

TRÁFICO AUTOSEMEJANTE*

Juan Andrés Pérez Mejía**
Jorge Mauricio Romero Parra***

Resumen

Durante décadas, la ingeniería de redes ha basado el análisis de tráfico en un mismo concepto. Pero, en los últimos años se han venido desarrollando teorías que rompen con estos enunciados y nos dan un nuevo enfoque para dicho análisis.

La mayoría del tráfico existente en las redes actuales se comporta de forma diferente a lo que se pensaba, es de tipo autosimilar o fractal; por lo tanto, se han empezado a desarrollar nuevos métodos para llevar a cabo su estudio y modelamiento. Esta autosemejanza implica un análisis estadístico o probabilístico de los sucesos recurriendo a diferentes herramientas; en este caso, los modelos no correlacionados, sin memoria, de dependencia a corto plazo, de dependencia a largo plazo, Wavelet Multifractales, entre otros.

En este caso se muestra, de forma general, cómo se puede lograr el análisis del tráfico por medio de tales modelos, con el fin de generar inquietudes en el lector.

* El presente artículo fue presentado en la Universidad de Los Andes en el Día de las Comunicaciones, obteniendo el segundo lugar en el concurso de Posters de dicho evento. Sus autores son estudiantes de octavo y décimo semestre, respectivamente, de la Facultad de Ingeniería de Telecomunicaciones de la Universidad Santo Tomás y miembros del Grupo de Investigación INVTEL de dicha Facultad.

** 1991779@correo.usta.edu.co

*** 2012379@correo.usta.edu.co

Palabras clave

Fractales, modelos de Tráfico, autosemejanza, dependencia a corto plazo, dependencia a largo plazo.

Abstract

During decades, traffic analysis has been based in a same notion. However, in the last years, new theories which break with those statements and gave us a new approach for this analysis are getting expounded.

Most of the traffic that now days exists in modern networks has a different performance than expected, now days traffic has a self-similar or Fractal behavior, thus, new methods to study and modeling traffic are getting expounded. Statistics and probability analysis of those events, helped by some tools, such as not correlate, no memory, SRD, LRD and Wavelet Models is needed due to this self-similarity.

This paper shows how traffic analysis can be done through these models in order to generate concerns in the reader.

Index terms

Fractals, traffic modeling, self-similarity, long range dependence, short range dependence.

Introducción

La necesidad de conocer el comportamiento del tráfico telemático subyace a toda discusión relacionada con el dimensionamiento y diseño de controles en redes de telecomunicaciones.

Se puede afirmar que la evolución del mundo de los servicios de telecomunicación desde hace una década ha venido determinada por dos hechos fundamentales:

- En primer lugar, por una mayor globalización de los servicios, que hasta ahora ha podido ser posible especialmente por el desarrollo de una mayor conectividad, pero sobre todo por la fulgurante expansión de un protocolo distribuido como Internet.
- En segundo lugar, se ha constatado la necesidad de una progresiva integración de los servicios que rompe de plano la separación tradicional de las redes físicas exclusivamente pensadas para un tipo de servicio (básicamente redes de distribución de TV, redes telefónicas y redes de datos). Esta integración obliga a compaginar sobre una misma red exigencias de calidad de servicio muy diversas, que van desde el bajo retardo y cierta tolerancia a pérdidas de los servicios de voz, a la tolerancia a retardos y las pérdidas nulas exigidas por muchos servicios de datos.

Cualquiera que sea el escenario definitivo concreto sobre el que se asiente este universo dominado por la integración y la exigencia de calidad, el hecho indiscutible es que existe la necesidad de definir diversos controles y mecanismos de gestión de tráfico que garanticen, de manera individualizada, la calidad exigida por cada servicio. Actualmente muchos de estos controles se encuentran en estado embrionario o, cuando menos, por optimizar de una manera eficiente.

Por su parte, la correcta definición de estos controles, así como el dimensionamiento de componentes de la transmisión (buffers, enlaces ...), exige un amplio conocimiento de los flujos de tráfico que van a ser transmitidos. Este conocimiento implica caracterizar, modelar e incluso predecir con precisión el comportamiento de dichos flujos. Por todas estas razones, el modelado de tráfico telemático se encuentra en el núcleo de la problemática actual de las redes integradas.

Modelamiento del tráfico

Se entiende por modelado de tráfico una abstracción matemática más o menos compleja que trata de imitar alguna o varias características estadísticas de un tipo de tráfico real o de un flujo concreto en particular. Dado que los modelos poissonianos tradicionalmente usados con el tráfico telefónico se han demostrado insuficientes para caracterizar las nuevas fuentes de tráfico, buena parte del esfuerzo investigador en estos últimos años se ha centrado en el modelado de muestras obtenidas en redes reales (Casilari et al.).

Desde el trabajo de Kleinrock a mediados de los 60, se encuentra la existencia de dependencia temporal que ha sido objeto de estudio por parte de numerosos autores, destacando el descubrimiento del enorme impacto que puede llegar a tener sobre las prestaciones de un sistema

de colas. A la vista de este hecho, han ido apareciendo nuevos modelos de tráfico de entrada que exhiben estructuras de correlación más o menos complejas, aplicados a casos en que el modelo del sistema de comunicaciones en estudio permite mantener la tratabilidad analítica. En cualquier caso, estos modelos (fundamentalmente markovianos) desprecian la correlación a partir de determinada separación en el tiempo, aunque ésta puede ser aumentada a costa de complicar el modelo con parámetros adicionales.

