



Editorial de la Universidad  
Tecnológica Nacional

## ESTUDIO SOBRE LA DISTRIBUCIÓN DE TRÁFICO AUTOSIMILAR DE REDES ETHERNET

Santiago Pérez, Higinio Facchini y Luis Bisaro

Cátedra Redes de Datos – Departamento de Electrónica  
GRID TICs (Grupo UTN de Investigación y Desarrollo en TICs)  
Laboratorio de Análisis de Tráfico y Seguridad (LATyS)  
Universidad Tecnológica Nacional, Facultad Regional Mendoza

### RESUMEN

El análisis de colas ha sido y es de enorme utilidad para los diseñadores de redes y analistas de tráfico, a efectos de planificar las capacidades de los elementos de red y predecir su rendimiento. Además, los modelos exactos del tráfico Ethernet son importantes para modelar las aplicaciones de capa superior y los buffers de memoria de switches. Estos análisis dependen de la naturaleza Poisson del tráfico de datos. Sin embargo, muchos resultados predichos a partir del análisis de colas difieren significativamente del rendimiento observado en la realidad. Diversos estudios han demostrado que para algunos entornos el patrón de tráfico es autosimilar, en lugar de Poisson. Este concepto está relacionado con otros más conocidos como son los fractales y la teoría del caos. Desde principio de los años 90 se comenzaron a publicar documentos referidos a la autosimilitud del tráfico de Ethernet. En este documento se analiza la distribución intertrama de secuencias de tráfico Ethernet de contenido diverso. Usando capturas de tráfico Ethernet realizadas en distinto momento, se ha analizado la distribución Pareto como posible modelo, y se ha comparado la bondad de ajuste de este posible modelo de distribución, y de las muestras entre sí usando el estadístico Kolmogorov-Smirnov. El presente trabajo desarrolla sucesivamente los siguientes apartados: 1) Introducción, 2) Modelos de Tráfico de Red, 3) Tráfico de Datos Ethernet Automisilar con Distribución Pareto, 4) Caso de Estudio experimental de tráfico Ethernet, 5) Análisis con distribución Pareto, y 6) Análisis de bondad de ajuste con la prueba de Kolmogorov-Smirnov

Palabras Clave: tráfico, autosimilitud, Ethernet, Pareto, bondad de ajuste

2012

Editorial de la Universidad Tecnológica Nacional – edUTecNe

<http://www.edutecne.utn.edu.ar>

[edutecne@utn.edu.ar](mailto:edutecne@utn.edu.ar)

© [Copyright] La Editorial de la U.T.N. recuerda que las obras publicadas en su sitio web son de libre acceso para fines académicos y como un medio de difundir el conocimiento generado por autores universitarios, pero que los mismos y edUTecNe se reservan el derecho de autoría a todos los fines que correspondan.

## **INTRODUCCIÓN**

La necesidad para que las redes de comunicaciones puedan proveer un espectro creciente de servicios requiere técnicas eficientes de análisis, monitoreo, evaluación y diseño de las mismas. El análisis es tradicionalmente encarado de forma incompleta, y siempre, con crecientes demandas de los usuarios e incertidumbre acerca de la evolución de los sistemas de red. Por ello, para cumplir los requerimientos de los usuarios y para proveer garantías sobre la confiabilidad y adaptabilidad, deben desarrollarse modelos de sistemas que puedan capturar las características de la carga de red real y conducir a predicciones precisas de performance del sistema, en una cantidad razonable de tiempo.

El análisis de tráfico es un componente vital para entender los requerimientos y capacidades de una red. En los años pasados han sido propuestos incontables modelos de tráfico para entender y analizar las características de tráfico de las redes. No obstante, no hay un simple modelo de tráfico que pueda eficientemente capturar las características de tráfico de todos los tipos de redes, bajo cada circunstancia posible. Consecuentemente, el estudio de los modelos de tráfico para identificar eventualmente el mejor modelo es una tarea crucial. La modelación correcta de tráfico es también un requerimiento básico para la planificación de la capacidad exacta.

Si los modelos de tráfico subyacente no capturan eficientemente las características del tráfico real, el resultado puede ser una subestimación o sobreestimación de la performance de la red. Esto perjudicaría totalmente el diseño de la red. Los modelos de tráfico son por lo tanto, un componente central de cualquier evaluación de performance de las redes, y ellos necesitan ser muy exactos.

Los modelos de tráfico se analizan y comparan en base al número de parámetros requeridos para describir el modelo, la tratabilidad, la estimación de parámetros, y como mejor el modelo representa el tráfico real, conocido como bondad de ajuste. Por ello, para evaluar la bondad de ajuste, debería definirse métricas que permitan cuantificar cuán cerca el modelo está a los datos reales. Las métricas definidas deberían también estar directamente relacionadas a las medidas de performance que van a ser predichas desde el modelo de tráfico.

