por Pulsos Codificados)	64 Kbps
G.726	ADPCM (Modulación por Pulsos Codificados Diferencial Adaptativa)	16, 24, 32, 40 Kbps
G.728	LD-CELP (Predicción lineal con excitación por código de bajo retardo)	16 Kbps
G.729	CS-ACELP (Estructura Algebraica Conjugada CELP)	8 Kbps
G.723.1	MP-MLQ (Cuantificación de Probabilidad Máxima de Multi impulso)	6.3 Kbps 5.3 Kbps
	ACELP (Predicción lineal con excitación por código algebraico)	6.3 Kbps 5.3 Kbps

Tabla 1. Estándares de codificación

2.2.4 *Calidad de servicio* en el marco de voz en redes NGN

Anteriormente se realizó una descripción de una red voz sobre IP, sus componentes más importantes y los diferentes codecs que sirven para la digitalización de la voz, aspecto necesario para su transporte en las redes de paquetes. Una visión general de la importancia de la calidad de servicio en redes IP y su estrecha relación principalmente con la voz (telefonía IP) en las redes NGN se da a continuación.

Cuando se habla de calidad de servicio se deben tener en cuenta diferentes aspectos intrínsecos de la red que permiten cuantificar su nivel, además de esto, el usuario busca también comodidad en el servicio, disponibilidad, seguridad etc. En lo relacionado con la voz, uno de los objetivos es conseguir un alto grado de inteligibilidad durante las conversaciones.

El crecimiento progresivo de Internet y la mejora de sus características lleva a que las redes de conmutación de paquetes se vuelvan más flexibles, motivo por el cual se esta

pensando en un nuevo concepto de red, donde converjan todos los servicios de comunicaciones ya sea de datos y/o servicios de tiempo real como voz y video, paulatinamente se esta evolucionando hacia este nuevo entorno.

Para asegurar la evolución de la telefonía IP dentro de estas nuevas redes es necesario que se ofrezcan mecanismos de QoS de manera automática.

En las redes de paquetes, el tráfico de voz necesita cierta calidad en lo referente a dar prioridad a los paquetes, tanto a nivel de acceso, como a nivel de interconexión, esto exige que se realice la optimización de la red en su conjunto y la organización del tráfico de la manera más adecuada posible, con el fin de responder de manera eficiente a los problemas que se presentan. Uno de estos problemas es la congestión, se desea conseguir la disminución de ésta por medio de técnicas que optimicen el funcionamiento de las redes para obtener retardos más pequeños, y aumento en el rendimiento

El problema de ofrecimiento de calidad de servicio en las redes de paquetes está asociado con la capacidad y mecanismos que utilizan los puntos de acceso, así como también en los puntos de interconexión y demás dispositivos para el tratamiento de los paquetes. Si un usuario necesita que se le garantice eficacia y eficiencia en el transporte de sus datos y aplicaciones por parte de un proveedor de servicios de Internet, puede hacer uso de los SLA (acuerdos de nivel de servicio) los cuales expresan los diferentes requisitos que se deben tener una determinada red para el soporte de varias aplicaciones.

Si en una red IP se quiere prestar el servicio de voz, se deben tener en cuenta algunos aspectos de calidad con los cuales se pueda lograr ofrecer una comunicación con alto grado de similaridad, comparable con el de las telefónicas tradicionales. Para lograr esto

es necesario superar una serie de dificultades técnicas propias de la transferencia de datos [8].

2.2.5 Problemas técnicos relacionados con las redes de paquetes

Entre los parámetros a considerar en las redes IP y que son causa del deterioro de la calidad ofrecida, en este caso en la comunicación de la voz, se encuentran:

- *Retardo*: En pocas palabras, el retardo es el tiempo total que se gasta entre la transmisión de un paquete y su llegada al receptor. En las redes IP muchos son los tipos de retardo entre los cuales se pueden considerar: retardos de codificación, retardos de empaquetado, retardos de transmisión, retardos de propagación, retardos de encolamiento, retardos de descompresión etc. [10]. La ITU recomienda un retardo en una sola vía de no más de 150ms para la mayoría de aplicaciones, con un límite de 400ms para una comunicación de voz aceptable.
- *Variación del Retardo (Jitter)*: es la variación estadística del tiempo en la llegada de paquetes, principalmente se debe al retardo en las colas. Es decir, se presentan cambios temporales en el arribo de los paquetes al receptor.
- *Pérdidas*. Se presentan primordialmente cuando se exhibe el fenómeno de congestión, además, se pueden dar por vencimiento del tiempo de vida de un paquete, caída o deterioro de un enlace etc. Esta pérdida de información afecta la calidad de las conversaciones en las redes IP.
- *Eco*: Su presencia se hace visible cuando existe poco acoplamiento entre los dispositivos electrónicos que se utilizan para el tratamiento de la señal analógica. Esto hace que se refleje parte de la señal transmitida hacia el transmisor.

Para ayudar a subsanar estos problemas de alguna manera, y con el fin de disminuir el deterioro de la calidad en los servicios, especialmente los de tiempo real, se han creado una serie de métodos y protocolos para ofrecer QoS en la redes IP. Algunos de ellos son:

- Reservación de recursos (RSVP, IntServ).
- Organización del tráfico por medio de prioridades (Diffserv).
- MPLS

Cada uno de estos métodos y protocolos tienen sus propias características. En ¹ se da una visión general de algunos de ellos y el entorno donde se aplican.

2.3 TELETRÁFICO Y TEORÍA DE COLAS

2.3.1 *Teletráfico*

La teoría de teletráfico ofrece una serie de herramientas matemáticas que pertenecen a la teoría de la probabilidad, cuyo fin es tratar las redes de telecomunicaciones en diversas etapas, desde su planeación hasta su operación y mantenimiento. En cada una de estas etapas se deben tomar una serie de decisiones con base en medidas de tráfico, para lograr un equilibrio entre capacidad del sistema y grado de servicio, en este ámbito la teoría es un instrumento adecuado.

El teletráfico abarca todo tipo de tráfico de comunicación de datos y tráfico de telecomunicaciones. Aunque también se puede aplicar a otros tipos de tráfico como el aéreo, vial etc. a la hora de diseñar un sistema de comunicaciones, el objetivo es desarrollar sistemas en lo posible de bajo costo y que respondan de una manera efectiva a la demanda actual y futura.

¹ Ver ANEXO C

2.3.2 Conceptos de tráfico.

El concepto de tráfico se utiliza normalmente para indicar la intensidad de tráfico, es decir el tráfico por unidad de tiempo.

Intensidad de tráfico. De la intensidad de tráfico se puede decir que es el número de recursos ocupados en un instante dado, con respecto a su totalidad. Para el calculo de la intensidad de tráfico media se utiliza la siguiente formula

$$Y(T) = \frac{1}{T} \int_0^T n(t) dt, \quad (2.1)$$

donde $n(t)$ indica el número de dispositivos ocupados en el tiempo t [11].

El tráfico trasportado en un lapso específico, se refiere al volumen de tráfico llevado por un conjunto de servidores durante un intervalo de tiempo T .

Tráfico ofrecido A. Es un valor teórico que solo puede calcularse en función del tráfico transportado, debido a que este no se puede medir, conceptualmente seria el tráfico total que se puede transportar teniendo en cuenta la capacidad del sistema.

Una formula para la valoración cuantitativa del tráfico ofrecido es la siguiente:

$$A = \lambda \cdot s \quad (2.2)$$

Donde:

λ = número medio de llamadas ofrecido por unidad de tiempo.

s = tiempo medio de servicio.

Tráfico perdido o rechazado: Si al tráfico ofrecido se le resta el tráfico transportado, da como resultado el tráfico rechazado [11].

Tráfico de múltiple tasa: “Si se tienen llamadas que ocupan más de un canal, y otras del tipo i que ocupan d_i canales, el tráfico ofrecido expresado en cantidad de canales ocupados se calcula con la siguiente ecuación:”[11]

$$A = \sum_{i=0}^N \lambda_i \cdot s_i \cdot d_i \quad (2.3)$$

Donde

N = número de tipos de tráfico.

λ_i = velocidad de arribos del tipo i .

s_i = tiempo medio de ocupación del tipo i .

Unidad erlang: es una unidad adimensional, es decir es independiente del tiempo de observación definida para la intensidad de tráfico.

En sistemas de transmisión de datos no se habla de tiempo de servicio, sino de necesidades de transmisión. Donde la utilización del sistema depende de alguna manera de su capacidad. Por lo tanto

$$\rho = \frac{\lambda \cdot s}{\varphi} \quad (2.4)$$

En la ecuación anterior ρ es la utilización del sistema, φ es la capacidad del sistema, λ es el número de tareas por unidad de tiempo y s tiempo medio de servicio. Se debe tener presente que la utilización debe estar en el intervalo $0 \leq \rho \leq 1$.

2.3.3 Teoría de colas

En la teoría de colas se incluye el estudio matemático de los sistemas de colas o líneas de espera (retardo/congestión), además provee una colección de modelos matemáticos para describirlas. Estos modelos ayudan a establecer el equilibrio entre costos del sistema y los tiempos promedio de la línea de espera para un sistema dado.

Debido al carácter aleatorio del proceso de llegadas y del tiempo de servicio, es muy difícil determinar que capacidad o tasa de servicio proporciona el balance correcto.

Problemas típicos de Teoría de Colas son: Compañía de telecomunicaciones, un hospital, una oficina de correos, etc. [12]. En la tabla 2 se pueden observar, problemas de este tipo en el campo de las redes de comunicaciones.

Sistema	Clientes	Servidor
Procesador	Programas o procesos	CPU, discos, dispositivos I/O, bus ...
MUX estadístico	Paquetes o celdas	Enlace de comunicaciones
Conmutador de circuitos	Llamadas	Canales
Red de acceso múltiple (LAN, LAN inalámbrica)	Paquetes o tramas	Medio (FO, UTP, RF)
Servicios Web	Requerimientos del cliente	Web Server

Tabla 2. Problemas típicos en teoría de colas.

2.3.3.1 *Definición* un sistema de colas está formado por llegadas de los clientes, una cola de espera y posibilidades de servicio con uno o varios servidores dispuestos en serie o generalmente en paralelo, los cuales deben presentar características similares. La cola se presenta cuando hay demanda de un determinado servicio y los servidores encargados de este, no están disponibles inmediatamente [12]. Como se observa en la figura 3.

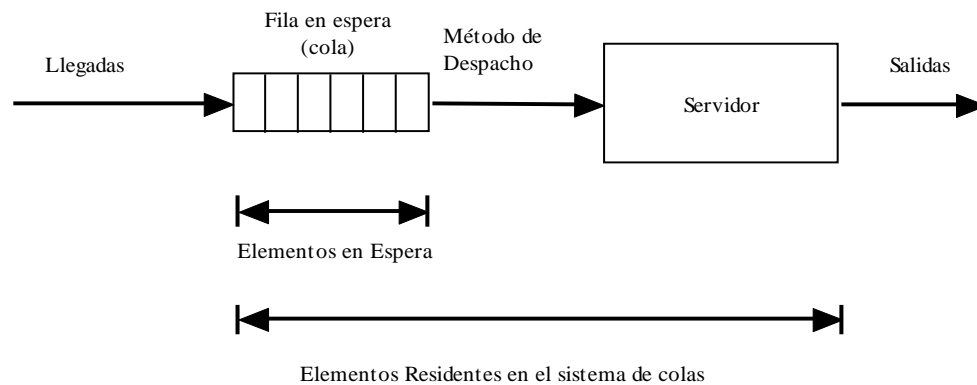


Figura 3. Sistema de colas

2.3.3.2 Conceptos básicos de los sistemas de colas: en general, una cola se forma en todo sistema cuando el tiempo de servicio es superior al tiempo entre llegadas. Siendo el *sistema*, el lugar donde se presta el servicio a los usuarios por medio de los *Servidores*, los cuales están encargados de satisfacer las solicitudes que se le presenten.

Los elementos más importantes que se deben tener en cuenta en un sistema de colas son: las llegadas, la cola, el servicio y la salida.

Las llegadas de la población al sistema, ocurren en forma aleatoria, es decir, no se puede establecer con precisión en que momento llegara un usuario. Del proceso de arribos interesa saber la distribución de probabilidad con que lo hacen, esta distribución tiene un parámetro muy importante, el cual corresponde a la tasa de arribos λ (número de entradas por unidad de tiempo). La unidad depende del tipo de sistema que se este analizando [13].

Las distribuciones de probabilidad consideradas para el proceso de llegadas y el tiempo de servicio en la teoría de colas son las siguientes:

M : Distribución de exponencial (markoviana) para el arribo de elementos al sistema.

D : Distribución determinística o degenerada (tiempos constantes), de antemano se conoce cuando ingresa un elemento de la población al sistema. Es considerada como un caso particular de los eventos de tipo aleatorio. Por ejemplo, un sistema que funcione con cita previa y que la consulta dure el tiempo previsto.

E_k : Distribución Erlang.

G : Distribución general. Es cualquier otra distribución en donde se conocen los parámetros de la distribución, como la media y la desviación estándar. Esto no es requisito indispensable en el sistema, pero se busca conocer al menos los parámetros más importantes que son los que aparecen en las expresiones finales de tiempos y longitudes.

Una característica importante de la cola es su población, y corresponde al número total de clientes que pueden requerir servicio en determinado momento (puede suponerse que el tamaño es infinito o finito, desde el punto de vista práctico el infinito significa una cantidad supremamente grande en comparación con otra). No es necesario que un elemento ingrese al sistema para que sea considerado como de la población del mismo [13].

Cuando en el sistema hay más de un usuario esperando ser atendido, al quedar uno de los servidores disponibles, hay que determinar cuál de los usuarios en la línea de espera será el que pase a recibir servicio. Es decir, es necesario establecer un proceso que permita ordenar la forma de atención de los elementos o usuarios en la cola; a este proceso se le conoce como disciplina de servicio [13]. Entre los modelos más utilizados se tienen los siguientes: *FIFO*, *LIFO* y *SIRO*.

Siendo *FIFO* (*First In First Out – Primero en entrar primero en salir*) la disciplina más usada y se toma por defecto, cuando no se especifica nada acerca de la disciplina de

servicio. Simplemente se da servicio “respetando” el orden de llegada, la cola esta ordenada según el orden de llegada de los usuarios. La disciplina *LIFO* (*Last In First Out* – *Última en entrar primera en salir*) atiende primero al último en arribar, por lo tanto la cola se ordena en orden inverso al de llegada. Es decir en forma de pila o stack, quien va entrando al sistema se coloca en la cabeza de la cola para ser atendido. Este tipo de disciplina es poco común y tiene aplicación en sistemas muy variados. En la disciplina *SIRO* (*Service In Random Order* – *Servicio en orden aleatorio*) se escoge aleatoriamente cuál de los usuarios en espera accederá al servicio, en esta no importa el orden de llegada.

La disciplina de servicio puede ser establecida mediante asignación de prioridades, la prioridad se asigna teniendo en cuenta algunas características de la población en espera, y esta es considerada en el momento de la atención.

La prioridad puede presentarse de dos formas de acuerdo a como se halla establecido en el sistema, *prioridad con truncamiento* o *prioridad sin truncamiento*. La prioridad con truncamiento implica la suspensión inmediata del servicio cuando llega un elemento con prioridad mayor, en la otra se espera hasta que el elemento sea atendido. Este método requiere políticas de asignación bien establecidas, las cuales eviten que un elemento con baja prioridad se quede indefinidamente en el sistema [13].

En un sistema de colas se tiene uno o varios *servidores*, los cuales están encargados de prestar el servicio. Los servidores se pueden ubicar en diferentes configuraciones, como por ejemplo: en serie, en paralelo o combinaciones de las dos.

Cuando están dispuestos en paralelo, todos prestan el mismo servicio, a diferencia de la combinación en serie, en la que, cada usuario que ingrese al sistema debe pasar por

todos los servidores activos, ya que todos prestan diferentes servicios; por este motivo se presenta una característica particular en este sistema, la cual consiste en la formación de colas al frente de cada servidor, debido a que el tiempo en cada uno de ellos no es necesariamente el mismo [13].

En la mayoría de los sistemas, el *tiempo de servicio* es una variable aleatoria² continua que depende del tipo de transacción que se vaya a realizar en el sistema. Por ejemplo, en las redes de datos, el tiempo de servicio, depende de la velocidad de los enlaces y la longitud del mensaje que se desea transmitir, en este caso el tiempo de servicio corresponde al tiempo que demora en llegar un mensaje transmitido desde el nodo *A* hasta el nodo *B* [13].

A la hora de diseñar los sistemas es necesario tener en cuenta que, sí el tiempo de servicio es mayor que el intervalo entre llegadas, la cola aumentará indefinidamente y el sistema puede llegar a colapsarse.

Un parámetro de interés a la salida del sistema es la tasa con que los elementos se reintegran a la población una vez han recibido el servicio. La tasa de salida depende del tiempo de servicio en proporción inversa [13].

En el análisis de redes es muy importante determinar los parámetros de la distribución de probabilidad de los tiempos, longitudes, arribo y salida del sistema y la cola. Se definen las siguientes variables aleatorias con sus respectivos parámetros, así, [13].

² Ver ANEXO A (conceptos básicos de probabilidad)

$E(n)$ = Valor esperado de la cantidad de clientes en el sistema. Es el número total de elementos en el sistema, incluyendo al que se encuentra en servicio.

n = Caracteriza el estado del sistema. Es una variable aleatoria discreta.

m = Longitud de la cola. Es una variable aleatoria discreta; no incluye el elemento que se encuentra en servicio.

$E(m)$ = Valor esperado de la longitud de la cola o valor esperado de la cantidad de clientes esperando servicio en la fila.

$\sigma(m)$ = Desviación estándar de la longitud de la cola.

$\sigma(n)$ = Desviación estándar de la cantidad de clientes en el sistema.

T = Tiempo en el sistema. Es una variable aleatoria continua. Incluye el tiempo de servicio.

$E(T)$ = Valor esperado del tiempo en el sistema.

w = Tiempo en la cola. Variable aleatoria continúa.

$E(w)$ = Valor esperado del tiempo en la cola.

$\sigma(T)$ = Desviación estándar del tiempo en el sistema.

$\sigma(w)$ = Desviación estándar del tiempo en la cola.

$\rho_n(t)$ = Probabilidad de que el sistema se encuentre en el estado n en el instante de tiempo t .

$\mu_n(t)$ = Tasa de salida del sistema cuando este se encuentra en el estado n en el instante de tiempo t .

$\mu_n(t) \Delta t$ = Probabilidad de una salida del sistema cuando se encuentra en el estado n en el instante t .

$\lambda_n(t)$ = Tasa de entrada al sistema cuando se encuentra en el estado n en el instante t .

$\lambda_n(t) \Delta t$ = Probabilidad de entrada en el instante t cuando el sistema se encuentra en el estado n .

A = Carga sobre el sistema medida en Erlangs. Representa el número de elementos que ingresan al sistema durante un tiempo de servicio.

ρ = Factor de utilización del servidor. Representa la carga neta sobre un servidor.

La probabilidad de estado, $\rho_n(t)$ sirve para determinar todos los parámetros definidos de la cola.

Fórmulas de Little: Estas relaciones son fundamentales pues permiten determinar los parámetros de una cola, tales como la longitud de la cola, longitud del sistema, tiempo en la cola y tiempo en el sistema, solo conociendo el valor de una de ellas. Las llamadas fórmulas de Little facilitan el cálculo de estos parámetros, puesto que no es necesario trabajar con las complejas fórmulas de probabilidades que permiten calcular los mismos parámetros [13].

A continuación se da un resumen para las expresiones de las longitudes de la cola y del sistema, los tiempos en la cola, sistema y la relación entre los tiempos en cola, sistema y servicio.

El valor esperado de la longitud de la cola ($E(m)$), es igual a la tasa promedio de llegadas (λ) por el valor esperado del tiempo en la cola ($E(w)$).

El valor esperado del tiempo que un usuario está en el sistema ($E(T)$), es igual al valor esperado del tiempo que un usuario permanezca en la cola ($E(w)$), mas el tiempo que se demore el *servidor* en prestar el servicio ($\frac{1}{\mu}$).

El número que se espera de clientes en el sistema $E(n)$, es igual a la tasa promedio de llegadas (λ) por el valor esperado del tiempo en el sistema $E(T)$.

λ = Tasa promedio de llegadas al sistema.

Por lo tanto,

$$E(m) = \lambda E(w) \tag{2.5}$$

$$E(n) = \lambda E(T) \tag{2.6}$$

$$E(T) = E[w] + \frac{1}{\mu} \tag{2.7}$$

Notación de Kendall: La notación de Kendall es usualmente utilizada para identificar las principales características de las colas, permite identificar la distribución de probabilidad que sigue el arribo de usuarios en la línea de espera, la distribución de probabilidad del tiempo de servicio, el número de servidores, la capacidad del sistema, el tamaño de la

población y la disciplina de servicio utilizada por el sistema. Es decir facilita el conocimiento de un sistema de colas con pocos parámetros [13].

La notación consta de seis elementos dispuestos de la siguiente manera:

$$1 / 2 / 3 / 4 / 5 / 6$$

Cada uno de los dígitos tiene su propio significado, así:

1: Indica la distribución que siguen los tiempos entre llegadas. Se utilizan las letras M , G y D para denotarlas, letra M (Markov) se utiliza cuando la distribución de entrada es de tipo exponencial, la letra G cuando es una distribución general y D si es determinística o constante.

2: Indica la distribución del tiempo de servicio y, por tanto, de las salidas del sistema. Al igual que en el primer dígito, también se utilizan las letras M , G y D para indicar el tipo de distribución.

3: Número de canales de servicio ó número de servidores que prestan el servicio, generalmente están dispuestos en paralelo, si es de otra forma se debe especificar.

4: Capacidad del sistema, tamaño máximo de la cola, longitud del buffer o capacidad de almacenamiento que se admite en el sistema en servicio. Por defecto, la cola tiene capacidad infinita.

5: Tamaño de la población (puede ser finito o infinito), corresponde al número de elementos que potencialmente requieren del servicio. Cuando el tamaño es infinito o muy grande en comparación con la capacidad del sistema no se indica en la casilla.

6: Disciplina de servicio que se maneja en la cola, es decir, el modo en que van a ser atendidos los usuarios que se encuentran en espera. Por omisión se considera una cola tipo FIFO.

Tipos de colas

Cola Tipo M/M/1: Es un tipo de encolamiento sencillo, donde según la notación de Kendall se tiene un único servidor, los arribos obedecen a una distribución exponencial, así como el tiempo de servicio [13].

Cola Tipo M/M/N: Aquí se tiene la misma organización que el tipo de cola anterior, es decir, tiempo entre llegadas de tipo exponencial, tiempo de servicios de tipo exponencial, pero se considera que se tiene N servidores idénticos [13].