Sin embargo, un trabajo presentado por Leland, Willinger, Taqqu y Wilson en ACM SIGCOMM '93 mostraba, tras la observación de exhaustivas mediciones realizadas sobre una red Ethernet en Bellcore (NJ, USA), que el tráfico (número de tramas en la red por unidad de tiempo) era de naturaleza fractal, "autosimilar" o "autosemejante" (self-similar), que implica la existencia de una acusada correlación a largo plazo (Long-Range Dependence -LRD-).

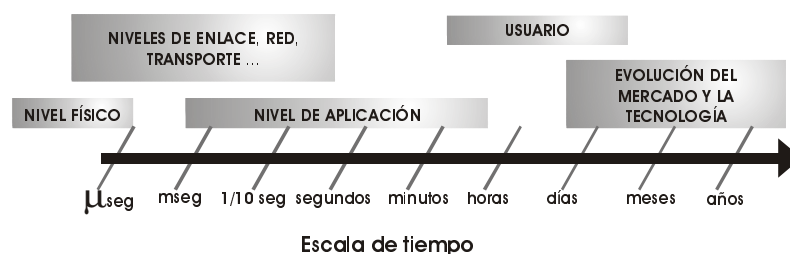
Posteriormente, el mismo carácter fractal ha sido observado en redes WAN, en los datos de SS7, en el tráfico debido a la World Wide Web (WWW), o incluso en el tráfico de video de tasa variable (Variable Bit Rate -VBR-). Todos estos hallazgos han contribuido enormemente a un replanteamiento muy significativo de la modelación estocástica del tráfico de entrada y del análisis de prestaciones, ya que muchos estudios han puesto de relieve el enorme impacto que la dependencia temporal a largo plazo puede llegar a tener sobre las prestaciones de las redes de comunicaciones, frente a otros modelos que se han venido usando hasta hace poco y que, o bien no presentan correlación temporal alguna por simplicidad analítica, o exhiben una estructura de correlación más complicada analíticamente (como los modelos markovianos o autorregresivos habitualmente utilizados), pero que podríamos denominar a corto plazo (Short-Range Dependence-SRD) (López, José).

En este sentido, al investigador le caben dos opciones: muestrear el tráfico en una red a la que tenga acceso y autorización para “espíar” (esto habitualmente sólo es posible para una red de área local LAN) o recurrir a las múltiples bases de datos con muestras de tráfico que ofrece Internet. Los ejemplos de este último tipo son numerosos, tanto para tráfico Ethernet, tráfico en redes de área metropolitana, tráfico TCP/IP correspondiente a servicios HTTP (Web) o tráfico de video VBR correspondiente a estándares MPEG y M-JPEG.

La principal ventaja de trabajar con bases de datos conocidas es la posibilidad de tener un marco común de experimentación en el que se pueden cotejar los modelos obtenidos con los correspondientes a otras propuestas. Por su parte, el muestreo puede ser directo, analizando el tráfico que directamente fluye por el enlace, o indirecto, estudiando de una manera *off-line*, sobre un servidor, el volumen de información (archivos en servidores FTP, tamaño de correos electrónicos, páginas Web, películas MPEG,...) a la que los clientes tendrán acceso. En el primer caso (medida directa), la muestra se ve fuertemente determinada por la torre de protocolos de la red concreta que se muestrea. En el segundo caso, el muestreo se produce a nivel de aplicación, por lo que el modelo resultante sería independiente de los niveles inferiores.

Otro aspecto clave a la hora de definir un modelo es decidir la escala de tiempos a la que el modelo debe representar convenientemente la realidad. A menudo resulta completamente inútil y costoso tratar de modelar las fluctuaciones que se producen a todas las escalas en cierto tráfico. La aplicación final del modelo establece una escala de interés sobre la que debe centrarse el esfuerzo modelador y que determina los factores claves a imitar y los que pueden ser obviados. Así, si el problema reside, por ejemplo, en dimensionar una red WAN para períodos de meses o años, el modelo debe ser capaz de predecir la evolución del mercado. Si, por el contrario, el dimensionado ha de hacerse a más corto plazo, tratando de analizar la hora cargada al cabo de un día, el elemento central es el usuario, mientras que para controles destinados a gestionar la calidad de la aplicación, ésta se puede convertir en el factor clave. Así mismo, si el control a diseñar trabaja a escalas de tiempo puede resultar fundamental modelar con cuidado los efectos que imponen los niveles inferiores de la torre de protocolos.

El objetivo básico de un modelo de tráfico es ser capaz de imitar el “comportamiento” del tráfico real. Sin embargo, esta cualidad (la correspondencia con la realidad) no es el único concepto que debe considerarse a la hora de elegir una estrategia de modelado. Así, resultan de especial interés otros aspectos como:



Márgenes de actuación a grandes rasgos de los factores genéricos de influencia en el tráfico telemático

- La tratabilidad analítica del modelo; esto es, la capacidad intrínseca del modelo de arrojar resultados analíticos, sin necesidad de simulación.
- La facilidad de implementación, ya sea vía software para proceder a simulaciones o vía hardware con el objetivo de disponer de un generador de tráfico sintético en tiempo real. Esta característica exige simplicidad no sólo a la propia estructura del modelo, sino también a la algoritmia y los costes computacionales que exige su implementación.
- La “parsimonia” del modelo. Con este concepto, que recurre al significado original de la palabra parsimonia, se describe la capacidad del modelo para ser descrito mediante un conjunto reducido de parámetros. De igual importancia, resulta la posible significación física de los parámetros. Se debe ser consciente de que al usuario final hay que ofrecerle una representación de la realidad comprensible, flexible, pero con una variabilidad acotada y controlable mediante unos parámetros que posean una conexión evidente con la realidad (por ejemplo: velocidad máxima, velocidad de pico, tamaño de ráfaga...).
- Capacidad de modelar otros tipos de tráfico, en especial tráfico agregado. En numerosas ocasiones se puede exigir de un modelo, en especial de los generalistas, la maniobrabilidad suficiente para que, mediante ciertos cambios de sus parámetros, sea capaz de imitar otros tipos de tráficos o incluso el propio tráfico agregado que resulta de multiplexar sobre un canal varias fuentes individuales.

