

# Estudio De Autosimilitud En La Red Ethernet De La Fundación Universitaria San Martín Sede Bogotá Con Base En La Transformada Wavelet (Julio 2010)

Mauricio González Hernández – emyieich@hotmail.com

**Resumen**— Este artículo proporciona información sobre el desarrollo de un estudio acerca de la red Ethernet de la Fundación Universitaria San Martín FUSM sede Bogotá, en la cual se obtienen características acerca del tráfico presente en la misma, haciendo uso de la transformada Wavelet, para que estas sirvan como base para posteriores estudios y modelados de la red.

**Palabras Clave**— Autosimilitud, Transformada Wavelet, Teletráfico.

## I. INTRODUCCIÓN

DURANTE la última década diversos estudios han permitido descubrir la complejidad existente en el tráfico: dependencia de largo alcance y autosimilitud. Esto no era necesario en las redes únicamente telefónicas, redes de colas modeladas con tráfico Poisson, además de que se basan en muchas suposiciones, no se ajustan al tráfico visto en el mundo real. Igualmente con la llegada de Internet, esta distancia entre lo real y modelado se hace cada vez más amplia, debido a que se ofrecen servicios con diferente calidad de servicio como tráfico de voz, datos, video, multimedia.

Uno de los factores significantes en la evolución de las telecomunicaciones ha sido la globalización de las prestaciones, la entrada de Internet y una mayor conectividad, diferentes tipos de servicio, los cuales llevan a exigencias de calidad de servicio muy variadas. Este cambio hace de las redes actuales un objeto de estudio de gran interés.

Esta área de investigación es vital en la tecnología, la cual está en constante evolución, y cada vez más forma parte en un nivel social, ya que la ausencia de implementación de características autosimilares repercutiría y afectaría la globalización y conectividad del mundo y de las personas, factor cada vez más exigido

Ante esta insuficiencia en el modelamiento de tráfico se pretende encontrar una herramienta mediante la cual se logren extraer de forma adecuada las características autosimilares de las redes, se pretende dar respuesta a esta dificultad con ayuda de la transformada Wavelet, para de esta forma lograr obtener

patrones de red.

## II. MARCO REFERENCIAL

### A. Dependencia de Largo Alcance [1]

La dependencia de largo alcance indica que por pequeños que sean los valores obtenidos en la función de autocorrelación, al aumentar la escala su valor no es menospreciable, produciendo consecuencias significativas, en otras palabras, su función de autocorrelación es sumable

Estos modelos poseen memoria, el tráfico presente en cualquier instante de tiempo depende del tráfico anterior, lo que conlleva que para escalas de agregación altas estos procesos tenderán a comportarse igual que en escalas más bajas. Estos modelos permiten capturar las características del tráfico actual.

La distribución exponencial ejerce como frontera para los modelos dependiendo de su alcance, si la función de autocorrelación decae como una exponencial negativa será de corto alcance, en cambio si decae a una velocidad más lenta que la exponencial será de largo alcance.

### B. Tráfico Autosimilar [2]

La naturaleza autosimilar presente en las redes actuales es inherente a las características de los fractales, cuya propiedad es la de prevalecer su forma sin importar la escala en que se encuentre. El pionero de estos estudios es Benoit Mandelbrot, los cuales posteriormente fueron identificados en las redes Ethernet, dando inicio al estudio del tráfico autosimilar.

#### 1) Autosimilitud Exacta

Se denomina autosimilitud exacta cuando un determinado patrón se reproduce exactamente a distintas escalas, como en el caso de los fractales, se crea a partir de una serie temporal determinista

#### 2) Autosimilitud Estocástica

En un proceso estocástico estacionario las características estadísticas no cambian con el tiempo, son invariantes frente a traslaciones temporales: la media no depende del tiempo y la función de autocorrelación depende solamente de una diferencia temporal.