Cola tipo M/M/N/K: Esta representación posee las mismas características que la cola $M/M/N$ pero se tiene una limitación dada por la letra k , donde según la notación de Kendall solo se pueden atender K usuarios simultáneamente [13].

Cola Tipo M/G/1: Este modelo de encolamiento responde a clientes cuyas llegadas están de acuerdo a los procesos de Poisson, se tiene un único servidor los tiempos de servicios no obedecen a una distribución exponencial. Consecuentemente la letra G hace referencia a un modelo general [13].

Colas tipo M/E_k/1: En las colas de este tipo el tiempo de servicio está regido por una distribución de Erlang, se considera un simple servidor, el arribo de los clientes esta de acuerdo a procesos de Poisson [13].

Colas tipo M/G/1/N: La notación de este modelo indica que es del mismo tipo que *M/G/1* con la salvedad que la cantidad de servidores presentes es N [13].

2.4 PROCESOS DE MARKOV.

Un proceso de Markov se define como un proceso estocástico, es decir, una familia de variables aleatorias delimitadas sobre un espacio de probabilidad.

Los procesos de Markov Proveen un medio flexible, poderoso y eficiente para la descripción y análisis de las propiedades de los sistemas (computacionales, redes, etc.) dinámicos. El rendimiento y la confiabilidad son medidas que fácilmente pueden ser obtenidas. Además, los procesos de Markov constituyen la teoría fundamental subyacente de los sistemas de colas [15].

Entre los procesos de Markov más utilizados se encuentran los MMPP, los cuales involucran dos procesos independientes, los procesos de Poisson y las cadenas de Markov. Por lo tanto es necesario dar una serie de definiciones que permitan entenderlos.

2.4.1 Cadenas de Markov.

Definición 1. La secuencia de variables aleatorias $\{X_t : t \in \mathbb{N}\}$ se dice es una cadena de Markov en tiempo discreto, si ella toma valores en algún espacio muestral discreto Ω ($X_t \in \Omega$ para todo t) y, para todo $t \in \mathbb{N}$ y para todo $i_k \in \Omega$, entonces

$$P[X_n = i_n \mid X_1 = i_1, X_2 = i_2, \dots, X_{n-1} = i_{n-1}] = P[X_n = i_n \mid X_{n-1} = i_{n-1}] \quad (2.8)$$

Es decir, una cadena de Markov en tiempo discreto es una serie temporal de valores discretos en la cual, el valor del estado presente solo depende del estado anterior. Los valores Ω que puede tomar la cadena de Markov se conocen como los estados de la cadena y usualmente son números enteros. Es decir $\Omega \subseteq \mathbb{Z}$ [14], [15].

Definición 2. Una cadena de Markov homogénea y en tiempo discreto, es una cadena de Markov en tiempo discreto, por lo cual la probabilidad condicional

$$P[X_n = i_n | X_{n-1} = i_{n-1}] \quad (2.9)$$

es independiente de n [14].

Las cadenas de Markov aquí discutidas son en tiempo discreto, en las cadenas de Markov homogéneas los estados están numerados con enteros. Cuando se hace referencia a cadenas de Markov se está refiriendo a las cadenas de Markov homogéneas en tiempo discreto con los estados representados por enteros, para los MMPP en tiempo continuo se necesita la matriz transiciones, pero esta es fácilmente obtenida por medio de la matriz de probabilidades de transición en tiempo discreto [14].

Definición 3. Las probabilidades de transición en tiempo discreto, de las cadenas de Markov homogéneas están dadas por P_{ij} en donde

$$P_{ij} = P[X_{n+1} = j | X_n = i] \quad (2.10)$$

Esta es la probabilidad de que la cadena estando en el estado i pase al estado j en un salto [13], [14].

Definición 4. La probabilidad de transición para el m -ésimo paso de una cadena de Markov discreta está dada por $P_{ij}^{(m)}$ donde [14], [15].

$$P_{ij}^{(m)} = P[X_{n+m} = j | X_n = i] \quad (2.11)$$

Definición 5. Un proceso modulado de Markov es un proceso o serie temporal $\{Y_t : t \in \mathbb{N}\}$ en el cual el estado de Y_t depende de la cadena de Markov que lo modula, o más generalmente, donde la función de densidad de Y_i es función de una cadena de Markov subyacente.

$$Y_t = g(X_t), \quad (2.12)$$

para alguna función $g(x)$ o, en general Y_t podría ser generada por el muestreo de una distribución de probabilidad la cual depende del estado de la cadena (fundamental) subyacente [14].

Definición 6. Una cadena de Markov es llamada irreducible, si todos los estados pueden ser alcanzados desde otro. Es decir, para todo $i, j \in \Omega$, existe algún m tal que $P_{ij}^{(m)} > 0$ [14].

Definición 7. El estado $i \in \Omega$ de la cadena de Markov es *periódico* con periodo γ , si los retornos al estado i son solo permitidos después de algún múltiplo del periodo γ [14], [15] entonces:

$$P_{ii}^{(m)} \begin{cases} \geq 0 & m \in \{\gamma, 2\gamma, 3\gamma, \dots\} \\ = 0 & \text{en otro caso.} \end{cases} \quad (2.13)$$

Definición 8. Sea $j \in \Omega$ un estado de la cadena de Markov y $n \in \mathbb{N}$ el numero de saltos, entonces $r_j^{(n)}$ es la probabilidad que el primer retorno al estado j ocurra después de n pasos o

$$r_j^{(n)} = P[X_{t+n} = i | X_t = i, X_{t+1} \neq i, X_{t+2} \neq i, \dots, X_{t+n-1} \neq i]. \quad (2.14)$$

Definición 9. Los términos recurrente y transiente están definidos en términos de la probabilidad r_j que el estado de la cadena nunca retorne a $j \in \Omega$, donde

$$r_j = \sum_{n=1}^{\infty} r_j^{(n)}. \quad (2.15)$$

Un estado $j \in \Omega$ con probabilidad $r_j = 1$ es recurrente, en otro caso, si $r_j < 1$ j es llamado un estado transiente [14].

Definición 10. El término tiempo medio de recurrencia M_i de un estado $i \in \Omega$ de una cadena de Markov esta dado por

$$M_i = \sum_{n=1}^{\infty} r_j^{(n)}. \quad (2.16)$$

En otras palabras, M_i es el valor esperado del primer tiempo de recurrencia. Un estado recurrente $i \in \Omega$ de una cadena de Markov se dice ser recurrente nulo si $M_i = \infty$ y recurrente no nulo si $M_i < \infty$ [14].

Definición 11. La probabilidad de encontrar el sistema en el estado $j \in \Omega$ en el tiempo n esta dado por

$$\pi_j^{(n)} = P[X_n = j], \quad (2.17)$$

y el limite de la probabilidad (si existe) esta dado por

$$\pi_j = \lim_{n \rightarrow \infty} \pi_j^{(n)}. \quad (2.18)$$

Los términos π_j son conocidos como las probabilidades de equilibrio para los estados.

Los siguientes dos teoremas son tomados de [16].

Teorema 1. "Los estados de una cadena de Markov irreducible son todos transientes, o todos recurrentes no nulos o todos recurrentes nulos. Si son periódicos, entonces todos los estados tienen el mismo periodo γ ".

Teorema 2. "En una cadena de Markov homogénea irreducible y aperiódica, los límites de las probabilidades de equilibrio π_j siempre existen y son independientes de la distribución de X_1 (el estado inicial de la cadena). Además se puede dar cualquiera de los dos casos siguientes:

(1) Todos los estados son transientes o todos los estados son recurrentes nulos y

$$\pi_j = 0 \text{ para todo } j \in \Omega, \text{ o}$$

(2) Todos los estados son recurrentes no nulos y para todo $j \in \Omega$, $\pi_j = 1/M_j > 0$.

Si ocurre el segundo caso en el Teorema 2, entonces las cantidades π_j son únicamente determinadas a partir de las ecuaciones

$$\sum_{j \in \Omega} \pi_j = 1, \tag{2.19}$$

y

$$\pi_j = \sum \pi_i P_{ij}. \tag{2.20}"$$

Definición 12. Un estado de la cadena de Markov se dice que es *ergódico* si este es irreducible, aperiódico y recurrente no nulo. Si todos los estados son ergódicos entonces la cadena es así misma ergódica y π_j es conocida como la probabilidad de equilibrio del estado j [14].

Definición 13. La matriz de probabilidades de transición P , es la matriz formada por los elementos P_{ij} , para todo $i, j \in \Omega$ y el vector probabilidad de equilibrio π es el vector de todas las probabilidades de equilibrio π_j para todo $j \in \Omega$.

Por ejemplo, si $\Omega = \{1, 2, \dots, n\}$ (cadena de Markov con n estados), entonces la matriz de transición está dada por:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & P_{n3} & \cdots & P_{nn} \end{bmatrix}, \quad (2.21)$$

y el vector de probabilidad de equilibrio

$$\pi = (\pi_1, \pi_2, \dots, \pi_n). \quad (2.22)$$

La ecuación (2.20) puede ser escrita en términos de P y π así:

$$\pi = (\pi_1, \pi_2, \dots, \pi_n) \begin{bmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & P_{n3} & \cdots & P_{nn} \end{bmatrix} = \pi P. \quad (2.23)$$

con la ecuación (2.19) éstas siempre formarían $n + 1$ ecuaciones para un estado n de la cadena y por consiguiente π esta siempre en principio determinada completamente por las ecuaciones (2.19) y (2.20) [14].

2.4.2 Procesos de Poisson

Los procesos de Poisson son útiles para modelar diversas situaciones en el mundo real, en este caso los procesos de llegada a los sistemas de colas. La ocurrencia de los

arribos es totalmente aleatoria en el tiempo, con tasa λ e independiente del pasado, además están distribuidos exponencialmente. Así:

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \dots,$$

Donde $N(t)$ es el número de arribos en el intervalo $[0, t]$.

2.5 MOVIMIENTO BROWNIANO

El movimiento browniano pertenece a la temática relacionada con los procesos estocásticos, siendo este uno de los más importantes. Gran cantidad de fenómenos aleatorios se pueden modelar por medio de él, ejemplo de esto es el movimiento que experimentan pequeñas partículas suspendidas en un fluido. La fundamentación matemática rigurosa de este proceso fue hecha por Norbert Wiener en 1923 [17]. El movimiento Browniano se puede definir de la siguiente manera:

Definición 14: Sea $B(t)$ un proceso estocástico de tiempo continuo tal que

- (i) $B(t)$ es gaussiano.
- (ii) $B(0) = 0$.
- (iii) $B(t)$ posee incrementos independientes y estacionarios.
- (iv) $E[B(t) - B(s)] = 0$, en donde t y s son tiempos distintos.
- (v) $Var[B(t) - B(s)] = \sigma^2 |t - s|$ ($Var[B(t)] = \sigma^2 t$)

Entonces $B(t)$ es un movimiento Browniano.

La función de densidad de probabilidad de $B(t)$ tiene la siguiente expresión

$$f_B(b, t) = \frac{1}{\sqrt{\sigma^2 2\pi t}} e^{-\frac{b^2}{2\sigma^2 t}} \quad (2.24)$$

De la definición 14, el movimiento Browniano es un proceso estocástico el cual se puede mirar como una colección de variables aleatorias o una función aleatoria, si se tienen incrementos independientes no existe dependencia en la cantidad de movimiento en un lapso y otro posterior no superpuesto. El término incrementos estacionarios se refiere a que el desplazamiento dentro de un intervalo solo depende de la longitud del intervalo y no del momento donde se origino dicho intervalo. Si $B(t)$ es gaussiano, la suma del gran número de pequeños desplazamientos se aproxima a una distribución normal en intervalo de tiempo específico.

La autocorrelación de $B(t)$ se puede expresar como $R_B(t_1, t_2) = [E(B(t_1)B(t_2))] = t_1$, como $B(t)$ tiene una media nula, la autocovarianza es igual a la autocorrelación, de modo que $C_B(t_1, t_2) = [E(B(t_1)B(t_2)) - t_1] = 0$ donde $t_1 < t_2$.

2.6 CARACTERÍSTICAS DEL TRÁFICO EN REDES

Como las aplicaciones que usa el público en las redes cambia, las propiedades estadísticas y la naturaleza del tráfico a través de la red cambia igualmente. Donde inicialmente se tenían redes de telefonía y redes de datos por separado, están evolucionando a redes donde existen servicios que demandan respuestas en tiempo real como la voz y el video.

Los primeros modelos de tráfico fueron derivados empíricamente de las trazas medidas de tráfico de la red pública. En el pasado, estas redes fueron en la mayoría de los casos

limitadas a las redes telefónicas de conmutación de circuitos. Como resultado, la mayoría de las propiedades capturadas fallan al caracterizar el tráfico encontrado en las modernas redes integradas [18].

En los modelos analizados en [19], [20], [21] se describe un determinado tráfico en función de parámetros estadísticos tales como: la tasa promedio de arribos, la desviación estándar, la varianza etc. que si bien dan información del comportamiento del tráfico, no capturan de una manera adecuada las características que presentan la redes integradas actuales debido a que estas presentan una mayor complejidad. Para solventar este problema, en la literatura referente al modelamiento en redes de telecomunicaciones se habla de autosimilitud. Antes de hablar de autosimilitud primero se toca de forma general el tema de Fractalidad y su relación con la redes de Telecomunicaciones [18].

2.6.1 Fractalidad

El término fractal fue introducido en la literatura por B. Mandelbrot [22], el cual esta ligado íntimamente con otro concepto importante como lo es la autosimilitud, donde la característica más relevante esta relacionada con la persistencia de cierta propiedad de un objeto. Un objeto es autosimilar o fractal si al realizar una serie de iteraciones o en otras palabras, si al realizar una magnificación³ se obtienen muestras que más o menos se conservan uniformemente organizadas sobre un amplio rango de escalas. En la figura 4 se puede observar este fenómeno, donde se ha realizado una ampliación (lado derecho) de la figura original (lado izquierdo).

³ En este contexto se entiende magnificación como el proceso de ampliación de un sector específico.

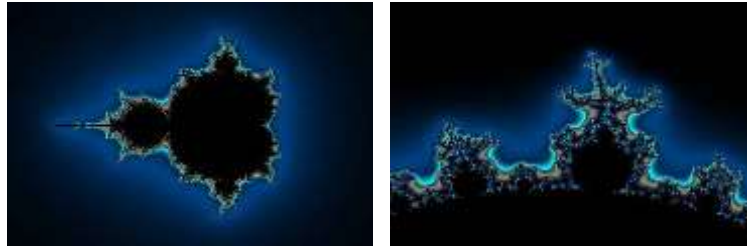


Figura 4 Conjunto de Mandelbrot [36].

El concepto de dimensión es una de las características representativas de los fractales. Cuando la dimensión de un objeto es representado por el número de coordenadas que son necesarias para describir un punto de este, se tiene: en \mathfrak{R}^2 , una línea se puede representar por dos coordenadas; en \mathfrak{R}^3 , por 3 coordenadas y en \mathfrak{R}^n por n coordenadas, en este caso se habla de dimensión euclidiana. Cuando el enfoque es solo tener en cuenta el número de coordenadas que sirven para representar un objeto, en lugar del espacio en el que el objeto reside, se habla de dimensión topológica en donde solo se necesita una coordenada para representar una línea, dos coordenadas para un plano y tres coordenadas para una superficie.

Cuando se habla de dimensión fractal la definición dada por Hausdorff es la más conveniente ya que su desarrollo matemático es relativamente fácil debido a que esta se basa en medidas, un gran inconveniente de esta definición se presenta para el cálculo de la dimensión de algunos objetos fractales [23].

En la determinación de la dimensión de Hausdorff D_f según se tiene en [17], “para ocupar un área con volumen L^{D_f} en un espacio de D-dimensiones, este es cubierto por

cubos cuyo volumen es ε^{D_f} . El número mínimo de cubos no solapados utilizados es

$M(\varepsilon) = L^{D_f} \left(\frac{1}{\varepsilon^{D_f}} \right)$ de aquí se obtiene que:"

$$D_f = \frac{\log M(\varepsilon)}{\log(1/\varepsilon)} \quad (2.25)$$

Una idea más intuitiva puede servir para entender mejor el concepto de dimensión fractal.

Tomando una línea en una dimensión y la dividimos en 3 partes. El factor de escala sería

$$r = \frac{1}{N} = \frac{1}{3} \quad (2.26)$$

Si hacemos esto en dos dimensiones se habla de un cuadrado este está dividido en

$N = 9$: donde:

$$r = \frac{1}{3} = \frac{1}{\sqrt{9}} \quad (2.27)$$

En tres dimensiones la figura sería un cubo en la cual $N = 27$ de ahí que:

$$r = \frac{1}{3} = \frac{1}{\sqrt[3]{27}} \quad (2.28)$$

De las tres ecuaciones anteriores se tiene que

$$r = \frac{1}{\sqrt[D]{N}} \quad (2.29)$$

Sacando logaritmo a ambos lados $\log(r) = -\log(N)^{1/D}$, de aquí

$$D = \frac{\log(N)}{\log(1/r)} \quad (2.30)$$

Un ejemplo de figura fractal es el llamado copo de Koch el cual se obtiene dividiendo un segmento en tres partes iguales, donde la parte de la mitad es reemplazada por dos

segmentos de igual longitud, si este procedimiento se realiza infinidad de veces el resultado de este procedimiento se puede visualizar en la figura 5. Aquí $N = 4$, $r = \frac{1}{3}$.

Aquí $D = 1,261$

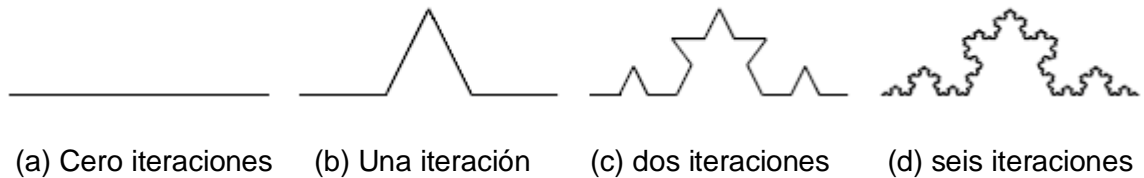


Figura 5. Dimensión fractal

2.6.2 Fractalidad en las Redes de Telecomunicaciones

Al hacer referencia al tráfico en las redes de telecomunicaciones se habla de autosimilitud estocástica fractal, es decir, si se hace una ampliación de una parte de un objeto, si bien no es idéntico al fragmento inicial, si se conservan las propiedades estadísticas en diferentes escalas. Mediciones de tráfico realizadas a redes en funcionamiento evidenciaron que estas poseen un comportamiento autosimilar o de larga dependencia. El grado de autosimilitud depende de la carga de la red, (Ej.: ráfagas sobre un amplio rango de escalas de tiempo [1]). En contraste con los modelos que solo describen la dependencia de corto rango, donde dominaba el concepto de que el tráfico se podía aproximar o suponer que es exponencial.

Los ambientes de red en los que se enfocaron los primeros estudios fueron los entornos más comunes, las redes LAN y WAN sobre redes IP, demostrando que la superposición de muchas fuentes generadoras de tráfico independientes e idénticamente distribuidas ON/OFF cada una de las cuales exhibe el fenómeno llamado efecto Noah “Noah Effect” resulta un tráfico agregado autosimilar, donde los periodos ON y OFF son independientes

unos de los otros. De este modo se conoce que el tráfico WWW (World Wide Web) muestra autosimilitud así como el video con tasa de bits variable (VBR video) [24]. Estas medidas llevan a que la variación en diferentes escalas de tiempo no sean aisladas, por el contrario, estas tienen alguna relación. En otras palabras un retraso o adelanto en determinado tiempo afecta directamente el comportamiento visto en un tiempo posterior.

Dado que es cierta la presencia de la naturaleza fractal o autosimilar en las redes de telecomunicaciones, es necesario una correcta caracterización de los diferentes tipos de tráfico ya sea datos, voz o video; por tal motivo, es de vital importancia, tanto la estadística como la inferencia científica, para realizar una correcta detección y cuantificación de estos fenómenos. Para realizar esto es necesario muestras de redes en funcionamiento o en su defecto simulaciones que se acerquen a redes reales. El objetivo principal es concluir si las muestras de datos están acorde con la estructura del modelo propuesto.

2.6.3 Autosimilitud

La autosimilitud se define como la apreciación de una estructura estadísticamente similar a lo largo de un amplio rango de escalas, es decir que la señal tiene un “aspecto” parecido (estadísticamente hablando aunque varíe la escala). Podemos ilustrar gráficamente la autosimilitud estocástica en la figura 6, en ella se puede observar el número de paquetes que transitan por un segmento de red Ethernet para distintas escalas de tiempo.

Interpretación visual: la figura 6 describe una Traza de tráfico Ethernet en 5 diferentes escalas de tiempo. Esta traza fue medida en el Bellcore (Bell Communications Research) network en agosto 1989 [2].

Iniciando con una unidad de tiempo de 100 segundos para la figura de la parte superior, cada mapa subsiguiente es obtenido a partir del anterior, incrementando la resolución del tiempo por un factor de 10. Además se concentra en un sub-intervalo escogido aleatoriamente, como indica la sombra más oscura en la figura 6. La unidad de tiempo correspondiente a la escala temporal más fina es 10 milisegundos, figura de la parte inferior. En el conjunto de trazas de la figura 6 es importante observar lo siguiente:

- ✓ Todos los esquemas muestran intuitivamente mucha similitud entre uno y otro en términos de su distribución, y son distintivamente diferentes del ruido blanco. Esto implica la conservación de características estadísticas a diferentes niveles de agregación.

- ✓ La propiedad de escala (eje y) y la ausencia de un tamaño de “ráfaga” en un rango en la escala de tiempo desde milisegundos a minutos y horas consiste de sub-periodos de ráfagas separados de sub-periodos sin ráfagas.

Esta escala invariante o rasgo autosimilar de las redes de tráfico modernas esta drásticamente diferenciada del tráfico telefónico convencional y de los modelos estocásticos tradicionales considerados en la literatura. Los últimos, típicamente producen esquemas de conteo de paquetes que son indistinguibles del ruido blanco después de la agregación sobre unos pocos cientos de milisegundos. La figura 6, prueba la presencia del fenómeno de autosimilitud en los paquetes de tráfico Ethernet. La presencia de gran variabilidad en diferentes escalas de tiempo, la semejanza estadística de las propiedades de segundo orden en el tráfico que circula por la redes actuales hacen necesaria la utilización de estos nuevos conceptos en el modelamiento de redes.