Una manera razonable de clasificar los modelos de tráfico es mediante el análisis de la autocorrelación que presentan. La función de coeficientes de autocorrelación $R_x(k)$ de cierta se-

ñal $X[n]$ es un indicador del grado de dependencia lineal que existe entre una muestra $X[i+k]$ y la k -ésima muestra anterior:

$$R_x(k) = \frac{1}{\sigma_x^2} E[(X[i+k] - \mu_x)(X[i] - \mu_x)]$$

Donde $E[\bullet]$ es el operador esperanza matemática, mientras que σ_x^2 y μ_x indican, respectivamente, la varianza y la media de la señal $X[n]$, la cual, en el caso de modelado de tráfico, lleva aparejado la evolución en el tiempo de una variable física (por ejemplo: número de usuarios conectados, duración de una llamada, número de bits transmitidos durante cierto intervalo, tiempo entre paquetes, tamaño del paquete...).

De acuerdo con la forma en que el modelo impone a esta función, se puede distinguir entre:

- *Modelos no correlacionados o ruidos blancos*

Se trata de señales en donde no existe relación de ningún tipo entre las muestras de la secuencia ($R_x(k)=0$ para todo $k \neq 0$), de tal modo que cada muestra es generada a partir de la misma función de distribución estadística, pero con total independencia de las anteriores. Estos modelos estadísticamente independientes se han probado válidos para modelar ciertos aspectos de la generación del tráfico, entre los que podemos mencionar el tiempo entre llamadas, la duración de la llamada o el tamaño de los ficheros transmitidos. Dentro de este grupo, habría que incluir a modelos clásicos muy extendidos en todos los ámbitos, como el de Poisson (en donde el tiempo entre eventos se distribuye exponencialmente), o los ruidos normales o gaussianos (Casilari et al.).

- *Modelos sin memoria*

Los tiempos entre llegadas, T , son independientes y exponencialmente distribuidos, de manera

que el tiempo que toca esperar hasta ver la próxima llegada es independiente del instante en que se empiece a observar, esto se conoce como la «falta de memoria» de la distribución exponencial.

Esta propiedad también aparece en modelos de tiempo discreto en los que, en cada ranura de tiempo, la llegada de un paquete se da con probabilidad p , independientemente de otras ranuras. El número de llegadas en intervalos de n ranuras es una variable binomial (n, p) , mientras que el número de ranuras N que toca esperar hasta ver la llegada del próximo paquete es una variable geométrica que también carece de memoria (Alzate, Marco y Peña, Néstor).

- *Modelos de renovación*

Los procesos de renovación son una extensión de los modelos de tráfico sin memoria, en los que los intervalos de tiempo entre llegadas de paquetes son independientes e idénticamente distribuidos, aunque no necesariamente exponenciales o geométricos.

El tiempo que falta esperar hasta ver la llegada del próximo paquete depende de hace cuánto tiempo llegó el último paquete. Aunque esta memoria incrementa la complejidad analítica de los procesos generales de renovación con respecto a los procesos sin memoria, sigue siendo nula la correlación entre los tiempos entre llegadas de paquetes consecutivos, lo cual hace que estos modelos sigan siendo analíticamente tratables para estudiar el desempeño de los elementos de la red (Alzate, Marco y Peña, Néstor).

- *Modelos de dependencia a corto plazo*

El carácter intermitente (o por ráfagas) del flujo de datos impuso desde un principio la necesidad de introducir correlación en los modelos de

tráfico, ya que la llegada de paquetes se encuentra fuertemente correlacionada. Las primeras soluciones de este tipo pasaron por sencillas cadenas markovianas de pocos estados. En un proceso markoviano, la correlación viene dada por el hecho de que el estado actual del sistema depende exclusivamente del estado anterior. En este grupo de procesos markovianos habría que incluir un modelo tan extendido como los procesos On-Off, aproximación simple y razonable de las fuentes de datos y voz. Una fuente On-Off consta de dos estados. Durante el estado On se emite tráfico a tasa constante mientras que durante el estado Off (período de silencio en el caso de la voz) la fuente permanece ociosa. Los tiempos de permanencia en cada estado en principio son exponenciales, aunque otras distribuciones son posibles (procesos semimarkovianos).

Otros modelos muy conocidos los constituyen los procesos modulados por Markov, en especial el MMPP (Markov Modulated Poisson Process) de dos estados. De acuerdo con este modelo, apropiado para modelar el agregado de fuentes On-Off, la generación de tráfico evoluciona entre dos estados con generación poissoniana, pero de tasa media distinta. Por el contrario, la necesidad de definir una matriz con las probabilidades de conmutación entre estados hace que el número de parámetros crezca geométricamente con el número de estados considerado. En ese caso, y a costa de perder la mencionada tratabilidad, se puede recurrir a los filtros ARMA (Auto Regressive Moving Average), en los que el valor de la muestra actual $X[n]$ se obtiene a partir de N_a valores anteriores de $X[n]$ y N_b valores de un ruido blanco aleatorio $w[n]$ denominado proceso de innovación, del modo

$$X_{ARMA}[n] = a_o + \sum_{i=1}^{N_a} a_i \cdot X_{ARMA}[n-i] + \sum_{i=0}^{N_b} b_i \cdot w[n-i]$$

Donde N_a y N_b se corresponden con los órdenes de la parte AR y MA, respectivamente, y los coeficientes a_i y b_i son diseñados para imitar los primeros coeficientes de autocorrelación (Casilari et al.).