La autosimilitud estocástica se refleja para los procesos estocásticos cuando las estadísticas del proceso permanecen al

cambiar la escala temporal, tanto desde el punto de vista cualitativo como cuantitativo, es decir, el comportamiento estadístico del proceso a corto plazo es igual a su comportamiento estadístico a largo plazo

### C. Serie Agregada [3]

Un proceso estocástico discreto que está definido en instantes  $t$ , se define como  $x(t)=\{x_t, t=0,1,2,\dots\}$ . A partir de esta serie original, una serie temporal estacionaria  $x$ , se define la serie temporal  $m$ -agregada  $x^{(m)}=\{x_k^{(m)}, k=0,1,2,\dots\}$  mediante la suma de la serie original en bloques adyacentes y no superpuestos de tamaño  $m$ :

$$x_k^{(m)} = \frac{1}{m} \sum_{i=k m - (m-1)}^{k m} x_i$$

Ecuación 1. Serie Agregada

Esta serie agregada es una técnica para comprimir la escala temporal, siendo así  $x^{(1)}$  la mayor ampliación posible, es decir, la mejor resolución posible de la serie temporal.

#### 1) Autosimilitud asintótica de segundo orden

Para esta definición de autosimilitud la autocorrelación del proceso agregado tiene la misma forma que la autocorrelación del proceso original, lo cual apunta que la naturaleza de ráfagas es igual sin importar la escala. [4]

$$\text{Var}(x^{(m)}) \sim \frac{\text{Var}(x)}{m^\beta}$$

Ecuación 2. Varianza para autosimilitud asintótica de segundo orden

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

Ecuación 3. Autocorrelación para autosimilitud asintótica de segundo orden

Con estas definiciones, la autocorrelación del proceso agregado autosimilar no tiende a cero (0) cuando  $m \rightarrow \infty$ ; caso contrario para los procesos estocásticos tradicionales en donde la autocorrelación si tiende a cero (0) cuando  $m \rightarrow \infty$ . La varianza de  $x^{(m)}$  decrece más lentamente que  $1/m$  cuando  $m \rightarrow \infty$ , decrece proporcionalmente a  $1/m^\beta$ .

### D. Parámetro Hurst [2], [6]

El parámetro Hurst  $H$ , también denominado parámetro de autosimilitud, es la medida clave y más importante en procesos autosimilares. Hurst mide la persistencia de los fenómenos estocásticos, es una medida de la dependencia de largo alcance a través de la correlación de las variables, expresando la persistencia de las series de tiempo.

En el intervalo de  $0,5 < H < 1$ , la autocorrelación decae de forma hiperbólica, forma característica de los procesos autosimilares. En el modelamiento y diseño de tráfico autosimilar este es el rango de mayor interés.

Un método de estimación es el de la gráfica Varianza-Tiempo, se fundamenta en la propiedad de decaimiento lento de la varianza de la serie agregada del tráfico, basándose en la ecuación 4.

$$\log[\text{Var}(x^{(m)})] \sim \log[\text{Var}(x)] - \beta \log(m)$$

Ecuación 4. Gráfica Varianza-Tiempo

Un fragmento de la ecuación es una constante independiente de  $m$ ,  $\log[\text{Var}(x)]$ ; en consecuencia si se representa  $\text{Var}(x^{(m)})$  Vs.  $m$  en una gráfica logaritmo-logaritmo, se produce una recta de pendiente  $-\beta$ , mediante la cual se infiere  $H$ .

### E. Transformada Wavelet [5]

En comparación con otras transformadas (Laplace y Fourier) realiza un análisis simultáneo en frecuencia y tiempo, este método se basa en la estimación del espectro reflejado en el promedio de la magnitud al cuadrado de los coeficientes Wavelet.

La transformada Wavelet discreta DWT es el conjunto de coeficientes definidos mediante el producto interno de la señal con los conjuntos de funciones  $\phi_{j,k}(t)$ , la función de escala; y  $\psi_{j,k}(t)$  la Wavelet madre.