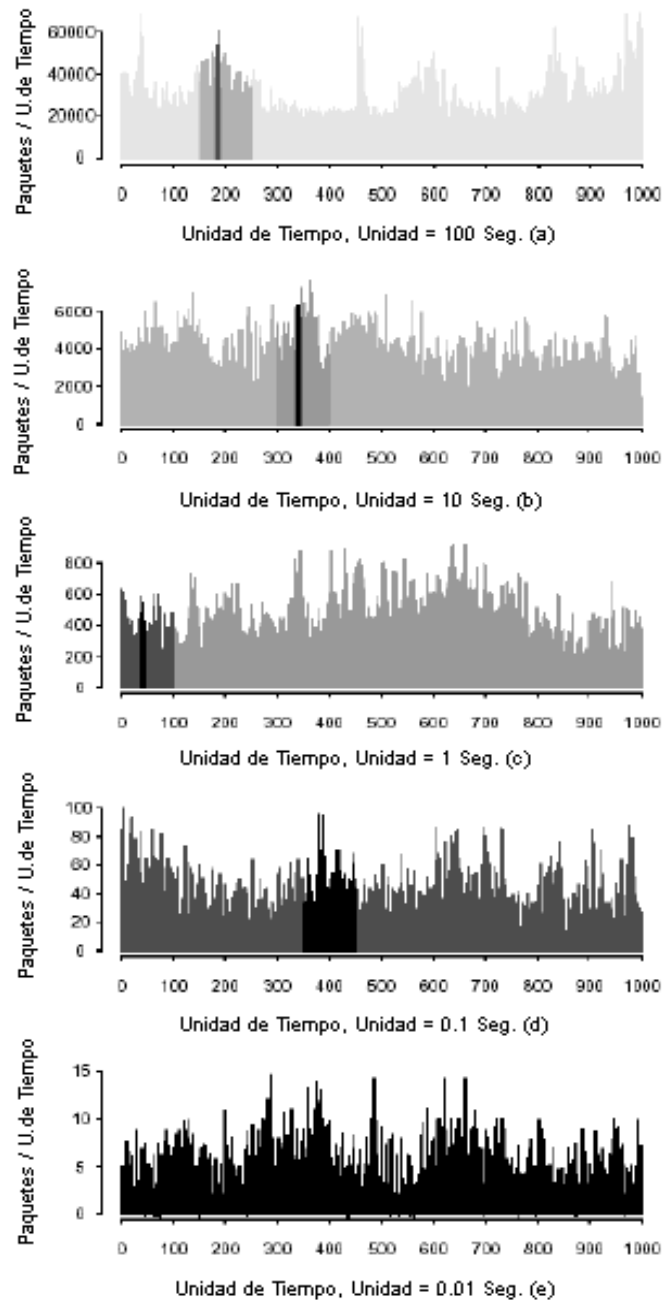


Figura 6. Ilustración gráfica de la autosimilitud [2].

Formulación matemática: El concepto de autosimilitud no es meramente una descripción intuitiva, está también puede ser descrita matemáticamente con el fin de explicar este fenómeno formalmente.

Inicialmente se puede hablar de los procesos conocidos como: exactamente autosimilares. Un proceso autosimilar exacto es un proceso estocástico de covarianza estacionaria.

El tráfico real en las redes no es exactamente autosimilar. Aunque las características de orden mayor difieren, se puede caracterizar matemáticamente por medio de unos modelos exactos o aproximados que son la mejor vía para describir la autosimilitud.

Consideremos un proceso estocástico en tiempo discreto o serie temporal $X(t)$, $t \in \mathbb{Z}$, donde $X(t)$ es interpretado como el volumen de tráfico medido en paquetes, en bytes o en bits, es decir, $X(t)$ es el tráfico total alcanzado en un instante de tiempo t . $X(t)$ es el proceso de incremento correspondiente a $Y(t)$, donde $Y(t)$ es tomado como un proceso “acumulativo”. Entonces $X(t) = Y(t) - Y(t-1)$.

Para propósitos de modelamiento de tráfico, se desea que $X(t)$ sea “estacionario” en el sentido de que su comportamiento y estructura sea invariante con respecto a una escala de tiempo [25].

$X(t)$ es un proceso estrictamente estacionario si $(X(t_1), X(t_2), \dots, X(t_n))$ y $(X(t_1+k), X(t_2+k), \dots, X(t_n+k))$ poseen la misma distribución conjunta para todo $n \in \mathbb{Z}$, t_1, t_2, \dots, t_n , $k \in \mathbb{Z}$.

Lo descrito anteriormente es una limitación, porque el criterio de estrictamente estacionario es muy restrictivo, y en un gran número de procesos físicos y de ingeniería

no se puede encontrar tal criterio. Por este motivo se presenta una forma de estacionaridad la cual requiere que la función de autocovarianza.

$$\gamma(r, s) = E[(X(r) - \mu)(X(s) - \mu)] \quad (2.31)$$

Sea invariante a la translación Ej.: $\gamma(r, s) = \gamma(r+k, s+k)$ para todo $r, s, k \in \mathbb{Z}$. Los primeros dos momentos se asumen que existen y son finitos, donde:

$$\mu = E[X(t)] \quad (2.32)$$

$$\sigma^2 = E[(X(t) - \mu)^2] \quad (2.33)$$

Para todo $t \in \mathbb{Z}$

Considerando un proceso acumulativo $Y(t)$, donde $t \in \mathbb{R}$, una definición de autosimilitud para procesos continuos en el tiempo en el sentido de distribuciones dimensionales finitas es la siguiente:

Definición 15: $Y(t)$ es un proceso autosimilar con parámetro de autosimilitud H , parámetro de Hurst $H(0 < H < 1)$ si para todo $a > 0$ y $t \geq 0$

$$Y(t) \stackrel{d}{=} a^{-H} * Y(at) \quad \forall a > 0, \quad (2.34)$$

donde " d " indica igualdad en las distribuciones de orden finito.

A la propiedad definida por la ecuación (2.34) se le denomina propiedad de escalado, en otras palabras invarianza de la distribución a la escala, por consiguiente $Y(at)$ es la versión escala en el tiempo de $Y(t)$. Normalizado por el factor a^{-H} . Como a es variable el término H permanece constante. Una forma de visualización de esta definición es que un proceso es autosimilar si, cuando en el eje x (eje de tiempo) es ampliada por un factor

a el eje y es ampliada por un factor de a^{-H} . Entonces el proceso visto tiene la misma distribución estadística [14].

Los procesos de incremento estacionario son de especial interés en el campo de la modelación de tráfico ya que facilitan el manejo matemático. Consideremos el “proceso agregado” con nivel de agregación m , obtenido del proceso original X_k mediante la suma agregada:

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(t), \quad k = 1, 2, \dots \quad (2.35)$$

Esto es, se hace una partición de $\{X(t), t \in Z\}$ en bloques no traslapados de tamaño m y promediamos los valores de cada bloque. Así de este modo se puede dar una definición de autosimilitud para una serie estacionaria.

Una serie estacionaria X_k se denomina exactamente autosimilar si para todos los niveles de agregación m se tiene:

$$X_k = m^{1-H} X_k^m \quad (2.36)$$

Autosimilitud asintótica: También suele decirse que existe autosimilitud asintótica de una serie estacionaria si la serie X_k satisface:

$$X_k = m^{1-H} X_k^m \text{ cuando } m \rightarrow \infty \quad (2.37)$$

Al mismo tiempo en el caso de aplicaciones ingenieriles la autosimilitud puede ser mostrada como momentos de primero y segundo orden que tienen un significado físico. Al habilitar estos momentos se obtiene información relevante acerca de los procesos.

Dado un proceso autosimilar de parámetro H , con momentos de segundo orden finitos, la función de correlación del proceso incremento estacionario $X(t) = Y(t) - Y(t - 1)$ es:

$$\rho(k) = \frac{1}{2} \left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right] = g(k), k > 0 \quad (2.38)$$

Analizando los distintos casos según el valor de H se obtiene.

$H < 0$: Un proceso auto-similar con incrementos estacionarios es un proceso no medible para valores negativos de H .

$H = 1/2$: el proceso es incorrelacionado.

$1/2 < H < 1$: la función de auto-correlación, $\rho(k)$ se comporta asintóticamente de la forma:

$$\rho(k) \sim H(2H - 1)k^{2H-2} \quad (2.39)$$

Es decir posee una caída hiperbólica.

$0 < H < 1/2$: en este caso la correlaciones son negativas y sumables verificándose que:

$$\sum_{k=0}^{\infty} \rho(k) = 0 \quad (2.40)$$

$H = 1$: $\rho(k) = 1 \forall k$, caso que tampoco posee interés práctico.

$H > 1$: $\rho(k)$ diverge a infinito, contradiciendo el hecho de que su valor debe estar entre -1 y 1.

Definiciones similares del concepto de autosimilitud en función del parámetro β , donde $\beta = 2H - 2$, son dadas a continuación:

✓ *Autosimilitud de segundo orden*: una serie estacionaria en covarianza X_k , con varianza σ^2 , se dice exactamente autosimilar de segundo orden con parámetro

$H = 1 - \frac{\beta}{2}$ ($0 < \beta < 1$) si se verifica cualquiera de las tres condiciones equivalentes:

i. Si su función de autocorrelación es:

$$\rho(k) = \frac{1}{2} \left[(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta} \right] = g(k) \quad \forall k > 0 \quad (2.41)$$

ii. La varianza de la series agregadas es:

$$\text{Var}(X_k^m) = \sigma^2 m^{-\beta}, \quad \forall m > 1 \quad (2.42)$$

iii. Su densidad espectral es de la forma:

$$f(\lambda) = c |e^{i\lambda} - 1|^2 \sum_{j=-\infty}^{\infty} |2\pi j + \lambda|^{\beta-3}, \quad \lambda \in [-\pi, \pi] \quad (2.43)$$

Donde c es una constante de normalización, tal que $\int_{-\pi}^{\pi} f(\lambda) d\lambda = \sigma^2$

Además ello implica que la función de auto-correlación de las series agregadas es igual al de la serie original, es decir:

$$\rho^{(m)}(k) = \rho(k) = g(k), \quad \forall k > 0, \forall m > 1 \quad (2.44)$$

✓ *Autosimilitud asintótica de segundo orden:* Una serie estacionaria con covarianza X_k , se dice asintóticamente auto-similar de segundo orden con parámetro

$H = 1 - \frac{\beta}{2}$ ($0 < \beta < 1$) si se verifica:

$$\lim_{m \rightarrow \infty} \rho^{(m)}(k) = g(k), \quad \forall m > 1 \quad (2.45)$$

También se dan algunas definiciones de autosimilitud en función de la autocovarianza.

Bajo la suposición de estacionaridad de segundo orden se llega a las siguientes definiciones de autosimilitud.

✓ *Autosimilitud de segundo orden:* $X(t)$ es exactamente autosimilar de segundo orden con parámetro Hurst $H(1/2 < H < 1)$ si:

$$\gamma(k) = \frac{\sigma^2}{2} \left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right) \quad (2.46)$$

Para todo $k \geq 1$. $X(t)$ es asintóticamente autosimilar de segundo orden si

$$\lim_{m \rightarrow \infty} \gamma^{(m)}(k) = \frac{\sigma^2}{2} \left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right) \quad (2.47)$$

En las ecuaciones anteriores (2.46) y (2.47) se puede observar que la autosimilitud de segundo orden captura las propiedades de la estructura de correlación exacta y autosimilar. Aquí $\gamma^{(m)}(k)$ denota la función de autocovarianza de la serie agregada.

2.6.4 Dependencia de Largo Rango

La dependencia de largo rango es una propiedad de las series temporales que exhiben una fuerte dependencia sobre los retrasos temporales. Es decir, la dependencia de largo rango refleja intuitivamente el fenómeno de persistencia en un proceso auto-similar, además de la existencia de ráfagas en todas las escalas de tiempo.

La dependencia de largo rango puede ser formalmente definida de dos maneras esencialmente equivalentes, en el dominio del tiempo y de la frecuencia [26].

Definición 16: Un proceso $X(t)$ es llamado proceso estacionario con dependencia de largo rango si existe un número real $H \in (0.5, 1)$ y una constante $C_r > 0$ tal que

$$\lim_{k \rightarrow \infty} \frac{\rho(k)}{C_r K^{2H-2}} = 1 \quad (2.48)$$

Donde H es la medida del grado de la dependencia de largo rango.

Definición 17: Sea $X(t)$ un proceso estacionario si existen un número real $H \in (0.5, 1)$ y una constante $C_s > 0$ tal que

$$\lim_{w \rightarrow \infty} \frac{S(w)}{C_s |w|^{1-2H}} = 1 \quad (2.49)$$

Entonces X es llamado un proceso estacionario con dependencia de largo rango.

C_s y C_r se relacionan de tal forma que $C_s = C_r \text{Var}[X(t)]\Gamma(2H-1)\sin(\pi-\pi H)$ donde $\Gamma(\bullet)$ Denota la función gamma.

Aunque la “autosimilitud” y la “dependencia de largo rango” no sean conceptos equivalentes, se suelen emplear indistintamente dentro del contexto del tráfico en las redes de telecomunicaciones.

También se puede dar una asociación de los procesos de segundo orden y los conceptos de dependencia de largo rango de la siguiente manera [27]:

Sea $\rho(k) = \frac{\gamma(k)}{\sigma^2}$ la función de auto-correlación. Para $0 < H < 1$, $H \neq 1/2$ se cumple

$$\rho(k) \sim H(2H-1)k^{2H-2}, k \rightarrow \infty \quad (2.50)$$

En particular, si $1/2 < H < 1$, $\rho(k)$ se comporta asintóticamente como $ck^{-\beta}$ para $0 < \beta < 1$ donde $c > 0$ es una constante y $\beta = 2 - 2H$, y se tiene

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty \quad (2.51)$$

Se puede decir entonces que la función de autocorrelación decae lentamente, esta propiedad hace que no pueda ser sumable. Lo cual es esencial en el dominio de la frecuencia donde la densidad espectral está dada por:

$$T(\nu) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \rho(k) e^{ik\nu} \quad (2.52)$$

Por consiguiente se requiere que la función de autocorrelación sea sumable para satisfacer la propiedad:

$$T(\nu) \sim c|\nu|^{-\alpha}, \quad \nu \rightarrow 0 \quad (2.53)$$

Donde $c > 0$ es una constante $0 < \alpha = 2H - 1 < 1$, así $T(\nu)$ diverge alrededor del origen, lo cual significa que las mayores contribuciones se presentan en las proximidades al origen.

2.6.5 Distribución de Cola Pesada (Heavy-Tailed)

Existe una íntima relación entre las distribuciones de cola-pesada y dependencia de largo rango. El caso más típico puede ser introducido de la siguiente manera: Una variable aleatoria Z tiene una distribución de cola pesada si la probabilidad:

$$\Pr[Z > x] = cx^{-\alpha} \quad \text{Cuando } x \rightarrow \infty \quad (2.54)$$

Donde α es llamado el parámetro de forma y se encuentra en el intervalo $0 < \alpha < 2$ y c es una constante positiva. La propiedad para las distribuciones de cola pesada consiste en el hecho de que si $0 < \alpha < 2$ tienen varianza infinita, si $0 < \alpha \leq 1$ ellas también tienen media ilimitada, es de especial interés en las redes de telecomunicaciones el intervalo $1 < \alpha < 2$. La distribución de cola pesada más utilizada es la distribución de Pareto.

Distribución de Pareto La distribución de Pareto tiene una especial importancia dada la naturaleza del tráfico en Internet. Su función de densidad acumulativa viene dada por:

$$F(x) = \begin{cases} 1 - \left(\frac{\kappa}{x}\right)^\alpha & x \geq \kappa \\ 0, & x < \kappa \end{cases} \quad (2.55)$$

Adicionalmente las siguientes relaciones se dan:

La función de densidad de probabilidad (pdf).

$$\text{pdf: } f(x) = \frac{\alpha \kappa^\alpha}{x^{\alpha+1}}, x \geq \kappa, \alpha, \kappa > 0. \quad (2.56)$$

$$\text{media } \bar{X} = \begin{cases} \frac{\alpha \kappa}{\alpha - 1} & \alpha > 1 \\ \infty, & \alpha \leq 1 \end{cases} \quad (2.57)$$

$$\text{Varianza } Var(X) = \begin{cases} \frac{\alpha \kappa^2}{(\alpha - 2)^2 (\alpha - 2)}, & \alpha > 2 \\ \infty, & \alpha < 2 \end{cases} \quad (2.58)$$

el coeficiente cuadrado de variación $c_x^2 = \frac{1}{\alpha(\alpha - 2)}, \alpha > 2$

En el caso de la distribución de Pareto, α es llamado el parámetro de forma y κ hace referencia al parámetro de escala, se puede notar en la ecuación (2.55) que la distribución puede tener tanto la media infinita, como la varianza infinita dependiendo del valor del parámetro de forma α [15].

2.6.6 Autosimilitud en el Tráfico de Redes de Telecomunicaciones

La separación del tráfico agregado en componentes individuales representando el flujo en pares activos de computadores fuente-destino, se puede lograr por que ellos tienen una cabecera que registra información de cada paquete en las redes Ethernet. En el nivel

individual el tráfico entre pares fuente-destino se puede modelar por medio de fuentes ON/OFF o modelos de trenes de paquetes [3]. Se consideran procesos ON/OFF estrictamente alternados, donde periodos de “actividad” son seguidos de periodos de “inactividad”. En estudios hechos en redes activas se demuestra que la superposición de muchos trenes de paquetes en escalas de tiempo largo, se comportan de manera auto-similar. Este comportamiento fue observado en trazas de redes LAN (Ethernet) y WAN [1] [3]. Además las duraciones de los estados ON y OFF básicamente son de cola-pesada lo que lleva a que el comportamiento sea autosimilar.

2.6.7 Teorema Límite Para tráfico Agregado

Considerando primero una simple fuente ON/OFF y centrándose en una serie temporal binaria estacionaria $\{W(t), t \geq 0\}$ como generadora. $W(t) = 1$ es un paquete en el tiempo t , y $W(t) = 0$ significa que no hay paquetes. Viendo $W(t)$ como un factor de utilidad se tiene uno de utilidad en un periodo ON y cero de utilidad en un periodo OFF, las longitudes de los periodos ON y OFF son independientes y pueden tener diferentes distribuciones. Los periodos OFF siempre siguen a los periodos ON y este par de periodos se definen como periodo entre arribos. (Tales procesos algunas veces son referidos como procesos de restablecimiento alternativo).

Se supone ahora que se tienen M fuentes ON/OFF con las mismas características descritas anteriormente. Aquí cada fuente m envía su propia secuencia de trenes de paquetes y cada una tiene su propia secuencia útil $\{W^{(m)}(t), t \geq 0\}$. La superposición o cuenta acumulada de paquetes en el tiempo t , es:

$$\sum_{m=1}^M W^{(m)}(t) \quad (2.59)$$

Re-escalando por un factor de T, considerando

$$W^*_M(Tt) = \int_0^{Tt} \left(\sum_{m=1}^M W^{(m)}(u) \right) du, \quad (2.60)$$

Como la agregación de la cuenta de paquetes acumulados en el intervalo $[0, Tt]$, donde es de interés el comportamiento estadístico del proceso estocástico $\{W^*_M(Tt), t \geq 0\}$ para M y T grandes, luego se tiene que comportamiento depende de las distribuciones de los periodos ON/OFF.

Considerando las posibles distribuciones el modelo de medidas de tráfico de paquetes agregados acumulativos se puede aproximar al movimiento Fraccional Browniano. Este hecho es explicado en [1] [3].

2.7 MODELOS DE TRÁFICO

Los modelos de tráfico juegan un rol importante en la caracterización y análisis del tráfico real en las redes. Estos modelos son usados en la planeación y dimensionamiento para incrementar el rendimiento y la utilización de los recursos de la red. Como la meta es proveer servicios NGN, el énfasis de las investigaciones y los desarrollos apuntan a redes de alta velocidad suministrando integración de servicios con ciertas especificaciones de QoS (Quality off Service). Consecuentemente, el rol de los modelos de tráfico precisos en el diseño y simulación de redes llega a ser siempre crucial. Los modelos de tráfico precisos afinan la comprensión de redes complejas y el comportamiento de redes reales bajo estas condiciones complejas. Esto permite estudios en los efectos de diferentes parámetros del modelo y configuraciones en el rendimiento de la red a través de la simulación.

Las modernas redes de datos albergan diversos tipos de tráfico y un gran número de fuentes del mismo tipo, las cuales se incrementan rápidamente, el resultado es un sistema de tráfico complejo desde un punto de vista estadístico, siendo una tarea difícil encontrar modelos de tráfico precisos que recojan todas estas características. A través de décadas de investigaciones en el campo del modelamiento de tráfico en redes, pocos modelos fueron capaces de capturar la mayoría de las características del tráfico real. Otro factor que adiciona complejidad para el desempeño de los modelos de tráfico precisos, es el tamaño de las trazas. El tamaño de las trazas de tráfico tiene una influencia directa en el intervalo de confianza estadística [18].

La calidad de un modelo de tráfico esta determinado por dos factores: rendimiento y la eficiencia computacional. Además se debe tener en cuenta la habilidad del modelo para capturar las propiedades estadísticas del tráfico en las redes. Los modelos de tráfico tradicionales tales como: Poisson, Markov Y ARMA (Auto Regressive Moving Average) son simples desde un punto de vista computacional, pero su desempeño representando tráfico en redes en general no es adecuado [21]. Últimamente modelos que tienen en cuenta el concepto de autosimilitud dan mucha mejor aproximación, aunque los parámetros del modelo son difíciles de determinar. A pesar de esto los modelos auto-similares son considerados en actuales investigaciones dado el inadecuado desempeño de los modelos tradicionales [18].

2.7.1 Procesos modulados de Markov (MMPP)

Los procesos modulados de Markov, son tal vez la clase de modelos de tráfico más estudiada. En estos procesos existe un proceso de Markov subyacente que evoluciona en el tiempo y cuyo estado actual modula las características estadísticas de las llegadas.

En cuanto a la cadena de Markov subyacente, existe un número determinado de estados y las llegadas ocurren siguiendo una determinada regla, cuando un proceso de Markov conmuta para un nuevo estado las llegadas ocurren de acuerdo con una regla definida por el nuevo estado [28]

Tal vez una de las más importantes clases de modelos de tráfico sean los procesos de Poisson modulados por Markov, también llamados MMPP, siendo este uno de los modelos de tráfico tratados en este trabajo. Las llegadas ocurren, en cada estado, siguiendo un proceso de Poisson con una tasa que puede ser diferente para cada estado de la cadena de Markov que lo modula. Un MMPP está completamente descrito por dos matrices, una que define las tasas de llegadas a los estados y la otra, en el caso de un MMPP en tiempo discreto es una matriz de probabilidades de transición, y en el caso de un MMPP en tiempo continuo es una matriz de tasas de transición entre estados. Como se dijo anteriormente la matriz de tasas en tiempo continuo, es fácilmente obtenida a partir de la matriz de probabilidades de transición para el caso discreto.