Los filtros ARMA (Procesos Autorregresivos de Medias Móviles) son la unión del Proceso AR (Autorregresivo), el cual es un filtro todo polos y cuya ecuación en diferencias que relaciona la entrada y la salida es:

$$x(n) + \sum_{i=1}^{N_a} a_i \cdot x(n-i) = w(n)$$

Proceso de MA (Medias Móviles), el cual es un filtro todo ceros y cuya ecuación en diferencias para la relación entrada-salida es:

$$x(n) = \sum_{i=0}^{N_b} b_i \cdot w(n-i)$$

Los procesos ARMA son entonces un filtro lineal que tiene polos y ceros finitos en el plano Z (Proakis, Jhon G. y Manolakis, D.).

Otro modelo autorregresivo es el modelo TES (Transform Expand Sample). Este modelo, además de invertir la función de distribución acumulativa de los tiempos entre llegadas, consiste en generar probabilidades de éstas mediante números pseudoaleatorios correlacionados, opera con aritmética módulo 1 y no es lineal. Por lo tanto, con este modelo se pueden ajustar la autocorrelación de muestras reales de tráfico y la distribución marginal de los tiempos entre llegadas (Alzate, Marco y Peña, Néstor).

En cualquier caso, los procesos markovianos y los filtros ARMA, así como alguna otra solución más compleja que se podría englobar en este grupo, como los procesos TES, tienen en común su incapacidad para aproximar la función de

autocorrelación para valores altos del retraso k . Efectivamente, estos procesos se diseñan para ajustar las dependencias a corto plazo o SRD (Short Range Dependence) y presentan indefectiblemente una caída exponencial de la función de autocorrelación (Casilari, A. et al.):

$$R_x(k) \propto \exp(-k); \quad \text{cuando } k \rightarrow \infty$$

- *Modelos de dependencias a largo plazo*

Las LRD, que se reflejan estadísticamente en coeficientes de autocorrelación con un decaimiento lento o hiperbólico, se asocian a la existencia de una variabilidad del tráfico en una amplia gama de escalas temporales que los modelos tradicionalmente empleados no son capaces de recoger y que pueden tener consecuencias graves para muchos esquemas de control, donde se supone una escala "natural" y acotada, en donde se manifiesta el carácter rafagueante del tráfico.

Por su parte, esta variabilidad de la señal se liga a cierto grado de autosemejanza o fractalidad que ha permitido aplicar en este campo las conclusiones y la analítica desarrollada por Mandelbrot y otros autores que han estudiado procesos fractales en otras ramas de la ciencia. Las LRD han venido a revolucionar, al menos parcialmente, el ámbito del modelado de tráfico telemático, ya que también se han detectado en otros tipos de tráfico: servicios Web, tráfico IP, tráfico de video, tráfico de enlaces metropolitanos y de red de área extensa. El argumento habitualmente esgrimido para explicar las LRD en series temporales es la existencia de una estructura multinivel de generación. Así, para justificar la LRD del tráfico Web se apela a la interacción del comportamiento del usuario (preferencias por ciertos ficheros, tiempo de análisis de páginas Web,...) con las políticas de caché de

los buscadores, el tamaño subexponencial de los objetos en las páginas Web y el agregado de tráfico que resulta en la LAN. En el caso de las secuencias de video, las LRD se suelen fundamentar por la existencia de una escala de tiempo superior cuya duración se ha probado subexponencial. Efectivamente se puede probar que un proceso cuyos niveles de actividad (tasa media de generación) oscilan de acuerdo con una temporización de colas pesadas presentará características LRD. Hasta la fecha existen dos técnicas genéricas en las que se podrían agrupar los modelos planteados para aproximar la naturaleza LRD de una señal de tráfico:

- *Modelos multinivel.* En estos modelos se imita directamente la variabilidad existente a una serie de niveles o escalas de tiempo. Cada nivel suele modelarse mediante un proceso "clásico" (habitualmente cadenas de Markov) que modula la generación a niveles inferiores. Su variabilidad en ese caso se limita a las escalas contempladas (modelos "quasi-fractales") presentando el inconveniente de que ofrecen estructuras complicadas, de escasa tratabilidad, cuya definición requiere un número amplio de parámetros, lo que perjudica la parsimonia del modelo.
- *Modelos autosemejantes.* Estos modelos ya responden a estructuras estrictamente fractales, es decir, de variabilidad a todas las escalas de tiempo. Se corresponden con modelos generalistas bien conocidos, como los ruidos gaussianos fraccionarios o FGN (Fractional Gaussian Noise), o los procesos ARMA integrados fraccionarios, o FARIMA. Mientras que los primeros generan señales fractales puras (con la misma forma de distribución estadística con independencia de la escala temporal desde la que son contemplados), los segundos son capaces de añadir el ajuste de las dependencias a corto plazo, o SRD, incorporando una típica es-

tructura autorregresiva de media móvil (ARMA). La principal ventaja de estos modelos, frente a la aproximación desarrollada por los del grupo anterior, es que son capaces de representar la LRD mediante un único valor, el parámetro H o de Hurst, que denota el grado de autosemejanza de la muestra a modelar.

$$R_{LRD}(k) \propto k^{-\beta}; \text{ cuando } k \rightarrow \infty \\ \text{con } 0 < \beta < 1$$

De esta manera, estos modelos presentan un alto grado de parsimonia. Por el contrario, su principal desventaja radica precisamente en la dificultad de una estimación ajustada de H, el cual es un valor acotado entre 0.5 y 1. Esta dificultad se explica por la propia naturaleza de los estimadores propuestos, los cuales se basan en la medición de comportamientos asintóticos o en la minimización de funciones que suponen previamente una forma al espectro y que introducen siempre cierto sesgo en la medida, ya de por sí limitada a un estrecho intervalo. Esta problemática se agrava cuando estos diversos estimadores se aplican sobre series de tráfico real finitas y con posibles problemas de no estacionariedad. A esta dificultad hay que añadir la complejidad computacional de la propia estimación, así como la complejidad de generación que la ejecución rigurosa de estos métodos requeriría, aunque se debe apuntar que existen soluciones aproximadas (que, por ejemplo, acotan el número de coeficientes a considerar) que pueden solventar parcialmente este problema a efectos prácticos (Casilari et al.).