Es posible medir promedios temporales y usarlos como estimaciones para los convenientes promedios estadísticos. Se estima la varianza del proceso  $dX(j,:)$  mediante el uso del siguiente estimador [3], [6]:

$$u_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_x(j,k)|^2$$

Ecuación 5. Estimador varianza

Donde  $n_j$  es el número de coeficientes de la octava  $j$ . La varianza de  $u_j$  decrece con  $1/n_j$  y lo convierte en un estimador no polarizado y asintóticamente eficiente. Esto quiere decir que  $u_j$  puede expresarse como una forma eficiente de comprimir en un solo número el comportamiento de segundo orden de  $X$  en la octava  $j$ . Esto conlleva a que el exponente de escala  $\alpha$  puede extraerse de la pendiente de la gráfica de  $y_j = \log_2(u_j)$  contra cada octava  $j$ , esta gráfica se conoce como Diagrama LogEscala.

## III. DESARROLLO

### A. Banco de Muestras

La red de la Fundación Universitaria San Martín sede Bogotá cuenta con una conexión de 6 Mbps (6 Megabits por segundo) a Internet de banda ancha, a través del proveedor de servicios de Internet Telecom. Esta conexión se encuentra dividida entre sus 2 sedes en Bogotá: sede central calle 80 y sede secundaria calle 60; y la clínica Mariano Alberto Alvear Orozco ubicada en la carrera 11 con calle 93ª-38.

La red en la cual se centrará este estudio es la red de la sede central calle 80, la cual cuenta con mayor carga de usuarios, porque en ella se encuentra 13 de los 14 programas de modalidad presencial ofrecidos por la Fundación Universitaria San Martín sede Bogotá.

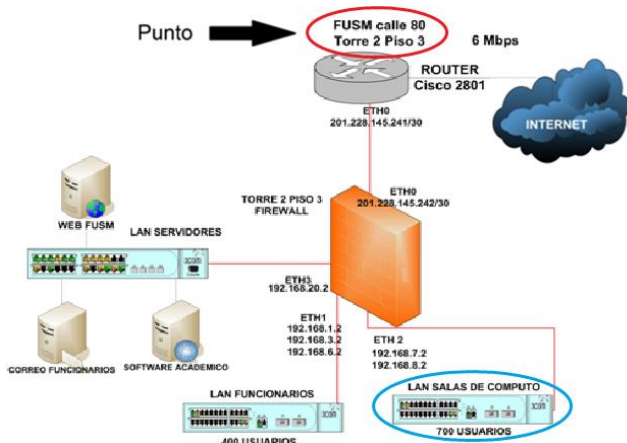


Fig 1. Topología Red FUSM

Se planeo efectuar la toma de la muestra a toda la red de la sede central por presentar el mayor índice de tráfico, consecuencia del elevado número de usuarios. Debido a que esta red está dividida en VLANs, de las cuales dos tienen privilegios, VLAN servidores y VLAN de funcionarios, consecuencia del tipo de datos que manejan en ellas (dirección de la página web, software académico, registros de notas, historial financiero, manejo de recursos y demás), por su carácter de importante y por cuestiones de seguridad en la División de redes y comunicaciones I&T de la Universidad, no pueden ser monitoreados bajo ningún concepto. El estudio de la red Ethernet de la Fundación Universitaria San Martín bajo los criterios de seguridad, es limitado únicamente a la VLAN que puede ser monitoreada, es decir, la VLAN de las salas de cómputo.

Al ser limitado el nivel de tráfico del estudio, el siguiente paso es la determinación del punto para la toma de las muestras, este punto se encuentra en la torre 2 piso 3 de la sede central, en este punto se encuentra ubicado un switch, encargado de la creación de las VLANs.

Ya determinado el punto para la toma de las muestras, un puerto del switch es configurado en modo espejo, en este modo el puerto refleja los datos presentes en la VLAN de las salas de cómputo a través de su puerto de enlace predeterminada (Gateway) para el muestreo, sin que afecten los datos que pasan ni el rendimiento de la misma.

Ya conectado al puerto en modo espejo del switch a través de un cable de conexión directa al puerto RJ 45 del computador, la toma de las muestras es realizada con el software Network Associates Sniffer Pro Versión 3.0. La primera parte es la selección de la NIC (Network Interface Card, Tarjeta de interfaz de red), al tener más de un adaptador. En la sección de "File" se elige "Settings", en esta parte se configura para tomar las muestras con Realtek PCIe FE Family Controller, es decir, por medio de una conexión de cable al puerto RJ-45.