Procesos de Poisson modulados por Markov en tiempo discreto (dMMPP): Un proceso de Poisson modulado por Markov en tiempo discreto puede ser visto como un proceso aleatorio Markoviano cuyos incrementos en cada instante tienen una distribución de Poisson con una media que depende del estado en el que se encuentre la cadena de Markov moduladora. Más exactamente, una cadena de Markov (homogénea) $(X, J) = \{(X_k, J_k), k = 0, 1, \dots\}$ es un dMMPP si y solo si para $k = 0, 1, \dots$,

$$P(X_{k+1} = m, J_{k+1} = j | X_k = n, J_k = i) = \begin{cases} 0, & m < n \\ p_{ij} e^{-\lambda_i} \frac{\lambda_i^{m-n}}{(m-n)!}, & m \geq n \end{cases} \quad (2.61)$$

Aquí S es el número de estados de la cadena, X representa el proceso Modulado (en este caso, el proceso de Poisson) y J la cadena de Markov Moduladora. Para todo $m, n \in \mathbb{N}$; $i, j \in S$; en el que λ_i , con $i \in S$, es real constante y no negativo, y $P = [p_{ij}]$ es una matriz estocástica e irreducible. Cabe mencionar que la distribución de $X_{k+1} - X_k$ dado que $J_k = j$ es de Poisson con media λ_j , siendo que λ_j representa el incremento medio del proceso X cuando la cadena de Markov moduladora esta en el estado j [28].

Siempre que (2.61) se verifique, se dice que: (X, J) es un dMMPP con un conjunto de estados moduladores S y con parámetros (matrices) P y Λ , y puede ser escrita como

$$(X, J) \sim dMMPP_s(P, \Lambda), \tag{2.62}$$

donde la matriz $\Lambda = [\lambda_{ij}] = [\lambda_i \delta_{ij}]$, $\delta_{ij} = 1$ cuando $i = j$ y $\delta_{ij} = 0$ cuando $i \neq j$. La matriz P es la matriz de probabilidades de transición entre los estados de la cadena de Markov moduladora J , y Λ es la matriz de tasas de llegadas del proceso de Poisson modulado. Si S esta constituido por n estados, se dice que (X, J) es un dMMPP de orden n (n -dMMPP). En particular, si $S = \{1, 2, \dots, n\}$ para algún $n \in \mathbb{N}$, entonces [28].

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & P_{n3} & \cdots & P_{nn} \end{bmatrix} \text{ y } \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{2.63}$$

y se escribe simplemente $(X, J) \sim dMMPP_n(P, \Lambda)$.

La distribución estacionaria de J esta designada por $\pi = [\pi_1, \pi_2, \dots, \pi_n]$, y pueden verificarse las siguientes condiciones: [28].

$$e(P - I) = 0 \tag{2.64}$$

$$\sum_i \pi_i = 1 \tag{2.65}$$

donde I designa la matriz identidad de dimensión $n \times n$, y e es un vector columna de longitud n con todos los elementos iguales a 1.

Una característica interesante de los MMPP es que una superposición de MMPPs es así mismo un MMPP. En la figura 7 se puede observar un diagrama representativo de un MMPP de dos estados [28].

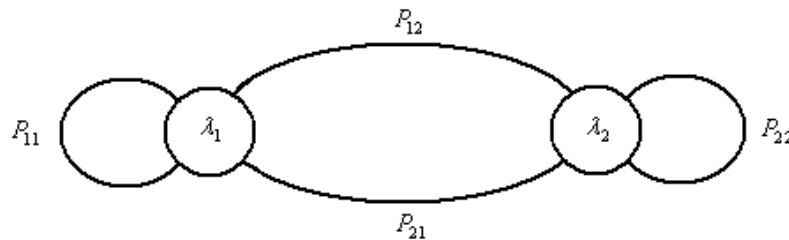


Figura 7. Diagrama de estados de un 2-MMPP [18].

2.7.2 Movimiento Browniano Fraccional

Antes de entrar a hablar del movimiento Browniano fraccional y de autosimilitud, se describen dos conceptos importantes que sirven para situarse dentro del contexto que se esta tratando. Los conceptos a estudiar son Estacionaridad y Ergodicidad.

2.7.2.1 Estacionaridad y Ergodicidad: El análisis de series temporales estocásticas depende de la estacionaridad de los procesos relacionados. Sin embargo, la

estacionaridad implica que las relaciones de los procesos de tráfico sean ergódicos. Ambos conceptos son principalmente estudiados para series temporales estocásticas.

Un proceso es estacionario, si sus propiedades estadísticas del proceso no cambian con el tiempo. Esto significa que el sistema llega a equilibrarse en el sentido de que mirando las relaciones de los procesos en un tiempo particular tiene las mismas relaciones para un proceso en el cual esta desplazado en el tiempo de la medición original. Aquí se presentan dos tipos de estacionaridad: “estacionaridad completa” y “estacionaridad de orden superior m ”. Para efectos prácticos la siguiente definición es usualmente la más usada [26] [29].

Estacionaridad completa: Un proceso estocástico $X(t)$ se dice estacionario si para cualquier t_1, t_2, \dots, t_n y cualquier k , la distribución de probabilidad conjunta de $X(t_1), X(t_2), \dots, X(t_n)$ es idéntica a la distribución de probabilidad conjunta de $X(t_1 + k), X(t_2 + k), \dots, X(t_n + k)$.

$$F_{X(t_1), \dots, X(t_n)}(X_1, \dots, X_n) = F_{X(t_1+k), \dots, X(t_n+k)}(X_1, \dots, X_n) \quad (2.66)$$

Estacionaridad de orden superior m : Un proceso estocástico $X(t)$ se dice que es estacionario de orden superior m si para cualquier t_1, t_2, \dots, t_n y cualquier k , los momentos conjuntos de orden superior m de $X(t_1), X(t_2), \dots, X(t_n)$ existen y son iguales a los correspondientes momentos conjuntos de orden superior m $X(t_1 + k), X(t_2 + k), \dots, X(t_n + k)$, es decir:

$$E[\{X(t_1)\}^{m_1} \{X(t_2)\}^{m_2} \dots \{X(t_n)\}^{m_n}] = E[\{X(t_1 + k)\}^{m_1} \{X(t_2 + k)\}^{m_2} \dots \{X(t_n + k)\}^{m_n}] \quad (2.67)$$

La ecuación anterior expone el estado de todos los momentos de orden superior m , los cuales son independientes del tiempo. Un caso particular de orden superior es cuando $m = 2$. Esto implica que la media y la varianza son independientes del tiempo.

1. La media del proceso $E[X(t)] = \mu$ es una constante.
2. El segundo momento del proceso $E[X^2(t)] = \mu_2$ también una constante.
3. Por lo tanto la varianza $Var(X(t)) = \mu_2 - \mu^2 = \sigma^2$ es también independiente del tiempo. Además se puede realizar la simplificación de las estructuras de covarianza de los procesos que incluyen la medida de tiempo únicamente.
4. $cov\{X(t)X(s)\} \equiv E[X(t)X(s)] - \mu^2$ porque la media es independiente del tiempo, entonces la única dependencia de tiempo que existe en la estructura es dada por t y s .

La ergodicidad es un concepto importante dentro del análisis de procesos estocásticos, en forma general, cuando todas las propiedades estadísticas de un proceso estocástico $X(t)$ se pueden determinar con probabilidad uno a partir de las medias temporales de una única realización, se dice que un proceso es ergódico. Aquí se necesita que el proceso sea estacionario. Por lo general en la mayor parte de los procesos solo se dispone de una señal muestra a la cual se le puede hallar la media temporal y a partir de esta media calcular las propiedades estadísticas del proceso estocástico [26] [29].

2.7.2.2 Modelos estocásticos auto-similares Dentro de los modelos estocásticos más comunes se encuentran el Movimiento Browniano Fraccional (Fractional Brownian Motion- FBM) introducido por Mandelbrot [30] y el Ruido Gaussiano Fraccional (Fractional Gaussian Noise- FGN). FBM es un proceso estocástico no estacionario que fue

desarrollado como una generalización del Movimiento Browniano. FGN esta relacionado con FBM, dado que FGN es la versión diferenciada (es decir el proceso incremento correspondiente).

El Movimiento Browniano Fraccional es un tipo de función aleatoria basada en el Movimiento Browniano estándar. Si $B(t)$ es un movimiento Browniano estándar y H es algún parámetro que esta en el rango $0 < H < 1$ entonces FBM de exponente H puede definirse como

$$B_H(t) = \frac{1}{\Gamma(H - 1/2)} \int_{-\infty}^t k(t-s) dB(s), \quad t > 0 \quad (2.68)$$

Donde la función $K(t-s)$ es

$$k(t-s) = \begin{cases} (t-s)^{(H-1/2)} - (-s)^{(H-1/2)} & \text{si } s < 0 \\ (t-s)^{(H-1/2)} & \text{si } 0 \leq s \leq t \end{cases} \quad (2.69)$$

La representación anterior fue propuesta por Mandelbrot utilizando una función con caída hiperbólica [30].

Movimiento Browniano Fraccional: El movimiento Browniano fraccional con exponente Hurst $0 < H < 1$ es un proceso estocástico gaussiano con media cero tal que:

1. $B_H(0) = 0$
2. $B_H(t) =_d N(0, \sigma |t|^H)$
3. $B_H(t+\delta) - B_H(t) =_d N(0, \sigma |\delta|^H)$

De esto se puede decir: que es un proceso estocástico de distribución normal con media cero y tiene una varianza escalada exponencialmente con H .

Ruido Gaussiano Fraccional: El ruido Gaussiano fraccional es uno de los modelos que despliega dependencia de largo rango y la mejor vía para introducir el ruido gaussiano fraccional es a través de su “padre” el movimiento browniano fraccional $\{B_H(t), t \geq 0\}$, que en otras palabras se puede describir como un proceso Gaussiano con media 0, incrementos estacionarios, varianza $E[B_H^2(t)] = t^{2H}$ y covarianza

$$E[B_H(s)B_H(t)] = \frac{1}{2} \{s^{2H} + t^{2H} - |s-t|^{2H}\} \quad (2.70)$$

El índice crucial es H , su valor se encuentra en el intervalo (0,1), y generalmente es llamado parámetro de autosimilitud.

El Ruido gaussiano fraccional $\{X_i, i \geq 1\}$ es un proceso de incremento del movimiento browniano fraccional, particularmente

$$X_i = B_H(i+1) - B_H(i), i \geq 1 \quad (2.71)$$

Con media cero, serie de tiempo gaussiana estacionaria cuya función de autocovarianza

Esta dada por $\gamma(k) = \frac{\sigma^2}{2} \left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right)$ donde $k \geq 0$

Los X_i 's están positivamente correlacionados cuando $0.5 < H < 1$ y se dice que ellos despliegan dependencia de largo rango o memoria larga. En este contexto el índice H mide la intensidad de dependencia de largo rango. La densidad espectral (Transformada de Fourier de γ) es [31].

$$f(\lambda) = C_H (2 \sin \frac{\lambda}{2})^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\lambda + 2\pi k|} \sim C_H |\lambda|^{1-2H} \quad \text{Para } \lambda \rightarrow 0. \quad (2.72)$$

Donde C_H es una constante.

En el campo de las redes en [32] se introduce un proceso estocástico $V(t)$ que es un modelo que satisface el almacenamiento de una entrada auto-similar que puede ser desocupado como una velocidad constante. Norros lo define de la siguiente manera:

Definición 18: "El modelo de almacenamiento estacionario el cual es browniano fraccional en la entrada de la red es un proceso estocástico $V(t)$, donde:

$$V(t) = \sup_{s \leq t} (A(t) - A(s) - C(t - s)), t \in (-\infty, \infty) \quad (2.73)$$

$A(t)$ es el proceso dado por:

$$A(t) = mt + \sqrt{am}Z(t), t \in (-\infty, \infty) \quad (2.74)$$

Donde $Z(t)$ es un movimiento browniano fraccional normalizado, el sistema tiene cuatro parámetros m , a , H y C los cuales tienen la siguiente interpretación: $m > 0$ es la velocidad media, $a > 0$ es el coeficiente de la varianza, $H \in [\frac{1}{2}, 1)$ es el parámetro de autosimilitud para $Z(t)$, y $C > m$ es la velocidad de servicio" [32].

3 MODELAMIENTO E IMPLEMENTACIÓN DE LA SIMULACIÓN

3.1 INTRODUCCIÓN

En el estudio y caracterización del comportamiento de tráfico de voz sobre IP en redes NGN (redes de próxima generación), es de especial interés realizar simulaciones de los modelos de tráfico utilizados, con el fin de constatar cual de estos modelos se aproxima con mayor exactitud a las características que presentan los paquetes de voz en las redes actuales y de próxima generación (autosimilitud y LRD) en donde la calidad de servicio es la meta para satisfacer a los usuarios.

Pensando en la simulación de estos modelos, la literatura actual proporciona tanto métodos exactos, como métodos aproximados. Debido a que la generación de tráfico auto-similar por medio de métodos exactos es computacionalmente muy costosa, se utilizan métodos aproximados para la generación de muestras que almacenan varias propiedades (entre las más importantes esta la dependencia de largo rango (LRD) y la autosimilitud), donde es de especial interés el caso $H > 1/2$.

A continuación se realiza una breve explicación de los métodos utilizados en este trabajo para generar secuencias tanto de FBM y MMPP.

3.2 MOVIMIENTO BROWNIANO FRACCIONAL

Haciendo referencia a la creación de algoritmos que respondan al comportamiento matemático del FBM (Fractional Brownian Motion) y FGN (Fractional Gaussian Noise) en este trabajo se usan el método de Cholesky [33] y el de Paxson [34].

3.2.1 Método Cholesky (Exacto)

Como su nombre lo indica el método de Cholesky hace uso de la descomposición de cholesky de la matriz de covarianza. Según esto la matriz de covarianza $T(n)$ se puede representar por la combinación de dos matrices $L(n) \times L(n)'$ ⁴, donde $L(n)$ es una matriz triangular inferior de $(n+1) \times (n+1)$. Denotando el elemento (i, j) para $L(n)$ por $l_{i,j}$, para $i, j = 0, \dots, n$, $L(n)$ se dice que es una matriz triangular inferior si $l_{i,j} = 0$ para $j > i$. Los elementos de $L(n)$ pueden ser calculados observando que el elemento (i, j) para $L(n)L(n)'$ y $T(n)$ debe ser igual para $j \leq i$ esto es:

$$\gamma(i - j) = \sum_{k=0}^k l_{ik} l_{jk}, j \leq i. \quad (3.1)$$

Esta ecuación se reduce a $\gamma(0) = l^2_{00}$ para $i = j = 0$ para $i = 1$ se obtienen dos ecuaciones

$$\gamma(1) = l_{10} l_{00} \quad (3.2)$$

$$\gamma(0) = l^2_{10} + l^2_{11} \quad (3.3)$$

Se puede observar claramente que $l_{i,j}$ no depende de n , y $L(n+1)$ puede ser calculada de $L(n)$, adicionándole una fila determinada por:

⁴ Producto cruz.

$$l_{n+1,0} = \frac{\gamma(n+1)}{l_{00}} \quad (3.4)$$

$$l_{n+1,j} = \frac{1}{l_{jj}} \left(\gamma(n+1-j) - \sum_{k=0}^{j-1} l_{n+1,k} l_{jk} \right), 0 < j \leq n \quad (3.5)$$

$$l^2_{n+1,n+1} = l^2_{00} - \sum_{k=0}^n l^2_{n+1,k}. \quad (3.6)$$

De las fórmulas anteriores se dice que $L(n)$ es única con la restricción adicional de que los elementos de la diagonal principal son estrictamente positivos. Donde $T(n)$ es una matriz exacta positiva.

En este método se halla un vector $X(n)$ de manera recursiva. Este vector es el resultado de multiplicar la matriz $L(n)$ por un vector $V(n)$ distribuido normalmente con media cero y varianza 1, es decir $X(n) = L(n)V(n)$. El vector resultante $X(n)$ debe tener media cero y la matriz de covarianza ser igual a:

$$Cov(L(n)V(n)) = E[(L(n)V(n)) (L(n)V(n))^T] = T(n) \quad (3.7).$$

La desventaja de este método y de la mayoría de los métodos exactos, se debe a que con ellos no se pueden generar gran cantidad de muestras, puesto que el consumo de recursos computacionales es grande, así como también el tiempo empleado para dar una respuesta determinada.

3.2.1 Método de Transformada de Fourier (Aproximado)

La base conceptual de este método se relaciona con el uso de la transformada rápida de Fourier para la creación de muestras sintéticas, las cuales deben responder a unas determinadas condiciones planteadas por Paxson [34]. Una visión general del procedimiento empleado se da a continuación.

Lo primero que se debe tener en cuenta, es que se necesita conocer el espectro de potencia del modelo que se quiere implementar. En este caso, concretamente se necesitan generar muestras que correspondan al Ruido Gaussiano Fraccional (FGN) donde el espectro de potencia viene dado por:

$$f(\lambda; H) = A(\lambda; H)[|\lambda|^{-2H-1} + B(\lambda; H)] \quad (3.8)$$

Para $0 < H < 1$ y $-\pi \leq \lambda \leq \pi$ donde:

$$A(\lambda; H) = 2\sin(\pi H)\Gamma(2H + 1)(1 - \cos\lambda) \quad (3.9)$$

$$B(\lambda; H) = \sum_{j=1}^{\infty} [(2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1}] \quad (3.10)$$

Aquí $f(\lambda; H)$ esta conformada por el parámetro $A(\lambda; H)$ el cual no tiene mayor complicación por ser el resultado de funciones finitas, mientras que el parámetro $B(\lambda; H)$ presenta inconvenientes para su calculo debido a que esta constituido por la sumatoria de infinitos componentes. Para subsanar esto Paxson utiliza la siguiente aproximación.

$$B(\lambda; H) \approx a_1^d + b_1^d + a_2^d + b_2^d + a_3^d + b_4^d + \frac{a_3^{d'} + b_3^{d'} + a_4^{d'} + b_4^{d'}}{8H\pi} \quad (3.11)$$

Donde:

$$d = -2H - 1$$

$$d' = -2H$$

$$a_k = 2k\pi + \lambda$$

$$b_k = 2k\pi - \lambda.$$

Con la aproximación dada por la ecuación (3.11), se conoce en su totalidad el espectro de potencia para el Ruido Gaussiano Fraccional. Observando la fórmula (3.8) los parámetros que se necesitan de entrada son la cantidad de muestras a generar y el valor del parámetro H (hurst) con ellos se obtienen componentes reales de señal $f(\lambda; H)$,

distribuidos asintóticamente de manera exponencial aleatoria con media equivalente a la potencial actual, a estos se les calcula la transformada de Fourier que se caracteriza por ser una serie de valores complejos, donde la fase se distribuye de manera uniforme entre 0 y 2π , posee simetría con respecto a su valor central, es decir, la mitad de los valores son los complejos conjugados de la otra mitad. Para la obtención de secuencia FGN en el dominio del tiempo se aplica la transformada de Fourier inversa.

Los métodos expuestos anteriormente, los cuales fueron implementados en matlab están basados en códigos existentes en el lenguaje de programación S-PLUS⁵.

3.3 PROCESOS DE POISSON MODULADOS POR MARKOV (MMPP).

En este modelo, el estado de la fuente cambia de acuerdo con la cadena de Markov continua (CTMC) homogénea multiestados. En cada estado de la fuente, el proceso de arribos es de Poisson con posibles tasas de llegadas dependientes del estado, es decir, los tiempos entre llegadas están distribuidos exponencialmente y la tasa a la cual llegan depende del estado en que se encuentre la cadena. De esa manera la CTMC modula el proceso de Poisson el cual introduce correlación entre los sucesivos tiempos entre llegadas. Recordar que para un proceso de arribos de Poisson ordinario los tiempos entre llegadas son independientes. El MMPP es ampliamente usado para modelar varias fuentes ISDN (Integrated Services Digital Network), tales como voz y video y también para caracterizar tráfico superpuesto [15]. En la figura 8 se presenta el diagrama de un MMPP de dos estados y en la figura 9 un ejemplo de cómo se pueden generar los arribos con los procesos de Poisson.

⁵ Disponibles en <http://ita.ee.lbl.gov/html/whats-new.html>.

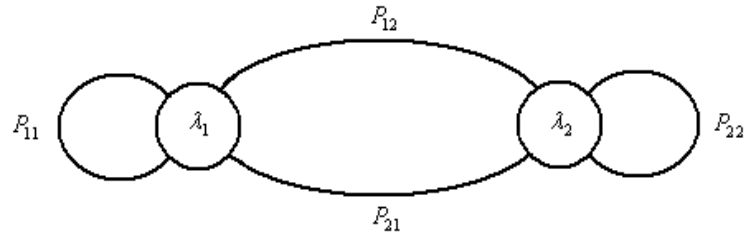
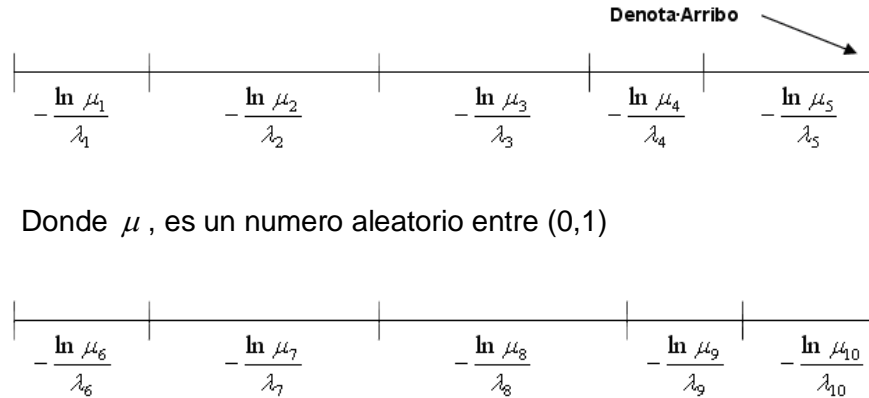


Figura 8. Diagrama de estados de un 2-MMPP [28].



Donde μ , es un numero aleatorio entre (0,1)

Figura 9. Representación gráfica de los arribos en un MMPP de 2 estados [16].

Generando eventos de un proceso de llegadas: Frecuentemente en las simulaciones, se necesita generar una secuencia de puntos aleatorios en el tiempo $\{T_i\}$ donde $i = 1, 2, 3, \dots$ para diferentes tipos de procesos de arribos estocásticos. Siendo $N(t) = \max\{i : T_i \leq t\}$ el número de arribos y $W_i = T_i - T_{i-1}$ será el tiempo entre llegadas entre el i -ésimo y el $(i - 1)$ arribo.