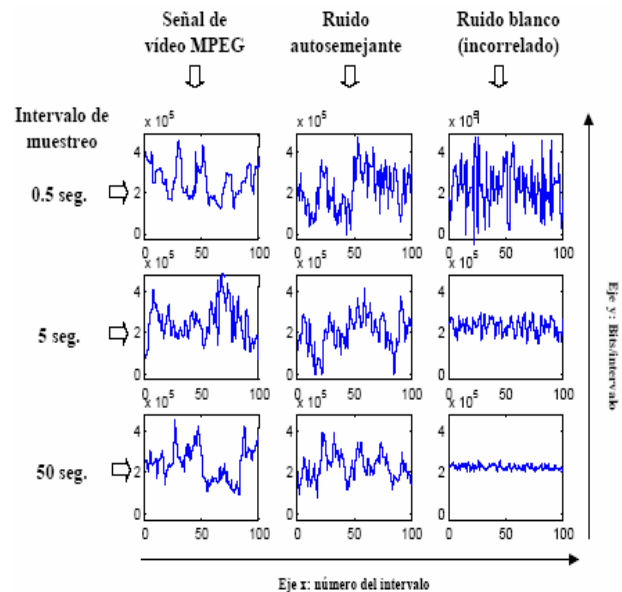
De esta manera, la aproximación a la naturaleza LRD exige decidir en un compromiso a favor de la parsimonia o de la complejidad de medida y

generación. En cualquier caso, hay que reseñar que el éxito de los modelos fractales vendrá determinado por las consecuencias prácticas que en la realidad posee la existencia de la LRD en el tráfico telemático.

Así, uno de los principales efectos de cargar un multiplexor con tráfico autosemejante o fractal es el hecho de que las pérdidas decaen muy lentamente conforme aumenta el buffer de contención; en oposición, los modelos clásicos (incorrelados o SRD) aumentar el buffer provoca caídas exponenciales de las pérdidas. Esto puede implicar la inutilidad de las técnicas de buffering al tratar con tráfico fractal, ya que aumentar el buffer provoca incrementos en el retraso sin obtener ganancias significativas en cuanto a pérdidas. Sin embargo, son muchos los autores que opinan que este fenómeno puede ser omitido, pues sólo comienza a ponerse de manifiesto con tamaños de buffer muy grandes, intolerables de por sí para muchos tipos de servicios, especialmente todos los que conllevan transmisiones en tiempo real, como muchos servicios de video.

Otra problemática asociada a la fractalidad es su escasa tratabilidad analítica, así como la convergencia más lenta de las medidas de simulación que imponen estos procesos de variabilidad a largo plazo (las simulaciones con procesos fractales pueden llegar a ver su duración incrementada en varios órdenes para obtener resultados con la misma fiabilidad que los que se pueden conseguir con procesos SRD o no correlacionados similares) (Casilari et al.).

A modo de resumen, podría decirse que la aparición de los modelos fractales en el campo de la modelación estocástica del tráfico ha resultado fundamental, debido a su capacidad de exhibir dependencia a largo plazo sobre todas las escalas temporales mediante el uso de pocos parámetros (modelación "parsimoniosa"). Como



consecuencia de ello, se ha abierto un amplio espectro de líneas de investigación que conducen, por un lado, a resolver nuevos problemas planteados en la aplicación de los modelos fractales en el análisis de prestaciones; y, por otro lado, a disponer de resultados y argumentos que revelen el impacto de la estructura de correlación (a largo y a corto plazo) sobre las prestaciones de las redes (López, José).

La autosimilitud estocástica, al igual que la determinística, se puede ilustrar gráficamente, como se ve en la figura anterior con la señal de video MPEG y con el ruido autosemejante. Considerando que el tráfico es una traza muestral de un proceso estocástico y restringiendo la similitud a ciertas estadísticas de las series de tiempo reescalizadas, se puede apreciar fácilmente autosimilitud exacta en los objetos matemáticos abstractos y autosimilitud aproximada en las realizaciones específicas (Alzate Monroy, Marco Aurelio).

Para medir la autosimilitud estocástica, se pueden utilizar las estadísticas de segundo orden que capturan la variabilidad de los procesos. De hecho, la invarianza a la escala se puede definir

en términos de la función de autocorrelación, pues el decrecimiento polinómico (en vez de exponencial) de esta función es la manifestación de la "dependencia de largo rango" que, en una importante clase de procesos, es equivalente a la autosimilitud.

Aunque los fenómenos de autosimilitud en el tráfico de las actuales redes de comunicaciones se descubrieron hace una década aproximadamente, en nuestro medio no se ha abordado formalmente el tema, pues todavía se le considera como un aspecto muy avanzado de la teleinformática. Sin embargo, los efectos que este fenómeno produce en el desempeño de las redes exigen un estudio inminente del tema. Si se quiere hacer un estudio serio del tema, se debe ubicar al nivel de abstracción matemático que se necesita; por eso, la presentación de los conceptos no puede dejar de ser formal (Alzate M., Marco Aurelio).