El estudio se realiza con la medida de los paquetes presentes en la VLAN de las salas de cómputo, para poder realizar estas muestras se selecciona en la parte superior del menú de Sniffer Pro, la sección de "History Samples", para recolectar el correspondiente historial.

El punto de toma de las muestras ubicado en un recinto cerrado en la torre 2 piso 3 de la sede principal calle 80, es de acceso controlado por la División de redes y comunicaciones I&T, dificultando el horario para realizar el muestreo, ya que depende de la presencia de uno de sus miembros y por consiguiente, se tuvo que tener en cuenta su disponibilidad de tiempo. Tratando de tener una igualdad en cuanto al día y la hora para realizar el procedimiento, se tomaron las siguientes muestras con sus respectivos intervalos de tiempo.

Tabla 1. Banco de muestras

Número De Muestra	Fecha	Hora Inicio	Hora Final	Nombre Archivo
Muestra 1	16/Marzo/2010	11:30 a.m.	12:30 p.m.	Muestra1.txt
Muestra 2	18/Marzo/2010	12:04 p.m.	01:04 p.m.	Muestra2.txt
Muestra 3	24/Marzo/2010	12:04 p.m.	01:04 p.m.	Muestra3.txt
Muestra 4	25/Marzo/2010	12:13 p.m.	01:13 p.m.	Muestra4.txt
Muestra 5	26/Marzo/2010	12:10 p.m.	01:10 p.m.	Muestra5.txt

### B. Estudio de Características de Distribución Poisson

En esta parte se usó la herramienta de software Oracle Crystal Ball versión 11.1.1, es una suite sofisticada de simulación, previsión y herramienta de análisis de riesgos mediante el uso de hojas de cálculo. Este software se ejecuta junto con Microsoft Office Excel 2007. El caso particular para el cual se utilizó es para realizar pruebas de bondad de ajuste entre las muestras de la red Ethernet de la Fundación Universitaria San Martín y la función de distribución de Poisson, por medio de la prueba de Chi Cuadrado.

### C. Estudio de Dependencia de Largo Alcance y Autosimilitud

Para realizar el estudio de dependencia de largo alcance y autosimilitud se desarrollaron en MATLAB 7.10.0 (R2010a) dos programas los cuales se encargan de generar las series agregadas de las muestras (primer código) y determinación por método de regresión lineal de mínimos cuadrados en la Gráfica Varianza-Tiempo del parámetro de autosimilitud Hurst (segundo código).

#### 1) Serie Agregada

El código en MATLAB se encarga de generar las series mediante el uso de la ecuación 1 (Serie Agregada), la cual genera bloques adyacentes no superpuestos de diferentes tamaños variando el valor de  $m$ . Este procedimiento es una técnica con la cual se comprime la escala para su comparación con la serie original y con otras escalas, de esta forma se logra analizar el efecto de la fractalidad a diferentes niveles, para este caso el tiempo.

Cada nivel genera su propia varianza, con el uso de la ecuación 2 (Varianza Para Autosimilitud Asintótica De Segundo Orden) el parámetro de autosimilitud Hurst es calculado.

Tras realizar este procedimiento, se obtienen todas las varianzas para cada muestra, estos datos son guardados en una variable como un archivo de datos MATLAB (.mat), es decir, se guarda el resultado de los vectores que contienen las varianzas de cada una de las series agregadas, esto con el fin de un uso posterior.