Proceso de Poisson: En un sistema no ranurado, un arribo ocurre en un intervalo $(t, t + dt)$ con probabilidad (λdt) . En este caso los tiempos entre llegadas $\{W_i\}$ son conocidos por estar distribuidos exponencialmente con una tasa λ y una media $(1/\lambda)$. Así, t_i puede ser generado recursivamente de la siguiente manera.

- ✓ Fijar $t_0 = 0, i = 0$

- ✓ Incrementar $i = i + 1$.
- ✓ Generar μ_i , número distribuido uniformemente entre 0 y 1 ($U(0, 1)$).
- ✓ $t_i = t_{i-1} + \left(-\frac{1}{\lambda} \ln \mu_i\right)$

Generación de variables aleatorias: La buena elección de un generador de números aleatorios es importante para la exactitud de la simulación. $U(0, 1)$ es un número aleatorio entre 0 y 1, generado con la función *rand* en Matlab.

Transformada inversa: Con este método las siguientes propiedades son utilizadas: si X , es una variable aleatoria con CDF (Función de distribución acumulativa) F , entonces la nueva variable aleatoria $Y = F(X)$ esta uniformemente distribuida en el intervalo $[0, 1]$. Así, para generar una variable aleatoria x de X primero un número aleatorio μ distribuido uniformemente en el intervalo $[0, 1]$ es generado y entonces la función F es invertida. De esta manera, $F^{-1}(\mu)$ da el valor requerido de x . La idea principal de este método es mostrada en la figura 10.

Este método puede ser utilizado para generar la distribución exponencial, la distribución uniforme, la distribución Weibull, la distribución de Pareto, así como varias distribuciones empíricas y discretas. Su mayor utilidad se da cuando la inversa de la CDF, $F(\cdot)$, puede ser fácilmente calculada.

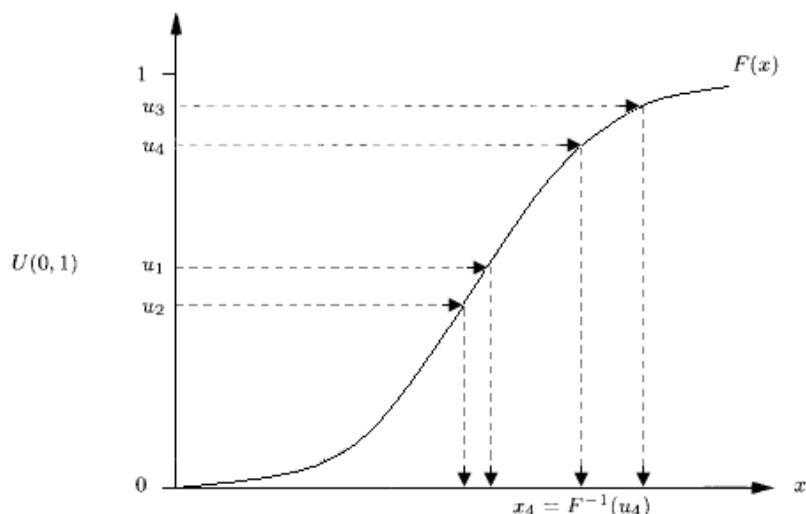


Figura 10. Método de la transformada inversa.

Distribución exponencial:

$$F(x) = 1 - e^{-\lambda x}, \quad (3.14)$$

Determinando μ a partir de $U(0,1)$, se genera x a partir de la distribución exponencial usando la fórmula

$$x = -\frac{\ln(1 - \mu)}{\lambda}. \quad (3.15)$$

Notar que $(1 - \mu)$, es también un número aleatorio, por lo tanto la anterior fórmula se puede simplificar a:

$$x = -\frac{\ln(\mu)}{\lambda} \quad (3.16)$$

A fin de verificar que la ecuación (3.16) provee una variable aleatoria de la distribución exponencial, se condujo un experimento. Se generaran repetidamente números aleatorios y correspondientemente se generan variables según la anterior fórmula. Tres casos se toman con 10, 100 y 1000 variables aleatorias e histogramas de estas observaciones y se grafican en la figura 11. Se superpone la función de distribución de probabilidad (pdf) de la distribución exponencial en cada caso. De las figuras es claro que incrementando el

número de variables aleatorias en las muestras el error cuadrático medio entre la función de distribución de probabilidad (pdf) teórica y las gráficas empíricas es menor.

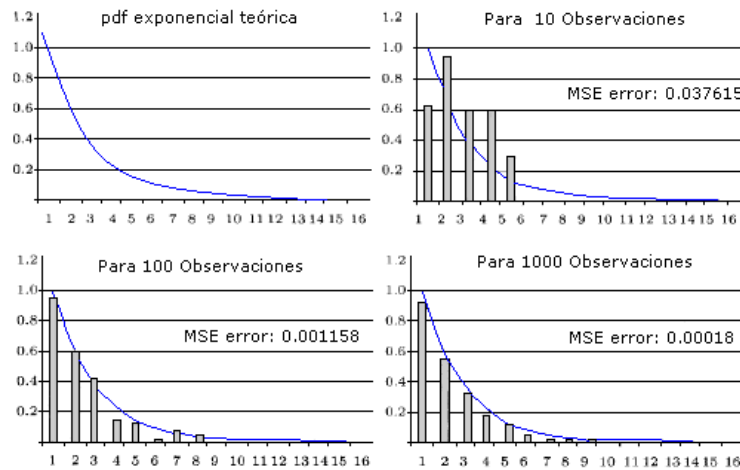


Figura 11. Pdf de una distribución exponencial e histogramas de una variable aleatoria.

Simulación proceso de Poisson modulado por Markov de dos estados (Markov Modulated Poisson Process two status - 2-MMPP). La simulación de los procesos de Poisson modulados por Markov de dos estados se realizó para el proceso de arribo de paquetes y para el tiempo entre llegadas por separado.

Teniendo en cuenta lo planteado en el capítulo 2, donde se establece que los MMPP son procesos de Poisson, en los cuales, la tasa de arribo depende del estado en el que se encuentra la cadena de Markov. Luego, para realizar la simulación se procede de la siguiente manera:

Para el *Tiempo entre llegadas*. (Ver figura 12)

- ✓ Inicialización de las variables a utilizar
- ✓ Mediante el proceso antes explicado se genera una variable aleatoria, la cual permite establecer en cual de los estados inicia la cadena.

- ✓ Se evalúa el valor de la variable estado, si estado es igual a uno, entonces se genera una variable aleatoria exponencial e_1 con tasa λ_1 , sino se genera una variable exponencial e_2 con tasa P_{12} .

Se evalúa la condición ($e_1 < e_2$ - Tiempo para que ocurra un arribo < tiempo para que ocurra una transición o cambio de estado), si es verdadera fijamos el próximo tiempo entre llegadas que será el nuevo e_1 .

Si es falso se cambia de estado.

- ✓ Cuando se produce el cambio de estado, se genera una variable exponencial e_3 con tasa λ_2 , y una variable exponencial e_4 con tasa P_{21} .

Se evalúa la condición ($e_3 < e_4$) si es verdadera se fija el próximo tiempo entre llegadas que será el nuevo e_1 .

Si es falso se cambia de estado.

Producir un arribo

- ✓ Para producir un arribo se ajusta los tiempos (tLlegada, transición y tFin Intervalo) usando el tiempo de llegada.
- ✓ Después se genera una variable aleatoria con el valor λ para el estado actual.
- ✓ Y por ultimo se incrementa la bandera arribos, la cual permite controlar los datos a graficar.

Producir Transición.

- ✓ Al igual que para los arribos se hace un ajuste de los tiempos (variables) pero con el tiempo de transición.

- ✓ Se realiza el cambio de estado en la cadena (se varia el valor de una bandera estado).
- ✓ Por ultimo se da un valor aleatorio al tiempo de transición.

Terminar intervalo.

- ✓ Se ajusta las variables de tiempo.
- ✓ Se incrementa una bandera (intervalos) la cual permite simular un determinado número de intervalos.

Se crea un vector con el número de arribos en cada intervalo.

Este proceso se ejecuta hasta que se complete el número de arribos que le indique el usuario. Como parámetros de entrada, necesita el número de arribos, las tasas de llegada λ_1, λ_2 , las probabilidades P_{12} y P_{21} . Estos parámetros se obtienen de las muestras de tráfico real de VoIP.

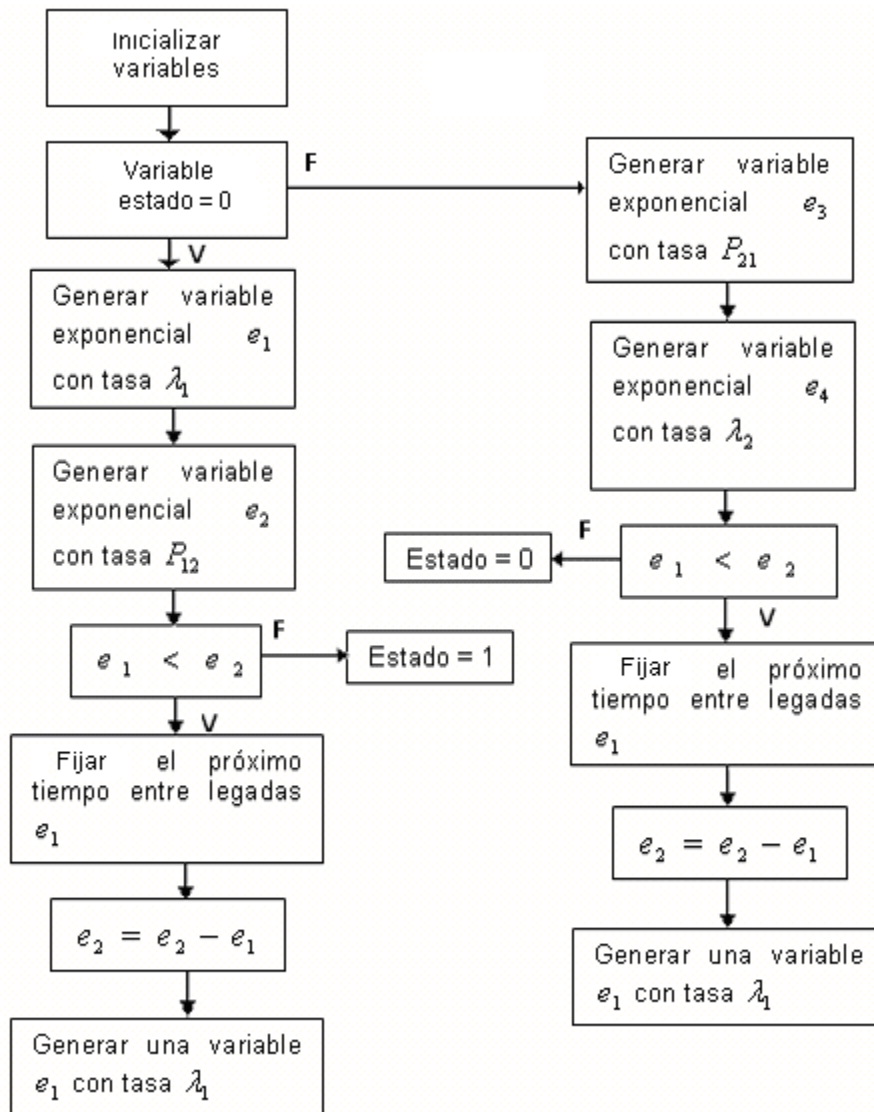


Figura 12. Diagrama en Bloques: Generar los tiempos entre llegadas.

Número de arribos de paquetes: para conocer el número de arribo de paquetes, se implemento un proceso de conteo para el tiempo entre llegadas (Ver figura 13). En el capítulo 4 se presentaran gráficas del tráfico sintético obtenido con estos dos métodos.

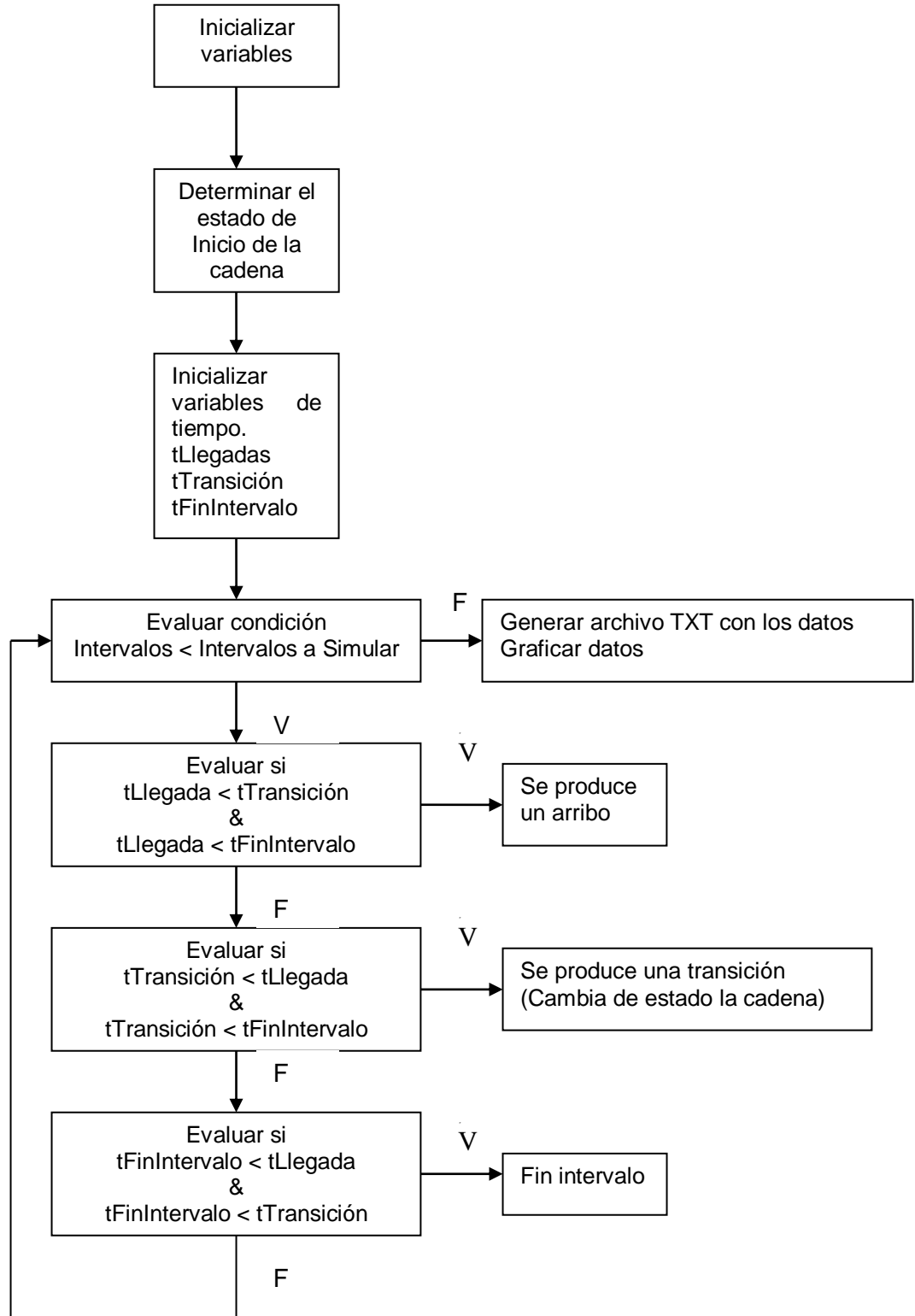


Figura 13. Diagrama en Bloques: Generar número de arribos.

4 PRUEBAS Y RESULTADOS

En este capítulo se realizan diferentes pruebas analíticas y experimentales con el fin de establecer cual de los dos modelos propuestos es el que mejor recoge las características del tráfico de la voz en redes IP de próxima generación. Inicialmente se extraen las principales características del archivo de datos de voz IP de una red real, los cuales dan una visión del comportamiento de los paquetes a nivel de transporte, para la comprobación de estas características se utilizan una serie de herramientas que permiten aproximar la función de densidad acumulativa de varias funciones matemáticas a la función de distribución acumulativa empírica de los datos reales, todo esto con el fin de observar gráficamente cual de estas funciones matemáticas se acerca más a los datos reales, para luego realizar una comprobación por medio del método de cuantiles.

El segundo paso dentro de este proceso es identificar el fenómeno de autosimilitud en los datos reales, la comprobación del grado de autosimilitud se realiza por medio de los métodos varianza-tiempo y periodograma.

Como un tercer paso se utilizan los modelos planteados en el capítulo tres, los cuales son los generadores de tráfico sintético tanto de los Procesos de Poisson Modulados por Markov (MMPP), como de Movimiento Fraccional Browniano (FBM) y su derivado Ruido Fraccional Gausiano (FGN). En este punto, teniendo en cuenta los datos obtenidos en la primera parte y las características de los métodos simulados, se escogerá el método que más se asemeje a los datos reales. A las muestras obtenidas por medio de estos métodos y a las muestras reales se les calcula el grado de autosimilitud por medio del

parámetro de Hurst. El valor de este parámetro se consigue a través de los métodos anteriormente descritos, los cuales son: el método de Varianza-Tiempo y el periodograma.

Análisis de Trazas de Tráfico de Voz IP en el contexto matemático existen una serie de distribuciones por medio de las cuales se pueden modelar el comportamiento de una variable aleatoria, en cada una de las distribuciones se tiene una serie de parámetros que las identifica de manera unívoca. De gran importancia dentro de este ámbito, es poder definir la función de distribución acumulativa o la función de densidad de probabilidad de la variable aleatoria comprometida en un experimento determinado, ya que estas dos funciones son la vía para la caracterización de una determinada distribución o modelo que caracteriza la v.a (variable aleatoria).

Los archivos de tráfico real⁶ de voz IP contienen la recopilación de una serie de muestras, uno de los eventos consignados dentro de los archivos es el tiempo entre llegadas de los paquetes de voz a un enrutador. En principio para la identificación de las características, se van a estudiar y analizar estos datos para su posterior aproximación a la distribución que mejor represente su comportamiento.

Los datos de tráfico de la voz se pueden representar por medio de una variable aleatoria que esta en función del tiempo (proceso estocástico), y su caracterización como se expreso anteriormente se puede hacer mediante su distribución y su densidad de probabilidad, si existe un conocimiento a priori de la función de probabilidad que da lugar al espacio probabilístico se puede hallar la función de distribución acumulativa de la variable aleatoria, dado que en este caso no se conoce la cdf (función de densidad acumulativa) se hace uso de la función de distribución empírica, para la cual se tienen que

⁶ Tomados de <http://tstat.tlc.polito.it/traces.shtml>

tomar un gran número de valores, para que la curva empírica se aproxime bastante a la función de distribución, entre más valores se tengan en cuenta para el cálculo de la función de distribución empírica mejor será su aproximación a la función de distribución acumulativa. A continuación se presentan en las figuras 14 y 15, las gráficas de funciones de distribuciones empíricas para diferentes tamaños de muestras (N=10000, N=20000, N=36000 y N=100000).

La figura 14 permite evidenciar el comportamiento y las variaciones relacionadas con la escogencia de la cantidad de arribos de paquetes de voz dentro de una red IP, la figura 15 muestra la ampliación de un sector de la figura 14, la cual permite observar mejor las gráficas para los diferentes tamaños de muestra.

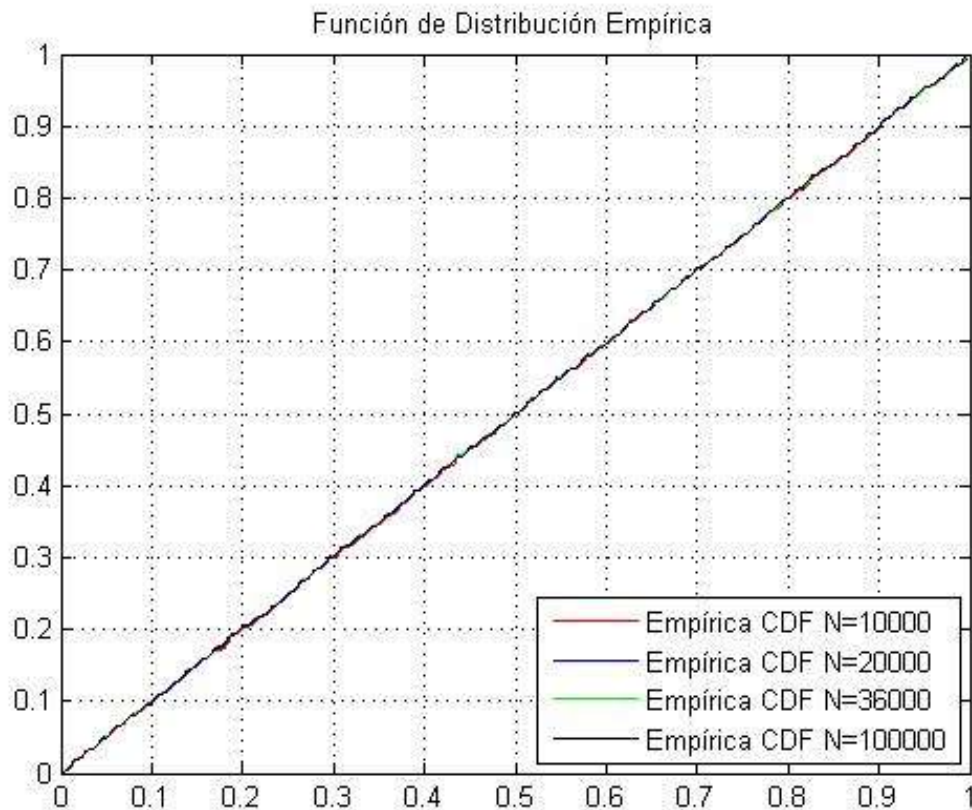


Figura 14. Función de distribución para diferentes tamaños de muestras.