Definición en tiempo continuo

Una definición común de procesos estocásticos en tiempo continuo está basada en un escalamiento directo de la variable de tiempo continuo. Un proceso estocástico $x(t)$ es estadísticamente autosimilar con parámetro H ($0.5 \leq H \leq 1$) si por cualquier real $a > 0$, el proceso $a^H x(at)$ tiene las mismas propiedades que $x(t)$. Esta relación puede ser expresada por las siguientes tres condiciones:

- $E[x(t)] = \frac{E[x(at)]}{a^H}$ media
- $Var[x(t)] = \frac{Var[x(at)]}{a^{2H}}$ varianza
- $R_x(t, s) = \frac{R_x(at, as)}{a^{2H}}$ autocorrelación

El parámetro H , conocido como el parámetro de Hurst, es la llave para medir la autosimilitud. Más precisamente, H es una medida de la persistencia de un fenómeno estadístico y es una medida de la longitud de la dependencia de rango largo en un proceso estocástico (Stallings, W., 1998) Un valor de $H=0.5$ indica la ausencia de autosimilitud. El valor más alto de H es 1, el mayor grado de persistencia de una dependencia de rango largo y la aleatoriedad desaparece. La condición $H < 0.5$ es artificialmente irreal, y con $H > 1$ se pierde la estacionariedad del proceso (Alzate M., Marco Aurelio, 2001).

Como un ejemplo basado en esta definición, se pasa a considerar el proceso de Movimiento Browniano Fraccional (FBM), que es una generalización del proceso de movimiento browniano. El proceso FBM es frecuentemente usado en el análisis del tráfico de datos autosimilar.

$$B_H(t) = Xt^H \quad (t > 0; 0.5 \leq H < 1)$$

Donde X es una variable aleatoria normalmente distribuida con media cero y varianza 1, y H es el parámetro del proceso. Así, el valor del proceso en el tiempo t es igual al valor de la variable aleatoria multiplicada por el intervalo de tiempo elevado por la potencia H .

Tenemos que $E[B_H(t)] = 0$ porque $E[X] = 0$. Igualmente, para cualquier variable aleatoria X y una constante a , $Var(aX) = a^2 Var(X)$. Por consiguiente, $Var[B_H(t)] = Var[t^H X] = t^{2H}$. Para $H=0.5$, el proceso FBM se reduce a un proceso de movimiento browniano ordinario.

La densidad de probabilidad de un movimiento browniano tiene la forma:

$$f_{B_H}(x, t) = \frac{1}{\sqrt{2\pi t^{2H}}} e^{-x^2 / 2t^{2H}}$$

Se puede ver que este proceso tiene incrementos estacionarios y que:

$$\text{Var}[B_H(t) - B_H(s)] = E[(B_H(t) - B_H(s))^2] = |t - s|^2$$

La autocorrelación de $B_H(t)$ es expresada como $R_{B_H}(t,s) = E[B_H(t)B_H(s)]$, y puede ser deducida de la siguiente manera:

$$E[(B_H(t) - B_H(s))^2] = E[B_H^2(t) + B_H^2(s) - 2B_H(t)B_H(s)]$$

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

$$E[B_H(t)B_H(s)] = \frac{1}{2} (\text{Var}[B_H(t)] + \text{Var}[B_H(s)] - \text{Var}[B_H(t) - B_H(s)])$$

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

Nuevamente, con $H=0.5$ se reduce la autocorrelación a la de un proceso de movimiento browniano.

Para ver que el proceso FMB es autosimilar, se debe satisfacer las tres condiciones anteriores. Considerando el proceso $B_H(at) = X(at)^H$, es fácil ver que $E[B_H(at)] = a^H E[B_H(t)] = 0$, lo que satisface la primera condición [Stallings, W., 1998].

$\text{Var}[B_H(at)] = \text{Var}[X(at)^H] = (at)^{2H} \text{Var}[X] = (at)^{2H}$, lo que satisface la segunda condición. La tercera condición se cumple de la siguiente forma:

$$R_{B_H}(at, as) = \frac{1}{2} ((at)^{2H} + (as)^{2H} - |at - as|^{2H})$$

$$R_{B_H}(at, as) = \frac{a^{2H}}{2} (t^{2H} + s^{2H} - |t - s|^{2H})$$

$$R_{B_H}(at, as) = a^{2H} R_{B_H}(t, s)$$

Considerando la correlación entre los incrementos desde $-t$ hasta 0 y los incrementos desde 0 hasta t :

$$E[(B_H(0) - B_H(-t))(B_H(t) - B_H(0))] = E[-B_H(-t)B_H(t)]$$

$$E[(B_H(0) - B_H(-t))(B_H(t) - B_H(0))] = -\frac{1}{2} \left((-t)^{2H} + t^{2H} - |-t-t|^{2H} \right)$$

$$E[(B_H(0) - B_H(-t))(B_H(t) - B_H(0))] = \frac{1}{2} (2t)^{2H} - t^{2H}$$

Para $H=0.5$, la correlación de los incrementos pasados y futuros desaparece, como se requiere para un movimiento browniano que tiene incrementos independientes. Pero, para $H>0.5$, se tiene una característica apreciable de persistencia (Stallings, W., 1998).