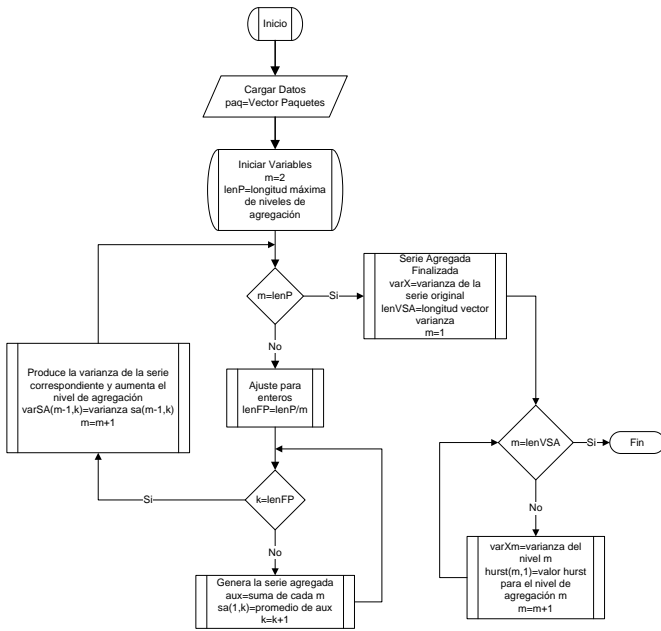


Fig 2. Diagrama de Flujo Serie Agregada

2) Parámetro Hurst

Con el código Serie Agregada se obtiene la varianza para cada nivel de agregación  $m$  ( $varSA\_T.mat$ ), con su correspondiente parámetro de autosimilitud Hurst. Sin embargo las muestras no pueden ser descritas con un diferente parámetro para cada nivel  $m$ , por esto al graficar Varianza Vs. Tiempo, es decir, cada varianza contra su respectivo  $m$  en una escala logarítmica todos los puntos tenderán a una línea recta.

Por ajuste de mínimos cuadrados se encuentran los valores de la recta que mejor se ajustan a estos puntos, la pendiente de esta recta tendrá un valor de  $-\beta$ , y empleando la relación Hurst-Beta se halla el valor del parámetro Hurst característica de cada muestra.

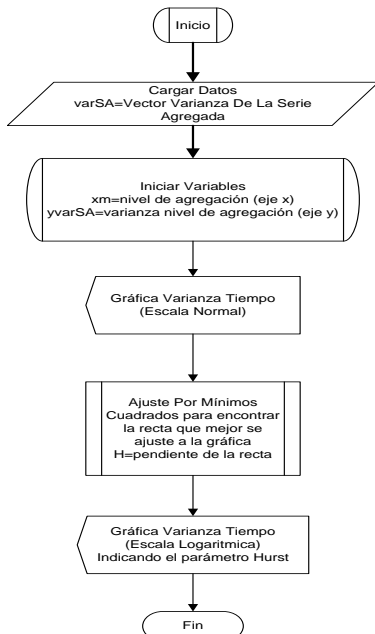


Fig 3. Diagrama de Flujo Parámetro Hurst

D. Uso de la Transformada Wavelet

Para el uso de la transformada Wavelet se hace uso de la ecuación 5. Lo primero es hacer un análisis multinivel Wavelet del cual para cada nivel se extraen los coeficientes de detalle, y con estos es posible sacar el Estimador Varianza  $y_j$ . Se gráfica el estimador varianza  $y_j$  contra sus correspondientes octavas y por método de mínimos cuadrados se halla la recta que mejor se ajuste a la gráfica, se hace una estimación variando las octavas.

Para este caso se empleo la Wavelet Haar, ya que tiene un buen índice de localización espacial y temporal por ser una Wavelet del tipo compacta. Se contrasta los resultados haciendo uso de la Wavelet db4.