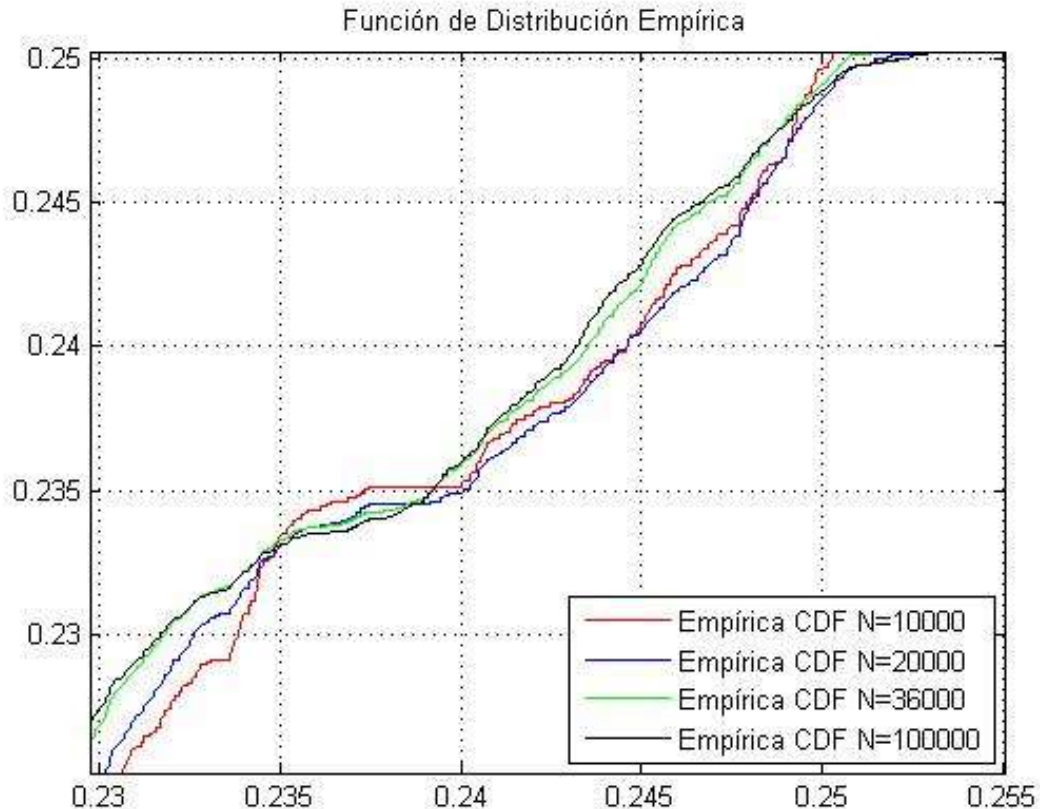


Figura 15. Ampliación de figura 14.

Teniendo la aproximación de la función de distribución acumulativa lograda a partir de la función de distribución empírica, el paso a seguir es encontrar una distribución matemática para la variable aleatoria esquematizada en las figuras 14 y 15.

En la escogencia de la distribución se debe tener en cuenta que ésta debe ajustarse a la función de distribución acumulativa en un mayor grado que sus similares, contra las cuales se realiza la comparación. Las distribuciones que se tomaran como base son:

Exponencial, Gumbel, Weibull, Pareto. La distribución exponencial es la más utilizada para modelar el comportamiento en las redes tradicionales, las distribuciones de Gumbel y de Weibull son utilizadas frecuentemente en la cuantificación de la distribución de las variables aleatorias independientes o débilmente dependientes, estas dos distribuciones

son también llamadas distribuciones de valores extremos, en este sentido también se utilizará una aproximación por medio de distribuciones de valores extremos generalizada, en lo que se refiere a la distribución de Pareto esta tiene inmersa diversas propiedades que se han encontrado en las redes modernas una de ellas la dependencia de largo rango. La figura 16 muestra cada una de estas distribuciones frente a la distribución acumulativa Empírica de los datos.

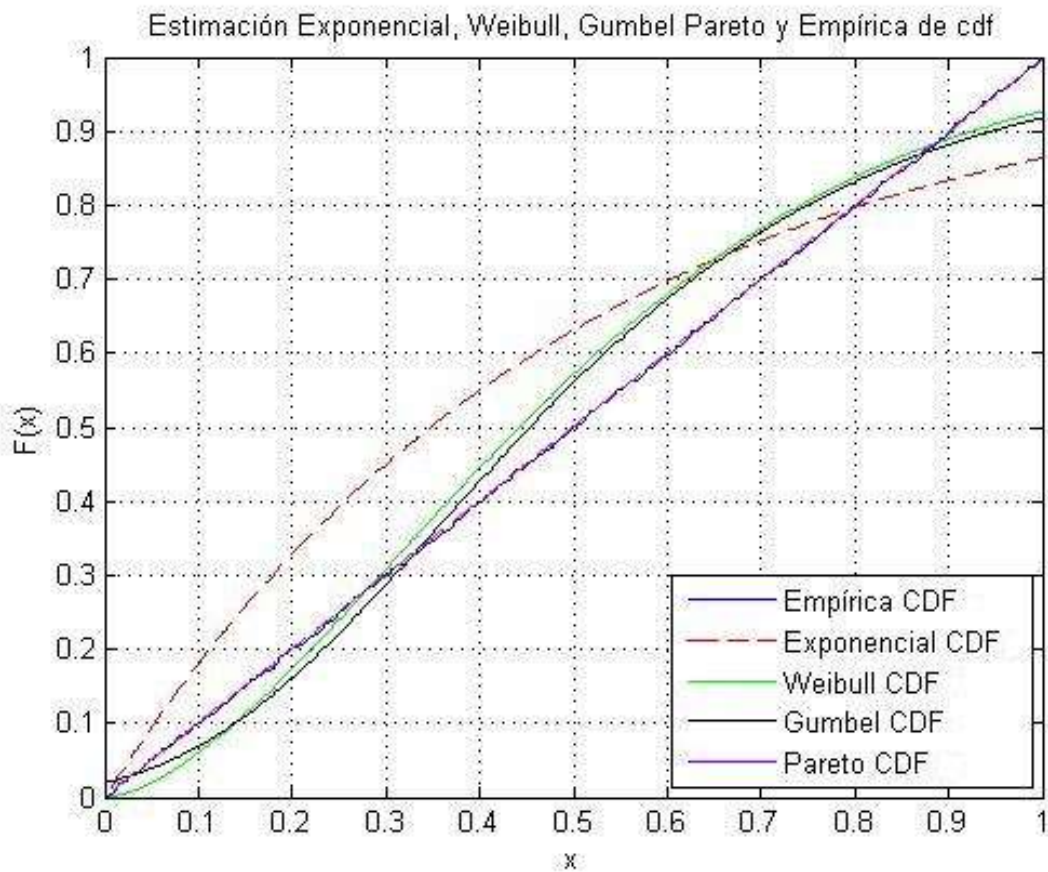


Figura 16. Las distribuciones: Exponencial, Gumbel, Weibull, y Pareto frente a la distribución acumulativa Empírica.

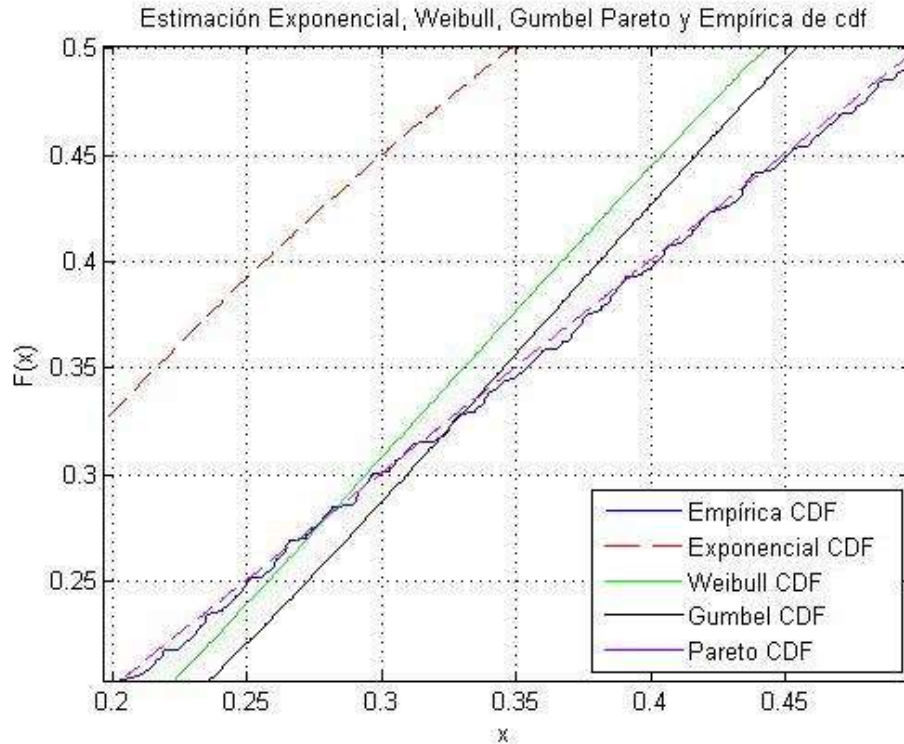


Figura 17. Ampliación de un sector de la figura 16.

En la figura 16 se puede distinguir como las diferentes distribuciones se aproximan a la función de distribución empírica de las muestras, con el fin de hacer una mejor visualización del ajuste en la distribución de Pareto presentada en la figura 16, se realiza una ampliación de un sector de esta gráfica. Esta ampliación se puede observar en la figura 17.

El ajuste visual a partir de la observación de la figura 16 y 17 conduce a la distribución de Pareto, como la distribución que mejor refleja el comportamiento de los paquetes de voz IP en redes NGN. Esto trae consigo una serie de implicaciones que ayudan a caracterizar el tráfico. La distribución de Pareto tiene una propiedad típica que sirve para representar fenómenos de cola pesada, esto es, su función de distribución complementaria decae más lentamente que la distribución exponencial. Por la tanto el tipo de tráfico de voz en redes reales puede ser auto-similar. Para tener un mejor criterio del comportamiento de

los paquetes de voz en redes NGN se consideró lo expuesto anteriormente, como la distribución escogida para caracterizar el tráfico de voz es la distribución de Pareto, se comprobara si el fenómeno de autosimilitud se conserva, en otras palabras se realiza la comparación de la distribución empírica frente a las diferentes distribuciones para diferentes tamaños de muestras. Lo que se espera en este caso es que la distribución de Pareto sea la que más se ajuste en todos los casos.

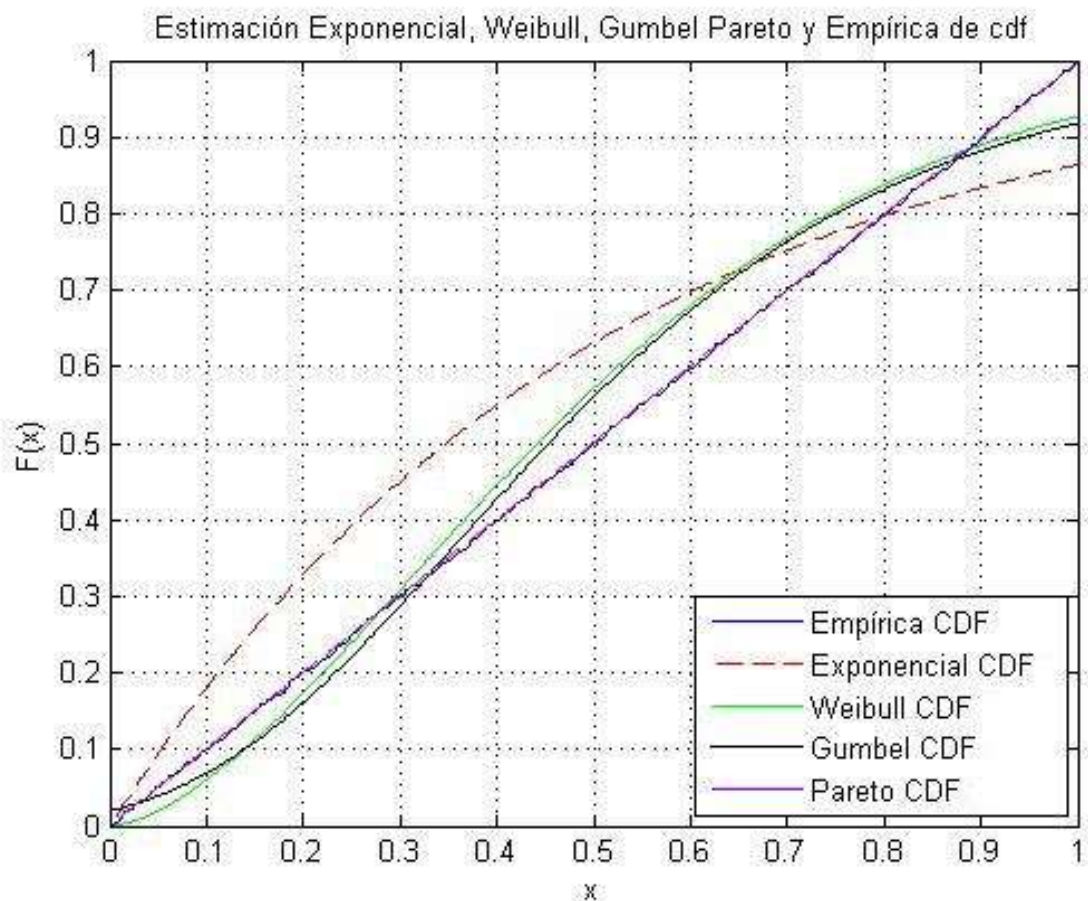


Figura 18. Las distribuciones: Exponencial, Gumbel, Weibull, y Pareto frente a la distribución acumulativa Empírica para 10.000 muestras

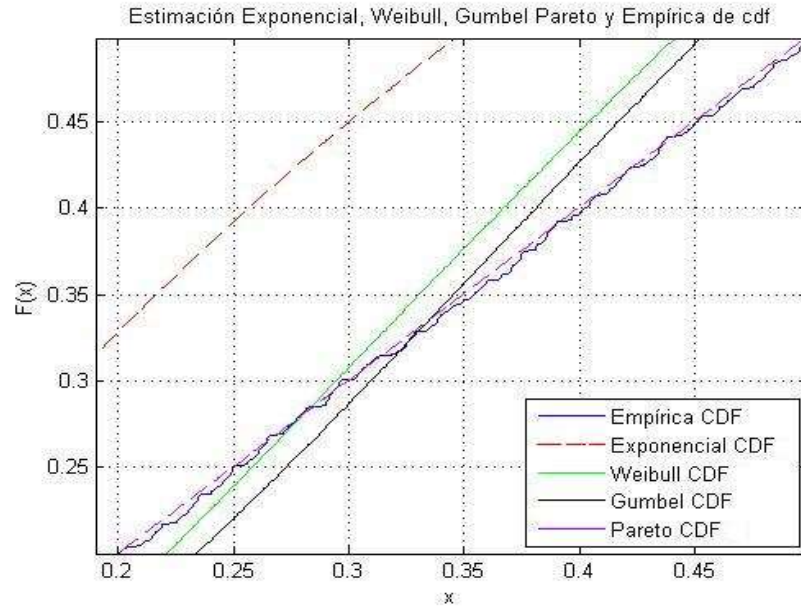


Figura 19. Ampliación de un sector de la figura 18.

En la figura 18 el número de muestras es de 10000, la figura 19 es una ampliación de la 18 en la cual se puede observar como se esperaba que la función de Pareto es la que más se asemeja a las muestras.

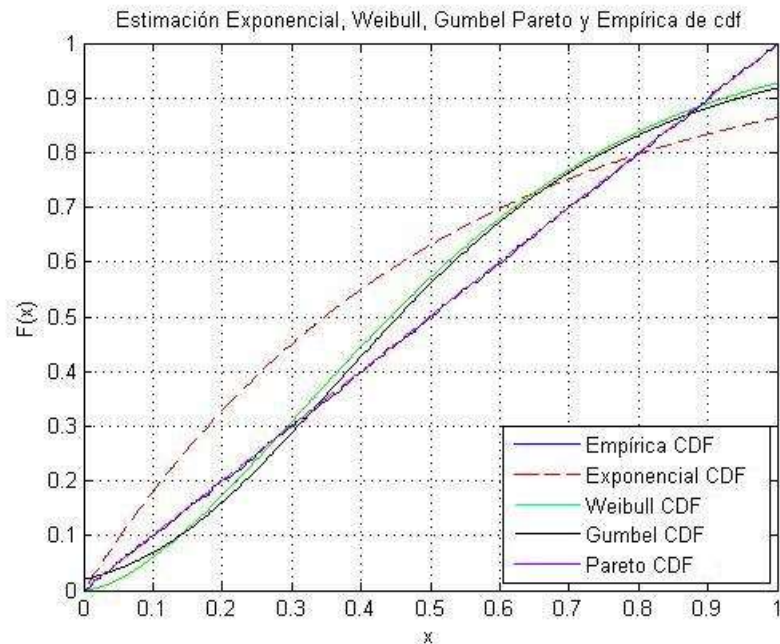


Figura 20. Las distribuciones: Exponencial, Gumbel, Weibull, y Pareto frente a la distribución acumulativa Empírica para 30.000 muestras.

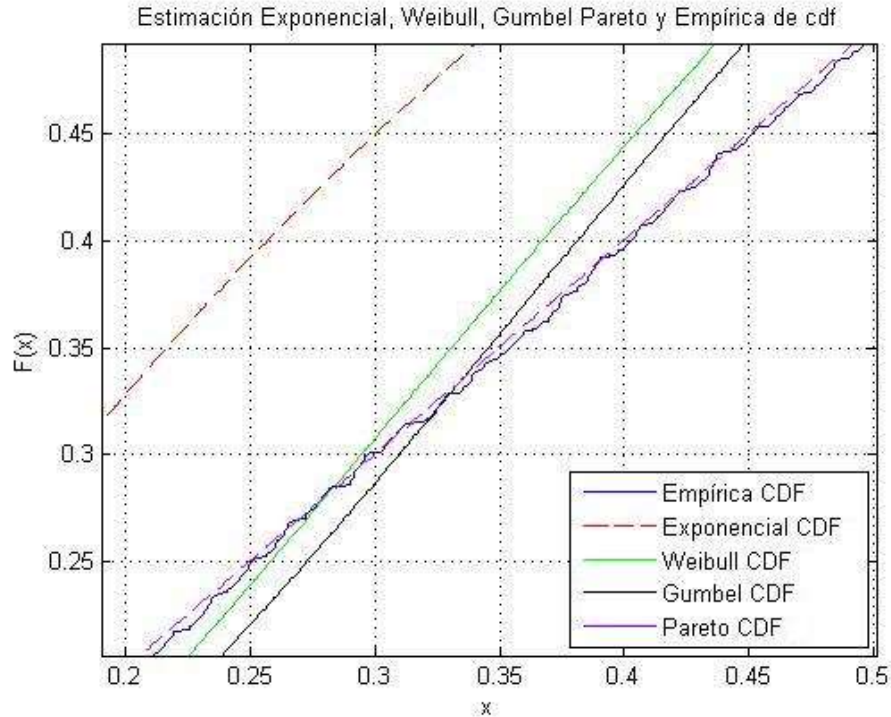


Figura 21. Ampliación de un sector de la figura 20

En la figura 20 para un número de muestras de 30000 la distribución de Pareto es la que más se aproxima, se puede observar mejor en la figura 21. La primera conclusión que se puede obtener es que se puede estar presentando el fenómeno de autosimilitud en los datos de voz IP, ya que para diferentes tamaños de muestra se conserva la distribución de Pareto como la mejor aproximación a la distribución acumulativa de los datos.

Otra técnica gráfica muy común que también se utiliza como método visual para una buena evaluación y una adecuada estimación con el fin de determinar si los datos de una población descenden, son originados o se ajustan a una distribución en concreto es la gráfica de quantile-quantile (q-q).

En una gráfica q-q se gráfica los quantiles de los primeros datos frente a los quantiles de los segundos datos. Quantile significa la fracción (porcentaje) de puntos bajo un valor dado. Ejemplo de esto es que si habla de un quantile de 0.3 es porque en este punto el

30% de los datos esta por debajo del mismo y el 70% esta por encima de este valor. Sabiendo lo útil que pueden resultar las gráficas de cuantiles-cuantiles, se utilizaran para realizar una nueva comprobación de si realmente la distribución de Pareto caracteriza el tráfico real de voz IP.

El procedimiento seguido fue el siguiente:

1. Por medio del toolbox Wafo, el cual esta disponible libremente en Internet para ser utilizado en Matlab, se puede observar en una misma gráfica la función de distribución acumulativa empírica a partir de una serie de datos frente a varias distribuciones, como las observadas en las figuras 16 a la 21, entre las cuales se tiene la distribución de Pareto generalizada. Además de esto dentro del proceso de ajuste el programa arroja una serie de valores dependiendo de la distribución con la cual se confronta los datos reales, específicamente para la distribución de Pareto los parámetros que se arrojan son el de escala y el de forma. Explícitamente en Wafo se maneja la distribución de Pareto generalizada, el resumen de los diferentes valores que arroja el toolbox para diferentes tamaños tomados de los datos de tráfico real de voz IP se muestra en la tabla 3.

Número de Muestras de Voz IP	Parámetro de Forma	Parámetro de Escala
10000	0.9928	0.9969
20000	0.9879	0.9925
36579	0.9844	0.9908

Tabla 3. Parámetros de escala y de forma, arrojados para la distribución de Pareto

2. Teniendo los parámetros de escala y de forma arrojados en el paso anterior, se generan una serie de muestras, en este caso se generan 36579.

3. Confrontación de las muestras generadas en el paso dos frente a las muestras de tráfico real de voz IP en un gráfico de cuantiles-cuantiles. La figura 22 muestra el resultado.

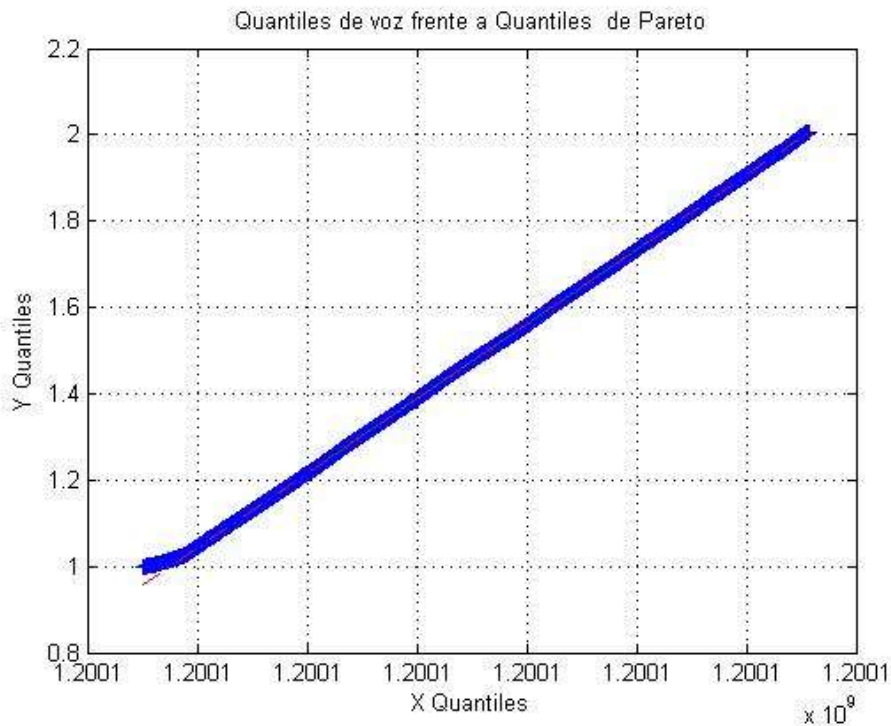


Figura 22. Cuantiles de voz frente a Cuantiles de Pareto

De la gráfica anterior se puede concluir que la apreciación es cierta en cuanto la distribución escogida es la adecuada para sintetizar las características de comportamiento de los tiempos de llegada de los paquetes de voz IP.