Definición en tiempo discreto

Para una serie de tiempo estacionaria x , se considera el "proceso agregado" con nivel de agregación m , $X^{(m)}$ obtenido del proceso original X mediante la siguiente suma normalizada:

$$X_k^{(m)} = \frac{1}{m} \sum_{i=km-(m-1)}^{km} x_i$$

Una forma de ver el proceso de agregación es como una técnica para compresión de la escala de tiempo. Esto es, se hace una partición de $X(t)$ en bloques no sobrelapados de tamaño m y se promedian los valores de cada bloque. Se puede considerar $X^{(0)}$, ser la más alta manifestación o la mayor resolución posible para estas series de tiempo. Por ejemplo, el proceso $X^{(3)}$ es el proceso reducido en magnificación por un factor de 3. Si las estadísticas del proceso (media, varianza, correlación, etc.) se preservan con la compresión, entonces estamos tratando con un proceso autosimilar.

Para un proceso ergódico, el tiempo promedio puede ser igual al promedio total, y la varianza del tiempo promedio puede ir a cero relativamente rápido a medida que m se vuelve más grande. Esto no pasa para un proceso autosimilar; la varianza también va a cero, pero lo hace mucho más lentamente que un proceso ergódico estacionario.

Un proceso X se dice que es exactamente autosimilar con parámetro β ($0 < \beta < 1$), si para todo $m = 1, 2, \dots$ se tiene:

$$\text{Var}(x^{(m)}) = \frac{\text{Var}(x)}{m^\beta} \quad \text{varianza}$$

$$R_{x^{(m)}}(k) = R_x(k) \quad \text{autocorrelación}$$

El parámetro β está relacionado con el parámetro de Hurst, definido como $H=1-(\beta/2)$. Para un proceso estacionario ergódico, $\beta=1$ y la varianza del tiempo promedio decaen a cero en una tasa de $1/m$. Para un proceso autosimilar, la varianza del tiempo promedio decae más suavemente.

Se dice que un proceso X es asintóticamente autosimilar si para todo valor de k :

$$Var(x^{(m)}) = \frac{Var(x)}{m^\beta} \quad \text{varianza}$$

$$R_{x^m}(k) \rightarrow R_x(k) \quad m \rightarrow \infty \quad \text{autocorrelación}$$

Con estas definiciones de autosemejanza, la autocorrelación del proceso agregado tiene la misma forma que el proceso original. Esto sugiere que el grado de variabilidad o de ráfaga puede ser el mismo a diferentes escalas.

Una interesante característica de la definición anterior es que la autocorrelación del proceso agregado autosimilar no va a cero cuando $m \rightarrow \infty$. Esto en contraste con los procesos estocásticos típicamente usados para el modelo de paquetes de datos, que tiene la siguiente relación:

$$R^{(m)}(\tau) \rightarrow 0 \quad m \rightarrow \infty$$

Una autocorrelación de la función $R(\tau)$, que es igual a cero, es consistente con el ruido blanco. En la figura anterior se puede ver que a medida que el nivel de agregación incrementa (incrementa m) no se tiene una estructura similar. Otra interesante característica es que la varianza $x^{(m)}$ decrece más suavemente que $1/m$ cuando $m \rightarrow \infty$; esto es, que decrece proporcionalmente a $1/m^\beta$ (Stallings, W., 1998).

Dependencia de rango largo

Una de las propiedades más significativas de los procesos autosimilares es referida a la dependencia de rango largo. Esta propiedad es definida en términos de la conducta de la autocovarianza $C(\tau)$ cuando τ incrementa.

Para muchos procesos, la autocovarianza decae rápidamente con τ . Por ejemplo, para un proceso de Poisson con incremento L y media λ , la autocovarianza para valores de $\tau > L$ es:

$$C(\tau) = R(\tau) - \lambda^2 = \lambda^2 - \lambda^2 = 0$$

En general, un proceso de dependencia de rango corto satisface la condición de que su autocovarianza decae exponencialmente:

$$C(k) \sim a^{|k|} \quad |k| \rightarrow \infty, \quad 0 < a < 1$$

Donde \sim denota que la expresión a ambos lados es asintóticamente proporcional a la otra.

Los tipos de modelos de tráfico típicamente utilizados en la literatura emplean sólo modelos de dependencia de corto rango, usando la ecuación:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad |x| < 1$$

Se puede ver que la sumatoria $\sum_k C(k)$ es finita.

En contraste, un proceso de dependencia de rango largo tiene una autocovarianza que decae hiperbólicamente:

$$C(k) \sim |k|^{-\beta} \quad |k| \rightarrow \infty, \quad 0 < \beta < 1$$

Donde β es el mismo parámetro definido anteriormente y que está relacionado con el parámetro de Hurst (Stallings, W., 1998).

La dependencia de rango largo refleja la persistencia del fenómeno en procesos autosimilares (Stallings, W., 1998). Hay diferencia conceptual entre autosimilitud y dependencia de rango largo. Hay procesos autosimilares que no poseen dependencia de rango largo y viceversa. Sin embargo, bajo la restricción $0.5 < H < 1$, existe una doble implicación entre autosimilitud y dependencia de rango largo; por eso, muchas veces se usan como conceptos idénticos (Alzate, M., Marco Aurelio).

Densidad espectral

Una formulación equivalente de la dependencia de rango largo se puede dar en el dominio de la frecuencia. Específicamente la densidad espectral de potencia obedece a la ley de potencia cerca del origen:

$$S(\omega) \sim \frac{1}{|\omega|^\gamma} \quad \omega \rightarrow 0, \quad 0 < \gamma < 1$$

Donde $\gamma = 1 - \beta = 2H - 1$. Esta divergencia de la densidad espectral de potencia cerca al origen indica una importante contribución de los componentes de baja frecuencia.