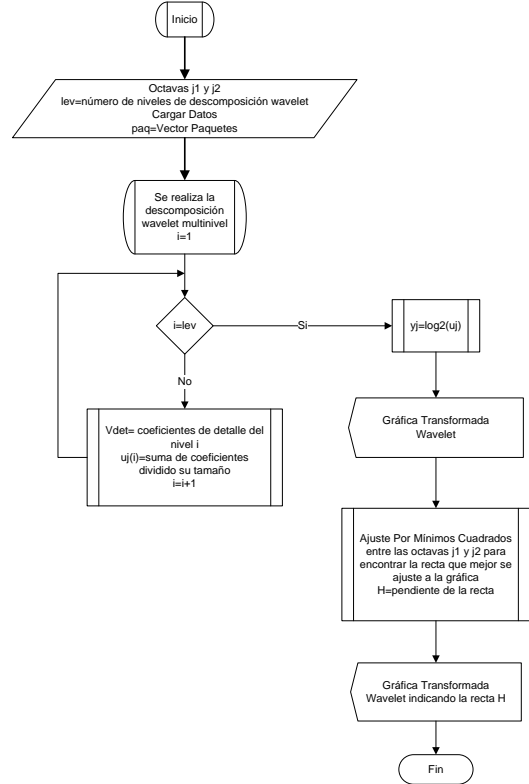


Fig 4. Diagrama de Flujo Transformada Wavelet

IV. RESULTADOS

A. Características Distribución Poisson

Gráficamente se puede observar que el mejor ajuste entre el valor observado de las muestras y el valor teórico hecho por la prueba de Crystal Ball no coinciden (se encuentra muy lejos la coincidencia de las gráficas). Con los valores estadísticos expuestos por Crystal Ball, se observa la diferencia entre las muestras y el mejor ajuste de distribución Poisson.

Tabla 2. Valores prueba bondad de ajuste

Muestra	Varianza Teórica	Varianza Real	Diferencia Varianza	Rate
Muestra 1	608.45	97826.95	97218,5	608,45
Muestra 2	654,71	147914,85	147260,14	654,71
Muestra 3	581,74	97691,70	97109,96	581,74
Muestra 4	571,58	99806,85	99235,27	571,58
Muestra 5	579,15	99844,09	99264,94	579,15

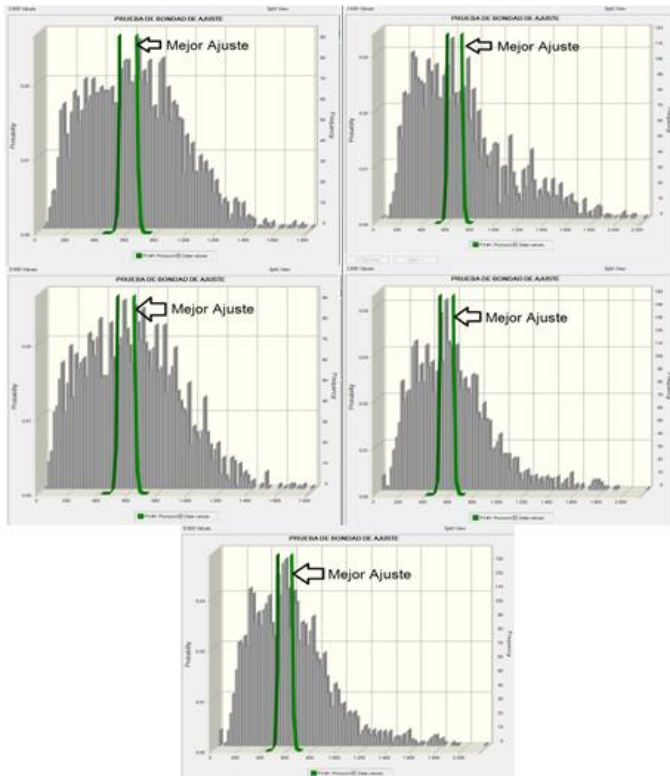


Fig 5. Prueba de Bondad De Ajuste. (Superior Izquierda Muestra 1, Superior Derecha Muestra 2, Media Izquierda Muestra 3, Media Derecha Muestra 4, Inferior Muestra 5)

La diferencia entre las varianzas es demasiado grande (valores entre 97218,5 y 28909,89) para que las muestras estén bajo la función de distribución de Poisson. El valor “Rate” expresa el grado en que el ajuste real concuerda a un ajuste teórico entre la distribución de Poisson y la prueba de bondad de ajuste, un valor igual a cero (0) significa una concordancia perfecta, valores mayores a 50 expresan que no hay concordancia entre los valores reales y teóricos.

La prueba de bondad de ajuste de Crystal Ball arroja valores de Rate exageradamente grandes, visto a través de las gráficas, es decir, ninguna de las muestras esta bajo una función de distribución de Poisson.