Una manera de conseguir una serie de muestras que se pueden utilizar como aproximaciones a muestras de tráfico real, es por medio de programas generadores de tráfico, diferentes técnicas y métodos se pueden utilizar. A las muestras generadas en este trabajo por medio de Cadenas de Markov y Movimiento Browniano Fraccional, se les realiza un análisis con el fin de encontrar el método más adecuado que refleje las

principales características que este poseen y de este modo dar un concepto sobre cual método es el mejor descriptor del comportamiento de la voz en redes reales.

Siguiendo con el análisis una comprobación más profunda se necesita para establecer si los datos obedecen a la autosimilitud, para este fin se tiene una serie de métodos que permiten asegurar si una colección de datos presenta dependencia de largo o corto rango, los métodos que se utilizaran son: Varianza-Tiempo y Periodograma⁷, los cuales nos permiten establecer si el parámetro de Hurst es mayor o menor que 0.5, este valor informa si se presenta el fenómeno de autosimilitud o no, de la siguiente manera, si los datos están por debajo de $H < 0.5$ el fenómeno dominante es la dependencia de corto rango y no hay manifestación de autosimilitud, por el contrario si $H > 0.5$ existe dependencia de largo rango y de autosimilitud.

En la figura 23 se muestra el resultado de evaluar los datos por medio del método de varianza-tiempo, el valor del parámetro de Hurst para diferentes tamaños de muestra se presentan en la tabla 4.

Cantidad de Muestras	Valor del Parámetro de Hurst
10000	0.8756
20000	0.8126
36579	0.7378

Tabla 4. Valor del parámetro de Hurst frente al número de muestras.

⁷ ANEXO D: Métodos de estimación del parámetro de Autosimilitud.

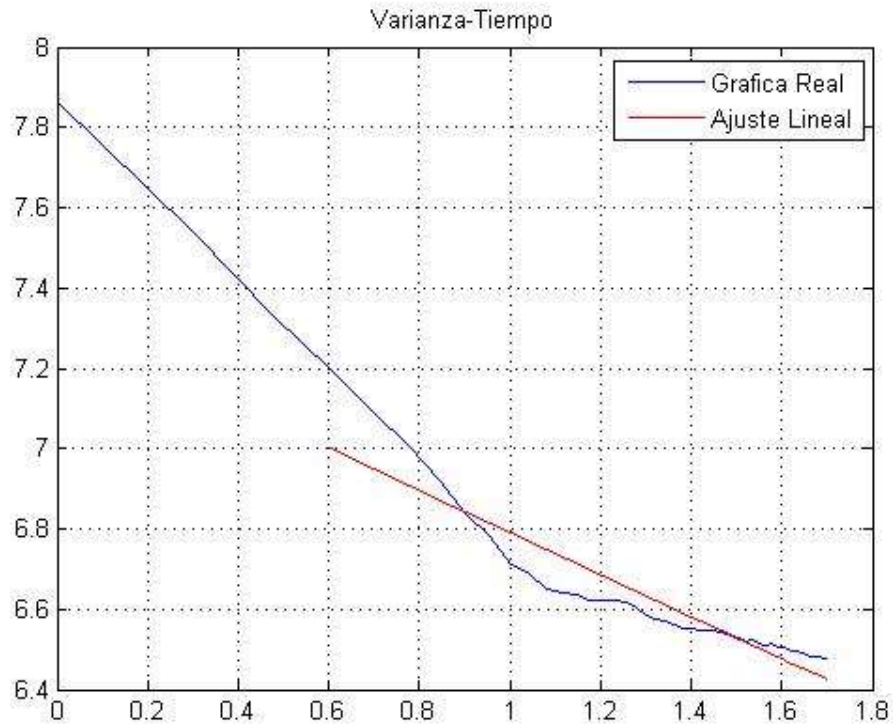


Figura 23. Gráfica Varianza-Tiempo

Seguidamente se muestra el método de periodograma en el cual también se estima el parámetro de Hurst como en el caso de Varianza-Tiempo, este parámetro se calcula con los valores cercanos al origen, Ver figura 24 y tabla 5.

Cantidad de Muestras	Valor del Parámetro de Hurst
10000	0.8824
20000	0.8695
36579	0.7044

Tabla 5. Valor estimado del parámetro Hurst por medio del Periodograma

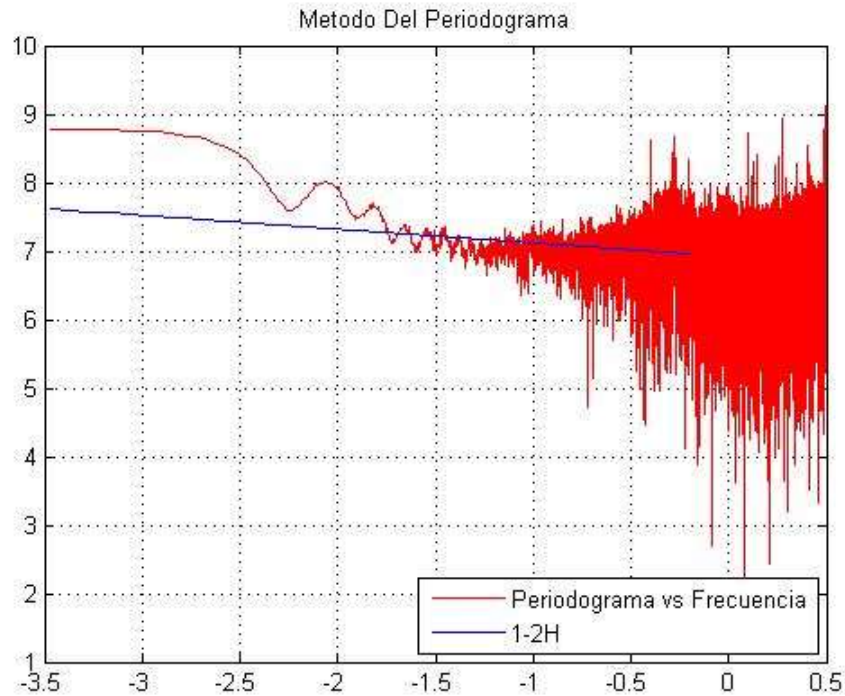


Figura 24. Gráfica de Periodograma de voz IP

Los datos obtenidos en las tabla 4 y 5 corroboran aun más que el tráfico de voz IP es auto-similar, también es de notar que los resultados obtenidos con los dos métodos anteriores son muy similares y exhiben el fenómeno de dependencia de largo rango.

En el capítulo 3 se indicó que para generar muestras de Movimiento Browniano Fraccional se iba a utilizar el método de Cholesky y el método de Paxson. Estos dos métodos necesitan unos valores de entrada, los cuales son la base de la generación de tráfico sintético, exactamente se necesitan dos datos, estos datos de entrada son de tipo numérico. El primer dato que se necesita como entrada es el que corresponde al exponente Hurst, y el otro corresponde a la cantidad de muestras a generar.

Con el método de Cholesky se genera una serie de muestras graficadas en la figura 25, los valores y el número de muestras se encuentran consignados en la tabla 6, a estas

muestras se les realizo la estimación del exponente Hurst, por medio de los métodos de Varianza -Tiempo y el Periodograma los resultados arrojados por estos métodos de estimación también se encuentran consignados en la tabla 6.

Valor de H utilizado para Generación de muestras	Número de Muestras	Valor calculado por el Método de Varianza-Tiempo	Valor calculado por el Método del Periodograma
0.6	5000	0.5809	0.6302
0.7	5000	0.7030	0.7113
0.8	5000	0.7568	0.8164

Tabla 6. Muestras generadas con el método de Cholesky y estimación de H

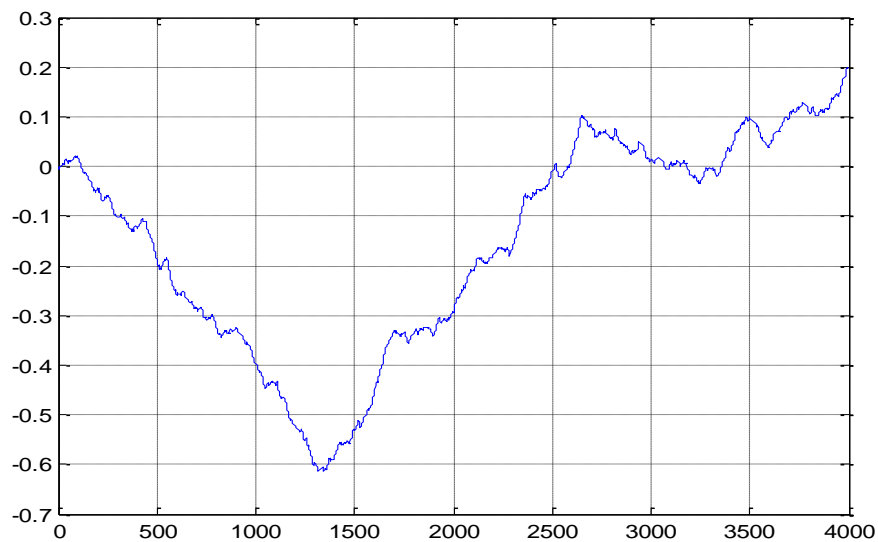


Figura 25. Gráfico generado por el método de cholesky

En las figuras 26 y 27 se observan las gráficas generadas por los métodos de Varianza -Tiempo y Periodograma a partir de las muestras arrojadas por el método de Cholesky con un parámetro de Hurst de 0.8.

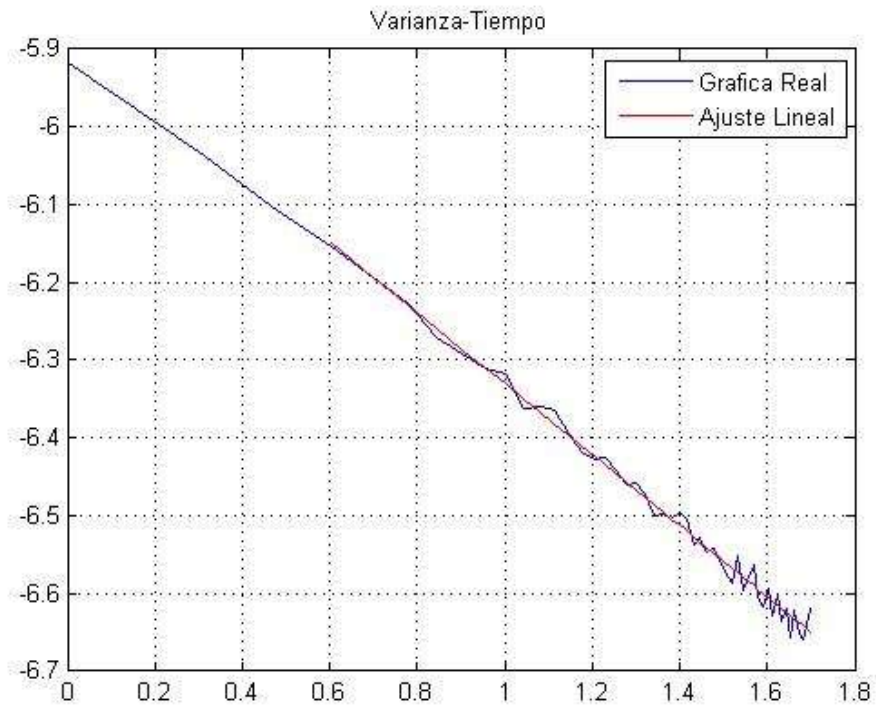


Figura 26. Método de Varianza tiempo para las muestras generadas con el método de Cholesky

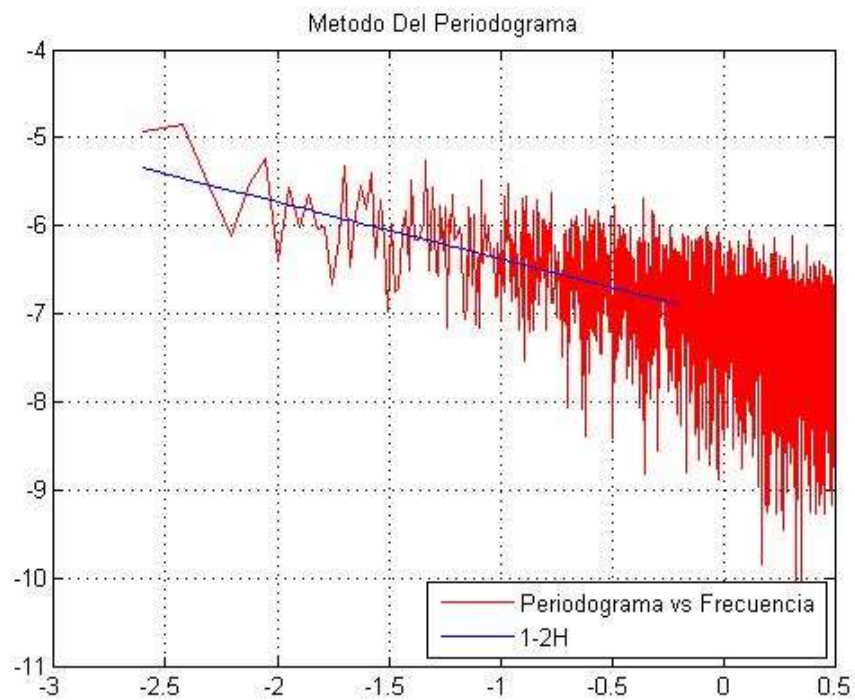


Figura 27. Método de Periodograma para las muestras generadas con el método de Cholesky

Con el método de Paxson se realiza el mismo procedimiento hecho anteriormente para el método de Cholesky, en la tabla 7 y en las figuras 28 ,29 y 30 se pueden ver estos resultados. La figura 28 presenta el gráfico arrojado por este método.

En este caso se tomo un valor de 0.7 para parámetro de Hurst, y con dicho valor se generaron las gráficas para los métodos de Varianza – Tiempo y el del Periodograma.

Valor de H utilizado para Generación de muestras	Número De Muestras	Valor calculado por el Método de Varianza-Tiempo	Valor calculado por el Método del Periodograma
0.6	100000	0.5989	0.6048
0.7	100000	0.6898	0.6987
0.8	100000	0.7929	0.7917

Tabla 7. Muestras generadas con el método de Paxson y estimación de H

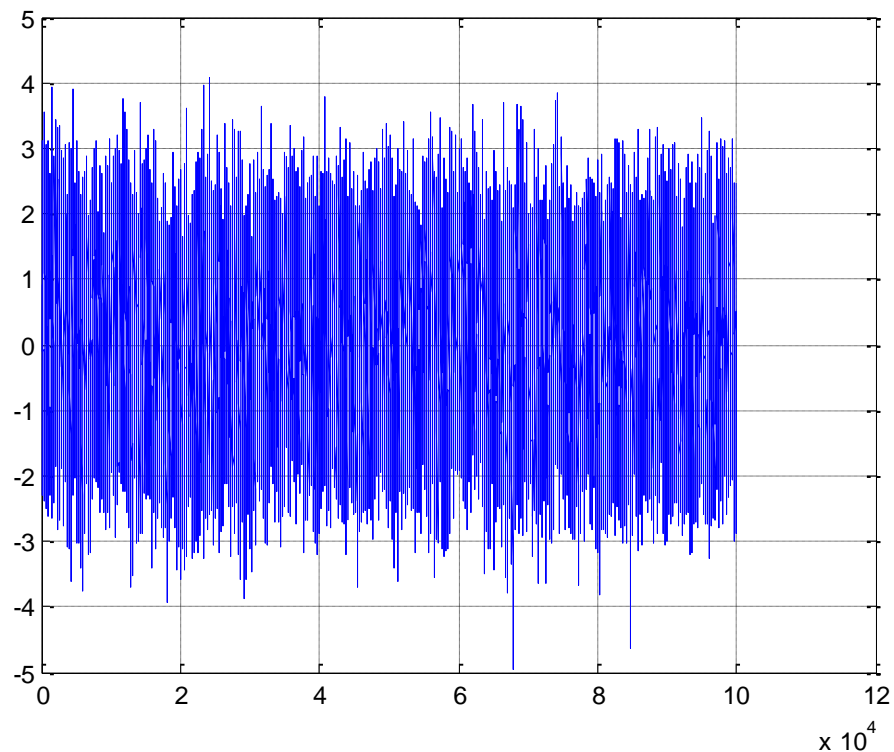


Figura 28. Gráfico generado por el método de Paxson

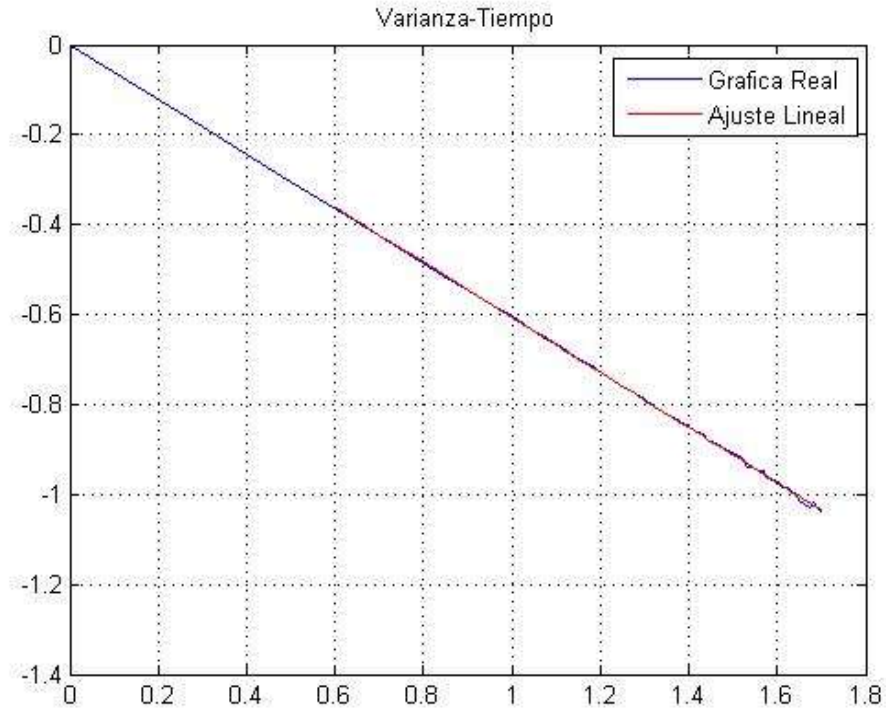


Figura 29. Método de Varianza tiempo para las muestras generadas con el método de Paxson.

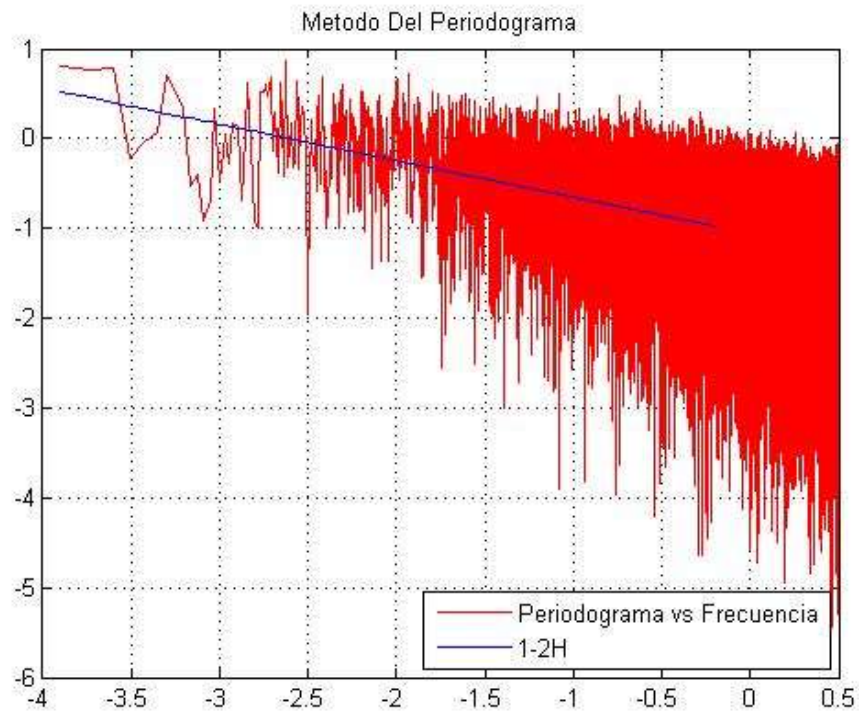


Figura 30. Método de Periodograma para las muestras generadas con el método de Paxson.

En cuanto a los procesos de Poisson modulados por Markov del capítulo 3, se presentaron dos algoritmos que permiten simular la generación de tiempos de arribos de paquetes (tiempo de arribo entre paquetes) y número de paquetes. Con estos se crearon unos archivos de tráfico para diferentes cantidades de muestras.

En estos algoritmos se simula un proceso de poisson modulado por markov de dos estados debido al alto costo computacional que implica un modelo con más estados. Estos dos algoritmos necesitan unos valores de entrada los cuales son esenciales para la generación del tráfico sintético, estos datos de entrada son, el número de muestras que se quieren generar, las tasas de arribo de cada estado (λ_1 y λ_2) y la probabilidad de permanencia en cada estado (w_1 y w_2). Para el cálculo de estos se utilizó el método de estimación de máxima verosimilitud (MLE) [35] y los programas Algorithm1 y NewAlgorithm.

Con los métodos expuestos anteriormente se generó una serie de trazas de tráfico, con diferentes cantidades de muestras las cuales son similares a las utilizadas para encontrar el valor del parámetro de Hurst en las muestras de tráfico real (tablas 4 y 5), estos se encuentran consignados en la tabla 8.

Para corroborar los resultados obtenidos con los dos métodos anteriores, se hizo uso del programa Algorithm1⁸ hecho en matlab el cual permite encontrar una traza MMPP de una traza de tráfico real de VoIP.

⁸ Disponible en <http://www.ittc.ku.edu/~sbali/MMPP/>

Número de Muestras	Valor calculado por el Método de Varianza-Tiempo	Valor calculado por el Método del Periodograma
10000	0.5409	0.5157
20000	0.5295	0.5046
36579	0.5290	0.5115
50000	0.5228	0.5191
80000	0.5406	0.5561

Tabla 8. Muestras generadas con el método MMPP y la estimación de H.