La densidad espectral para un proceso estocástico en tiempo discreto es:

$$S(\omega) = \sum_{k=-\infty}^{\infty} R(k) e^{-j2k\omega} \quad S(0) = \sum_{k=-\infty}^{\infty} R(k)$$

En contraste, los procesos de dependencia de rango corto son caracterizados por una densidad espectral finita cuando $\omega > 0$. Esto ocurre cuando $\gamma = 0$ ó $H = 0.5$ (Stallings, W., 1998).

Distribuciones de colas pesadas

Las distribuciones de cola pesada pueden ser usadas para caracterizar densidades de probabilidad que describen el proceso de tráfico como paquetes a intervalos de tiempo y longitud de ráfagas.

La distribución de una variable aleatoria X se dice que es de cola pesada, si:

$$1 - F(x) = \Pr[X > x] \sim \frac{1}{x^\alpha} \quad x \rightarrow \infty, \quad 0 < \alpha$$

En general, una variable aleatoria con distribución de cola pesada exhibe una alta varianza o incluso infinita varianza.

La distribución de cola pesada más simple es la distribución de Pareto con parámetros k y α ($k, \alpha > 0$), con una densidad y una función de distribución:

$$f(x) = F(x) = 0 \quad (x \leq k)$$

$$f(x) = \frac{\alpha}{k} \left(\frac{k}{x}\right)^{\alpha+1} \quad F(x) = 1 - \left(\frac{k}{x}\right)^\alpha \quad (x > k; \alpha > 0)$$

Y el valor de la media es:

$$E[x] = \frac{\alpha}{\alpha - 1} k \quad (\alpha > 1)$$

El parámetro k especifica el mínimo valor que la variable aleatoria puede tener. El parámetro α determina la media y la varianza de la variable aleatoria: si $\alpha \leq 2$, entonces la distribución tiene varianza infinita; y si $\alpha \leq 1$, tiene media y varianza infinitas.

La distribución de Pareto ha sido observada en una amplia variedad de fenómenos de las ciencias sociales, físicas y de comunicaciones (Stallings, W., 1998).

- *Modelos FARIMA*

Si la salida de un filtro ARMA se integra multiplicando su transformada Z con un integrador $(1-z^{-1})^d$, donde $d \in (-0.5, 0.5)$, se obtiene un pro-

ceso autorregresivo integrado de promedio móviles (FARIMA).

En particular, si d se encuentra en el intervalo abierto $(0, 1/2)$, los procesos FARIMA poseen propiedades LRD. Además de eso, poseen simultáneamente propiedades SRD debido a que se pueden descomponer en un proceso ARMA y otro FARIMA. El correspondiente proceso acumulativo es autosemejante con $H = d + 0.5$ (Alzate, M. y Peña, N.).

- *Modelos wavelet multifractales*

El análisis Wavelet es un método eficaz para el estudio de procesos autosemejantes, esto se debe a que está orientado al estudio multireso-

lución de señales, como la lectura de encefalogramas, con lo cual se puede analizar el comportamiento de la señal a diferentes escalas de tiempo simultáneamente.

Por medio de la transformada Wavelet se puede transformar un proceso autosimilar con una estructura de correlación compleja en una secuencia de procesos gaussianos, independientes entre sí y no autocorrelacionados.

La ventaja que presenta la Transformada Wavelet es que es posible hacer que la varianza de los coeficientes Wavelet en cada escala conserven la información del grado de autosemejanza del proceso, analizado todo esto por medio del procesamiento digital de señales multitasas.

Gracias a los procesos Wavelet, específicamente el Modelo Wavelet Multifractal (MWM), es posible hacer una sinterización de los procesos multifractales en el dominio de la escala (Alzate, M. y Peña, N.).

Conclusiones

Se presenta en este trabajo una introducción a los modelos de tráfico, describiendo las características que éstos deben tener para que sean modelos tratables desde todo punto de vista, y haciendo una caracterización de estos dependiendo de sus cualidades. Inherentemente a estas cualidades se ve la evolución de estos modelos de tráfico, y precisamente por esto se hace mayor hincapié en los modelos de dependencia de rango largo, ya que los modelos tradicionales son insuficientes para caracterizar el tráfico en las actuales redes de datos. Se introduce a los principales conceptos y modelos que son una manifestación de la complejidad de las redes de telecomunicaciones que actualmente el hombre ha desarrollado gracias a su ingenio investigador.

Se pretende con este trabajo generar interés en este tema, que es un área de estudio nueva y que los ingenieros que se desempeñan en el área de las comunicaciones no pueden dejar pasar.

Referencias

ALZATE, Marco y PEÑA, Néstor. *Modelos de tráfico en análisis y control de redes de comunicaciones*. Universidad Distrital-Universidad de Los Andes.

ALZATE MONROY, Marco Aurelio. «Introducción al tráfico autosimilar redes de comunicaciones». En: *Revista Ingeniería*. Universidad Distrital Francisco José de Caldas, v.6, n.2, 2001.

ARACIL, J. "Características del tráfico en la Internet e implicaciones para el análisis y dimensionamiento de redes de ordenadores". En: *Novática*, n. 124, Noviembre/diciembre, 1996, pp. 18-26.

CASILARI, A.; REYES LECUONA, A.; DÍAZ, Estrella y SANDOVAL, F. *Modelado de tráfico telemático*. Universidad de Málaga, Departamento de Tecnología Electrónica, E.T.S.I. Telecomunicación.

LÓPEZ ARDAO, José Carlos. *Contribución al análisis del impacto de la correlación en las prestaciones de las redes de alta velocidad*. Universidad de Vigo: Departamento de Tecnologías de las Comunicaciones.

PROAKIS, John G. y MANOLAKIS, Dimitris G. *Tratamiento digital de señales*.

STALLINGS, W. *High speed networks TCP/IP and ATM design principles*. Prentice may, Inc., 1998.