### B. Autosimilitud y Dependencia de Largo Alcance

En la figuras se muestran las obtenciones del parámetro Hurst por método gráfico Varianza-Tiempo, se observa la línea roja que hace el mejor ajuste para la respectiva muestra

Tabla 3. Parámetro Hurst por gráfica Varianza-Tiempo

Muestra	Parámetro Hurst
Muestra 1	0,93975
Muestra 2	0,51565
Muestra 3	0,65567
Muestra 4	0,62234
Muestra 5	0,62226

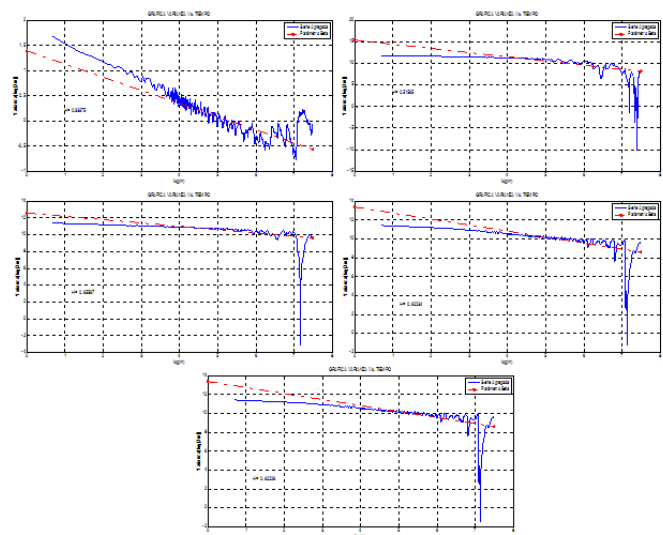


Fig 6. Obtención Parámetro Hurst por Método de Gráfica Varianza-Tiempo. (Superior Izquierda Muestra 1, Superior Derecha Muestra 2, Media Izquierda Muestra 3, Media Derecha Muestra 4, Inferior Muestra 5)

### C. Transformada Wavelet

Para el uso de la transformada wavelet se hace uso de la ecuación 5 (Estimador Varianza). Lo primero es hacer un análisis multinivel wavelet del cual para cada nivel se extraen los coeficientes de detalle (diferencia), y con estos es posible sacar el Estimador Varianza  $y_j$ .

Se gráfica el estimador varianza  $y_j$  contra sus correspondientes octavas y por método de mínimos cuadrados se halla la recta que mejor se ajuste a la gráfica, se hace una estimación variando las octavas.

Para este caso se empleo la wavelet Haar (parte superior de cada muestra) ya que tiene un buen índice de localización espacial y temporal (parte inferior de cada muestra). Se contrasta los resultados haciendo uso de la wavelet db4

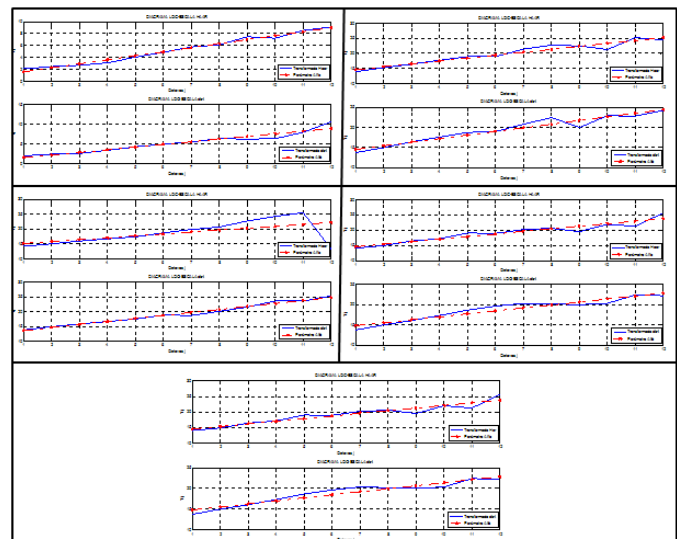


Fig 6. Obtención Parámetro Hurst por Transformada Wavelet. (Superior Izquierda Muestra 1, Superior Derecha Muestra 2, Media Izquierda Muestra 3, Media Derecha Muestra 4, Inferior Muestra 5)