En la figura 31 se puede ver una traza de tráfico real, en la figura 32 podemos observar una muestra sintética generada con el método MMPP. En las figuras 33 y 34 se observan las gráficas generadas por los métodos de Varianza -Tiempo y Periodograma a partir de las muestras arrojadas por los métodos para un número de muestras igual a 36579, para este número de muestras de tráfico real también se halló el parámetro H (tablas 4 y 5).

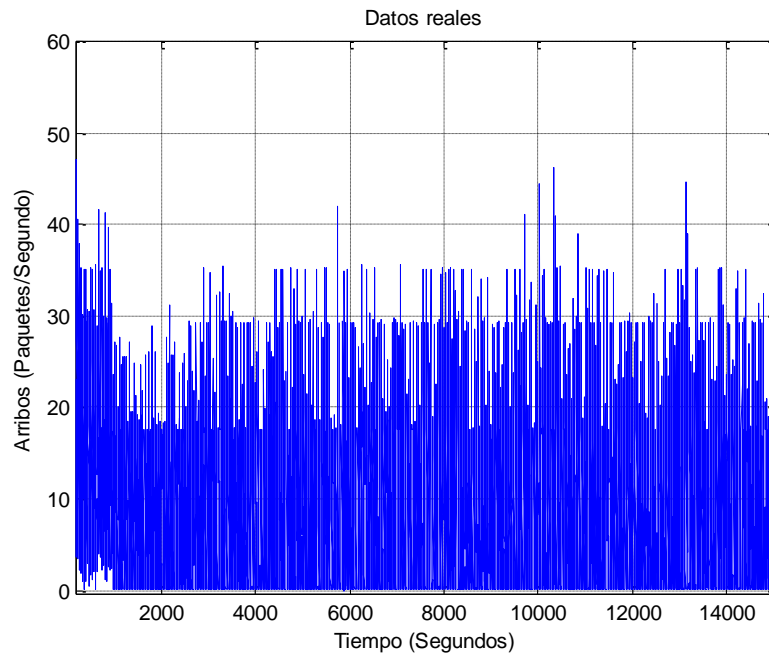


Figura 31. Grafica de 14.000 muestras de tráfico real.

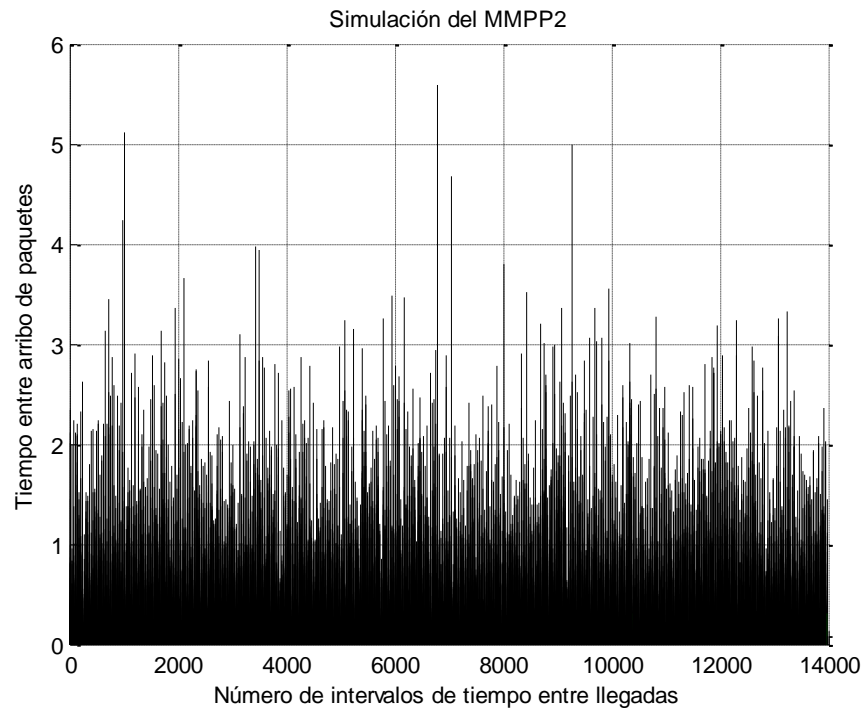


Figura 32. Muestras sintéticas generadas con el modelo mmpp.

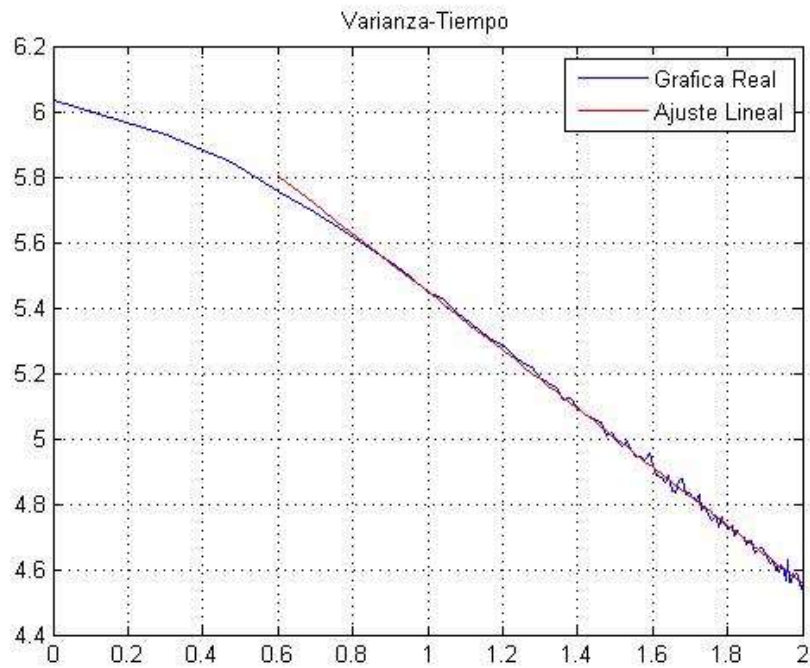


Figura 33. Método de Varianza tiempo para las 36579 muestras generadas con el método MMPP

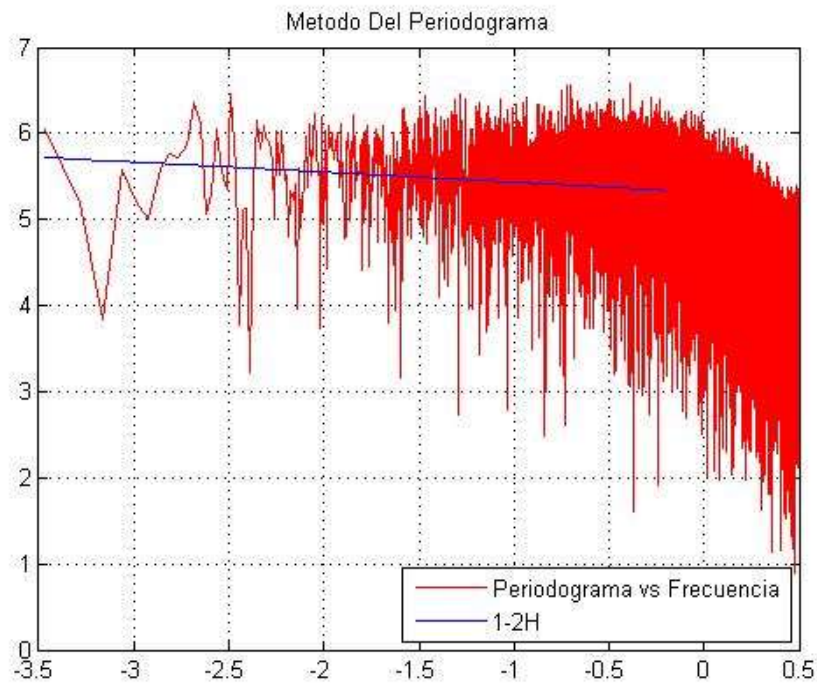


Figura 34. Método de Periodograma para las 36579 muestras generadas con el método MMPP

Como conclusión se establece que en el tráfico de voz sobre IP se encuentran presentes los fenómenos de autosimilitud y dependencia de largo rango, también se puede decir que el modelo que mejor representa el comportamiento de este tipo de tráfico, es el movimiento browniano fraccional por su eficacia en la generación de muestras sintéticas que se aproximan al tráfico real de voz IP.

5 CONCLUSIONES

- Un estudio de los parámetros y del comportamientos de los diferentes tipos de tráfico en las redes actuales se hace necesario, ya que el conocimiento de estos ayudan o permiten el diseño de mecanismos que respondan de una manera más eficiente a los requerimientos de calidad que se exigen actualmente, este estudio también ayuda en las tareas de dimensionamiento y en la evolución de las redes hacia los requerimientos futuros, de manera especial se requiere una respuesta efectiva, eficaz y óptima a los servicios que exigen respuesta en tiempo real, en este caso el tráfico de voz IP.
- Un nuevo enfoque ha surgido en los últimos años para el análisis de los diversos fenómenos presentes en la redes de comunicaciones en las cuales los conceptos y métodos que se utilizaban no respondían satisfactoriamente, debido a esto diversos estudios se han hecho con el fin de encontrar técnicas que representen satisfactoriamente las características del tráfico agregado, entre los conceptos que se manejan, se tienen: el concepto de autosimilitud, fractabilidad, dependencia de largo rango etc. En las redes NGN se habla de estos debido a que permiten explicar el comportamiento del tráfico convergente.
- La estadística es importante para poder establecer y cuantificar el grado de autosimilitud y la dependencia de largo rango. Sirve además para estudiar el inmenso volumen de tráfico circulante por las redes actuales, dar una inferencia del modelo que

presumiblemente obedece a los datos generados, en general sirve para realizar el ajuste del tráfico desconocido a un modelo matemático específico.

➤ Se puede afirmar que la agregación de tráfico de voz en las redes reales NGN posee un comportamiento auto-similar lo cual implica que no existe total independencia entre los paquetes de datos. En otras palabras, el tráfico que se encuentra en la red en un instante cualquiera se ve afectado de alguna manera por el tráfico del intervalo anterior. También se puede decir que los paquetes de voz se encuentran mejor caracterizados por la distribución de Pareto, la cual tiene la propiedad de ser una distribución de cola pesada y de caída hiperbólica. Esto se puede observar en las figuras 14 a 21, en las cuales se realiza la caracterización del tráfico de voz sobre IP, mediante la comparación de la función de distribución acumulativa empírica de los datos reales frente a las funciones de distribución acumulativa de varias distribuciones de probabilidad. El método de cuantiles-cuantiles (Figura 22) permite corroborar el resultado expuesto anteriormente.

➤ Entre los modelos que se presentaron en este trabajo, el que mejor responde a las características de las trazas de VoIP en las redes NGN es el método FBM (utilizado cuando se presenta el fenómeno de autosimilitud), los parámetros analizados en las trazas que fueron generadas con este, corresponden en gran medida con las características observada en las trazas reales de voz IP. En las tablas 4, 5, 6 y 7 se encuentran consignados los valores del parámetro de hurst calculados para las muestras de tráfico real y para las muestras generadas con los métodos de Cholesky y Paxson (métodos FBM) respectivamente, los valores que ahí se encuentran permiten realizar esta afirmación. Aunque los valores encontrados en las estimaciones del parámetro de hurst para las muestras de tráfico hechas con el modelo MMPP (tabla 8) se encuentran en el

rango $0.5 < H < 1$ que implica autosimilitud no ofrecen una buena aproximación a los valores obtenidos para el tráfico real. Sin embargo, podría tratarse de lo expuesto en algunos estudios [4], [5], [6], en los cuales se afirma que los modelos simples tales como el MMPP aunque no son modelos totalmente auto-similares poseen una correlación limitada. Por lo tanto, el modelo que permite caracterizar de una forma efectiva el tráfico de VoIP en el núcleo de las redes NGN son los métodos basados en el Movimiento Browniano.

- Los conceptos de autosimilitud, dependencia de largo rango, distribución de cola pesada están ligados de algún modo; cuando se observa la presencia de alguno de estos fenómenos es casi seguro la presencia de los demás, debido a su estrecha relación.

- Una de las formas y herramientas por medio de las cuales se puede dar una estimación cuantitativa del grado de autosimilitud es por medio del cálculo de un parámetro llamado el exponente de Hurst, el cual es de interés cuando se encuentra entre un valor superior a 0.5 y menor que 1, debido a que es en este intervalo en donde se puede afirmar que un determinado tráfico tiene dependencia de largo rango.

- Distribuciones matemáticas como la de Pareto son muy utilizadas hoy en día ya que estas distribuciones poseen características de cola pesada y esto implica una relación íntima con la dependencia de largo rango, las diferentes pruebas que se realizaron en este trabajo de grado arrojaron que la representación óptima por medio de la cual se puede hacer referencia al comportamiento del tráfico de voz en redes actuales NGN es relacionando este tráfico con las características propias de la distribución de Pareto.

- Entre los trabajos futuros que se pueden plantear están:

Estudio comparativo de la exactitud para modelar comportamiento auto-similar por medio de diferentes métodos que se utilizan para la generación de muestras de Movimiento Fraccional Browniano.

Estudio de las adaptaciones de los modelos basados en MMPP propuestos para recoger las características de autosimilitud del tráfico actual, con el fin de comprobar su correcto desempeño frente a modelos auto-similares.

Análisis del impacto de real audio en el núcleo de las redes, esto con el fin de establecer una serie de políticas de gestión en las redes que permitan que este servicio no afecte de manera negativa a los servicios de tiempo real.

GLOSARIO

ARMA: Promedio Variable Auto regresivo

ATM: Modo de Transferencia Asíncrono

CDF: Función de Densidad de Probabilidad Acumulativa

CMTC: Cadena de Markov de Tiempo Continúo.

DiffServ: Diferenciación de Servicios

dMMPP: Proceso de Poisson modulados por Markov de Tiempo Discreto.

FFT: Transformada de Fourier Rápida

FIFO: Primero En Entrar Primero en Salir

H.323: es una recomendación del ITU-T (Union Internacional de Telecomunicaciones), que define los protocolos para proveer sesiones de comunicación audiovisual sobre paquetes de red.

IntServ: Servicios Integrados.

IP: Protocolo Internet.

ISDN: Red de Servicios Integrados.

ISP: Proveedor de Servicios de Internet.

ITU: Union Internacional de Telecomunicaciones

ITSP: Proveedor de servicios de Telefonía e Internet.

FBM: Movimiento Fraccional Browniano.

FGN: Ruido Fraccional Gaussiano.

LAN: Red de Área Local.

LIFO: Último en Entrar Primero en Salir.

LRD: Dependencia de Largo Rango.

MGCP: Es un protocolo de control de dispositivos, donde un gateway esclavo (MG, Media Gateway) es controlado por un maestro (MGC, Media Gateway Controller, también llamado Call Agent).

MEGACO: Define el mecanismo necesario de llamada para permitir a un controlador Media Gateway el control de puertas de enlace para soporte de llamadas de voz/fax entre redes RTC-IP o IP-IP

MMPP: Procesos de Poisson Modulados por Markov.

MUX: Multiplexor

NGN: Red de Nueva Generación.

POTS: Sistema de Teléfono Antiguo

QoS: Calidad de Servicio

RSVP: Protocolo de reserva de recursos.

RT: Tiempo Real.

RTP: Protocolo de Tiempo Real

SIP: Protocolo de Información de Sesión

SDP: Protocolo de Descripción de Sesión.

SLA: Acuerdos de Nivel de Servicio.

TCP: Protocolo de Control de Flujo

UDP: Protocolo de Datagrama de Usuario

VBR: Tasa de Bits Variable

VoIP: Voz sobre IP

WAN: Red de Área Amplia.

REFERENCIAS

- [1] Walter Willinger, Murad S. Taqqu, Robert Sherman and Daniel V. Wilson. "Self-Similarity Through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level". Paper. Departamento de matemática. Universidad de Boston.
- [2] W.E Leland, M.S Taqqu, W. Williger, and D.V Wilson, "On the Self-Similar Nature of Ethernet Traffic," Article. IEEE/ACM Trans. Networking, vol. 2, issue 1, February 1994, pp. 1-15.
- [3] Walter Willinger, Vern Paxson and Murad S. Taqqu. "Self-similarity and Heavy Tails: Structural Modeling of Network Traffic". Paper. AT&T Labs-Research, Lawrence Berkeley National Laboratory, Department of Mathematics, Boston University.
- [4] M. Grosslauser, J. Bolot, "On the relevance of long-range dependencies in network traffic", IEEE/ACM Transaction on Networking 7 (1999)
- [5] S. Ben Fredj, T. Bonald, A. Proutiere, G. Régnie', J. Roberts, "Statistical Bandwidth Sharing: a Study of Congestion at Flow Level, in: Proceedings of ACM SIGCOMM 2001", San Diego, CA, USA, August 2001.
- [6] T. Bonald, A. Proutiere, G. Régnie', J. Roberts, Insensitivity Results in Statistical Bandwidth Sharing, in: Proceedings of International Teletraffic Conference (ITC) 2001, San Paolo, Brazil, November 2001.

- [7] Angus Ma. SmartBits, Performance Analysis System, "Voice over IP (VoIP)" Spirent Communications, 2001. Documento en pdf.
- [8] Grupo de expertos sobre telefonía IP del UIT-D, "Informe esencial sobre telefonía IP". Unión Internacional de Telecomunicaciones UIT 2003. Documento en pdf.
- [9]. Maurice David Woernhard. "Generating Synthetic VoIP Traffic for Analyzing Redundant OpenBSD-Firewalls". Tesis de Maestría. Universidad de Oslo. Departamento de Informática. Mayo 23, 2006. Documento en pdf.
- [10] Ling Fen Sun, "Speech Quality Prediction for Voice over Internet Protocol Networks". School of Computing, Communications and Electronics Faculty of Technology. Universidad de Plymouth. Enero 2004. Documento en pdf.
- [11] Villy B. Iversen "TELETRAFFIC ENGINEERING HANDBOOK". Junio 20,2001
- [12] Ivo Adan. Jacques Resing. "Queueing Theory" Department of Mathematics and Computing Science Eindhoven University of Technology. The Netherlands. 2001
- [13] Rafael Camerano Fuentes, "Teoría de colas: Aplicación a las telecomunicaciones". Libro, Biblioteca de catedráticos, Fondo de publicaciones de la universidad distrital Francisco José de Caldas, primera edición 1997.
- [14] Richard G. Clegg. "The Statistics of Dynamic Networks". Tesis de Doctorado Universidad de York Departamento de Matemáticas. June de 2004. Documento en pdf.
- [15] Gunter Bolch, Stefan Greiner, Hermann de Meer, Kishor S. Trivedi, "Queueing networks and Markov chains". Libro, A John Wiley & Sons, inc., Publication. Segunda edición. 2004.

-
- [16] L. Kleinrock. "Queueing Systems". Libro. Volumen 1. John Wiley & Sons, 1975.
- [17] Oleg I. Sheluhin, Sergey M. Smolskiy, Andrey V Osin. Self-Similar Processes in Telecommunications. John Wiley & Sons. Ltd 2007.
- [18] Selwin Jakobus Emiel Roon, "Resource dimensioning in a mixed traffic environment". Tesis, Faculty of Engineering, University of Pretoria, September 2004. Documento en pdf.
- [19] J.W Roberts, "Traffic theory and the internet", IEEE Communications Magazine, vol. 39, issue 1, January 2001, pp 94-99
- [20] Z. Hulicki, "On some aspects of modelling and dimensioning in broadband access networks", Proceedings of the 31 st Annual Simulation Symposium, 5-9 April 1998, pp 106-1409.
- [21] A. Adas, "Traffic models in broadband networks", Article, Communications Magazine, IEEE, Vol 35, issue 7, July 1997, pp 82-89.
- [22] B. B. Mandelbrot. "The Fractal geometry of nature. Freeman". Libro. Freeman and Company. New York, 1983.
- [23] Kenneth Falconer. "Fractal Geometry, Mathematical Foundations and Applications" John Wiley & Sons. Ltd. 2003.
- [24] Lakshmi Yellanki. "Impact of short-range and Lon-range Dependence on Variable Bit Rate Video Traffic Performance". Paper. Departamento de Ciencias de la Computación. Universidad de Saskatchewan Canadá. Documento en pdf
- [25] Kihong Park¹, Walter Williger². "Self-similar Network Traffic: An overview."
- ¹Laboratorio de sistemas de Red, Departamento de Ciencias de la Computación.

Universidad de Purdue, West Lafayette. ² Centro de Investigación en Ciencias de la Información, AT&T laboratorio de investigación, Florham Park, NJ. Paper. Documento en pdf.

[26] Telecoms Research Group. "The Application of Non-linear Dynamics to Teletraffic Modelling". Tesis de doctorado. Queen Mary and Westfield College, University of London. Julio de 1999. Documento en pdf

[27] Izzeldin Ibrahim Mohamed Abdelaziz "Modeling and Performance Evaluation of Self-Similar Behavior of Mpeg-4 Video Traffic Generators. University technology Malaysia abril 2006.

[28] Paulo Jorge Salvador Serra Ferreira, "Modelado de tráfico en redes de telecomunicaciones: modelos Markovianos y basados en sistemas de Lindenmayer". Tesis doctorado, Departamento de Electrónica y Telecomunicaciones de la Universidad de Aveiro. 2005. Documento en pdf.

[29] W. Stallings, "High Speed networks and Internets: Performance and quality of service", Libro, Segunda edición. Prentice Hall, 2001.

[30] B.B. Mandelbrot and J.W. Van Ness. "Fractional Brownian motions, fractional noises and applications". SIAM Review: 10, 1968, 422-437

[31] Murad S. Taqqu and Vadim Teverovsky. "On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series". Paper. Documento en pdf.

[32] Ilkka Norros. "A storage model with self-similar input". Paper. VTT Telecomunicaciones. Otakaari. Finlandia. Documento en pdf.

[33] Ton Dieker. "Simulation of fractional Brownian motion". Tesis de maestría Departamento de Ciencias Matemáticas Universidad de Twente The Netherlands. Año 2002. Documento en pdf.

[34] Vern Paxson. Fast, "Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic". Paper. Grupo de investigación de Red. Laboratorio Nacional de Berkeley Lawrence. Documento en pdf.

[35] A. Andersson and T. Ryden, Maximum Likelihood Estimation of a Structured MMPP with Applications to Traffic Modeling, 13th ITC Specialist Seminar, Monterey, CA, 2000.

[36] Juan Pablo Braña. "Curso de Introducción a la geometría Fractal". Curso dictado a través del sitio Web www.fractaltec.org. Año 2003. Resumen de este curso, en pdf.