Tabla 4. Parámetro Hurst por transformada Wavelet muestras

Muestra	Haar	Haar	db4	db4
	j1=1, j2=12	j1=1, j2=9	j1=1, j2=12	j1=1, j2=9
Muestra 1	0,8375	0,8369	0,8345	0,7995
Muestra 2	0,9818	0,6251	0,9466	0,9793
Muestra 3	0,8206	0,5153	0,5086	0,9490
Muestra 4	0,9212	0,8989	0,8630	0,9275
Muestra 5	0,9215	0,8987	0,8629	0,9274

El análisis Wavelet del banco de muestras refleja una fuerte característica autosimilar en las muestras, lo cual implica que el modelamiento de los recursos de la red debe ser enfocado para ajustarse a las características de dependencia de largo alcance con el fin de mejorar la eficiencia, generando una ganancia en el tiempo al disminuir los tiempos de servicio del usuario sin disminuir la calidad (incluso aumentándola).

## V. CONCLUSIONES

- La determinación del parámetro de autosimilitud Hurst se realiza mediante estimadores gráficos, métodos que hacen la aproximación a la recta que mejor se ajuste a los datos por regresión lineal de mínimos cuadrados, lo cual conlleva a la dificultad de escoger el rango en el cual se realice el ajuste.
- Las trazas obtenidas por los laboratorios Bellcore, BC-Aug89 y BC-Oct89, presentan dependencia de largo alcance, ya que por el método de regresión lineal de mínimos cuadrados obtenido por la gráfica Varianza-Tiempo tiende a seguir la pendiente característica de la estimación del parámetro Hurst; y matemáticamente por medio del estimador Wavelet los valores obtenidos se encuentran dentro del rango de autosimilitud  $0.5 < H < 1$ . Al obtener estos resultados por medio de los procedimientos descritos sirven como marco de referencia para el banco de muestras al permitir confiabilidad en los procedimientos.
- Las pruebas gráficas realizadas con MATLAB R2010a muestran características similares a través de diferentes escalas de tiempo y persistencia en el modelo de su función de autocorrelación, con lo que es posible deducir que existe algún tipo de persistencia en el tiempo, pero esta prueba no es suficiente para afirmar que tipo de dependencia está vinculada con el comportamiento de las muestras.
- La transformada wavelet es una herramienta utilizada en varios campos que van desde aplicaciones en la medicina hasta el procesamiento de imágenes, en este caso fue utilizada para el estudio de tráfico en una red Ethernet. Resultó ser adecuada para estos procesos pero presenta inconvenientes por tratarse de una herramienta de uso relativamente reciente, especialmente en el análisis multiresolución, donde sus estudios se remontan cerca del año 1995, debido al nivel matemático requerido y por el corto material disponible para su aplicación en el análisis de tráfico.
- Con el trabajo fue posible identificar las características del tráfico de una parte de la red Ethernet de la Fundación Universitaria San Martín sede Bogotá, descartando en un principio la característica de memoria nula. Se tiene el limitante de la cantidad baja de paquetes muestreados, sin embargo, se lograron establecer unos parámetros iniciales de

autosimilitud, con los cuales se puede diseñar y modelar la red de comunicaciones para optimizar su rendimiento, mediante la predicción de su comportamiento en largo tiempo

## REFERENCIAS

- [1] M. A. Alzate, Modelos de tráfico en análisis y control de redes de comunicaciones, Revista Ingeniería. vol.9, 2004, pp. 63–87.
- [2] W. Stalling, Redes en Internet de alta velocidad: Rendimiento y calidad de servicio. España, Pearson Prentice Hall, 2004.
- [3] M. A. Alzate, Uso de la transformada Wavelet para el estudio de tráfico fractal en redes de comunicaciones, Revista Ingeniería. vol.7, 2002, pp. 11-24.
- [4] W. Willinger, Self-Similar network traffic and performance evaluation, New Jersey, John Wiley & Sons, 2000.
- [5] J. Lévy-Véhel, Fractals in engineering: New trends in theory and applications, Londres, Springer Science+Business Media, 2005.
- [6] M. A. Alzate, Introducción al tráfico autosimilar en redes de comunicaciones, Revista Ingeniería. vol.6, 2001, pp. 6-17.