## **MODELOS DE TRÁFICO DE RED**

Los modelos de tráfico permiten a los diseñadores de red hacer suposiciones acerca de las redes a diseñar, basados sobre la experiencia pasada, y también la predicción de performance para requerimientos futuros. Los modelos de tráfico son usados de dos formas fundamentales: 1) Como parte de un modelo analítico, ó 2) Para activar una Simulación de Eventos Discretos. Algunos modelos característicos son el Modelo de Distribución Poisson y el Modelo de Distribución Pareto.

Uno de los modelos de tráfico más ampliamente usado y antiguo es el Modelo de Poisson. La distribución de Poisson sin memoria es el modelo predominante usado para analizar tráfico en redes de telefonía tradicional, y por extensión en las redes de datos, como discutido por Frots y Melamed (1994). En un proceso Poisson, los tiempos interarribos están distribuidos exponencialmente con un parámetro frecuencia  $\lambda$ :

$$f(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!},$$

La distribución tiene una media y una varianza igual al parámetro  $\lambda$ . La distribución Poisson puede visualizarse como una forma limitada de la distribución binomial, y también es usada ampliamente en los modelos de colas. La función de distribución de probabilidad y la función de densidad del modelo están dadas por:

$$F(t) := 1 - e^{-\lambda \cdot t}$$

$$f(t) := \lambda \cdot e^{-\lambda \cdot t}$$

El proceso de distribución Pareto, que es discutido en detalle posteriormente, produce tiempos interarribos independientes e idénticamente distribuidos. Una característica importante del modelo es que la distribución Pareto, del tipo hiperbólico, tiene valores grandes de varianza, respecto a Poisson, y varianza y media infinitas para ciertos valores de sus parámetros.

Aunque no hay un simple modelo que pueda ser usado efectivamente para modelar tráfico en todas las clases de redes, puede mostrarse que el modelo Poisson no responde bien para ciertos tipos tráfico. Por el contrario, por ejemplo, en casos de redes de alta velocidad, con demanda no esperada sobre las transferencias de paquetes, los modelos de tráfico basados en Pareo son excelentes candidatos, dado que el modelo tiene en cuenta la correlación de largo alcance en los tiempos de arribo.

## **TRÁFICO DE DATOS ETHERNET AUTOSIMILAR CON DISTRIBUCIÓN PARETO**

El artículo fundamental del estudio de los datos de tráfico autosimilar es «On the Self-Similar Nature of Ethernet Traffic» (La naturaleza autosimilar del tráfico de Ethernet), que posteriormente sería corregido y aumentado por Leland et al. (1994). Este documento contradujo la idea de que un simple análisis de colas basado en la suposición de que el tráfico fuera de Poisson pudiera modelar adecuadamente todo tráfico de red. Empleando una masiva cantidad de datos y un cuidadoso análisis estadístico, el artículo manifiesta que, para el tráfico de Ethernet, se requiere un nuevo planteamiento de modelado y de análisis.

Esto es, el tráfico Ethernet muestra cierta cantidad de picos o ráfagas. De este modo, el tráfico de Ethernet posee un aspecto similar para escalas grandes (horas y minutos) y para escalas pequeñas (segundos y milisegundos). Este tráfico autosimilar es muy diferente de lo que se observa en el tráfico telefónico, y en los modelos estocásticos usando Poisson de los análisis y diseño de redes de datos.

En una simulación Poisson, con alta resolución, el tráfico tiene bastantes picos. Pero a medida que se van agregando los datos mediante escalas de tiempo cada vez mayores, el patrón de tráfico se suaviza. Por lo tanto, es de esperar que la varianza de los datos se reduzca por cada nivel, a diferencia de lo que sucede en un tráfico autosimilar, como es Ethernet.

Por ello, en las simulaciones se prefiere modelar los periodos de tiempo de tráfico, con distribuciones de varianza infinita, utilizando en particular la distribución de Pareto. Esto da como resultado una distribución de elevada varianza, con muchas ráfagas muy cortas, muchas ráfagas largas y algunas ráfagas muy largas. Esto ha permitido determinar el origen de las discrepancias, por ejemplo, entre el tiempo real de espera y el tiempo estimado de espera obtenidos mediante el uso de la teoría de colas convencional usando Poisson.

## CASO DE ESTUDIO EXPERIMENTAL DE TRÁFICO ETHERNET

Como en otros trabajos publicados previamente por Pérez et al (2010), para este caso de estudio, se adopta la distribución Pareto, considerándola más apropiada para modelar el tráfico autosimilar Ethernet. Y se pretende demostrar que a partir de dos muestras Ethernet distintas, las mismas independientemente responden al modelo Pareto, y establecer el valor de los parámetros de dicha distribución. Y además, que las muestras definen comportamientos prácticamente similares. Es decir, que el tráfico mantiene su comportamiento general con distribución Pareto más allá del tiempo.

### *Colección de las muestras o trazas*

Los métodos de colección de tramas de red Ethernet, son el punto de partida para el entendimiento del comportamiento del tráfico y de los nodos de red. Para el trabajo se utilizó el programa Wireshark (ex Ethereal). Es el sniffer más usado con el que se pueden analizar 480 protocolos distintos. Tiene una interfaz flexible con opciones muy ricas de filtrado (Figura n° 1).

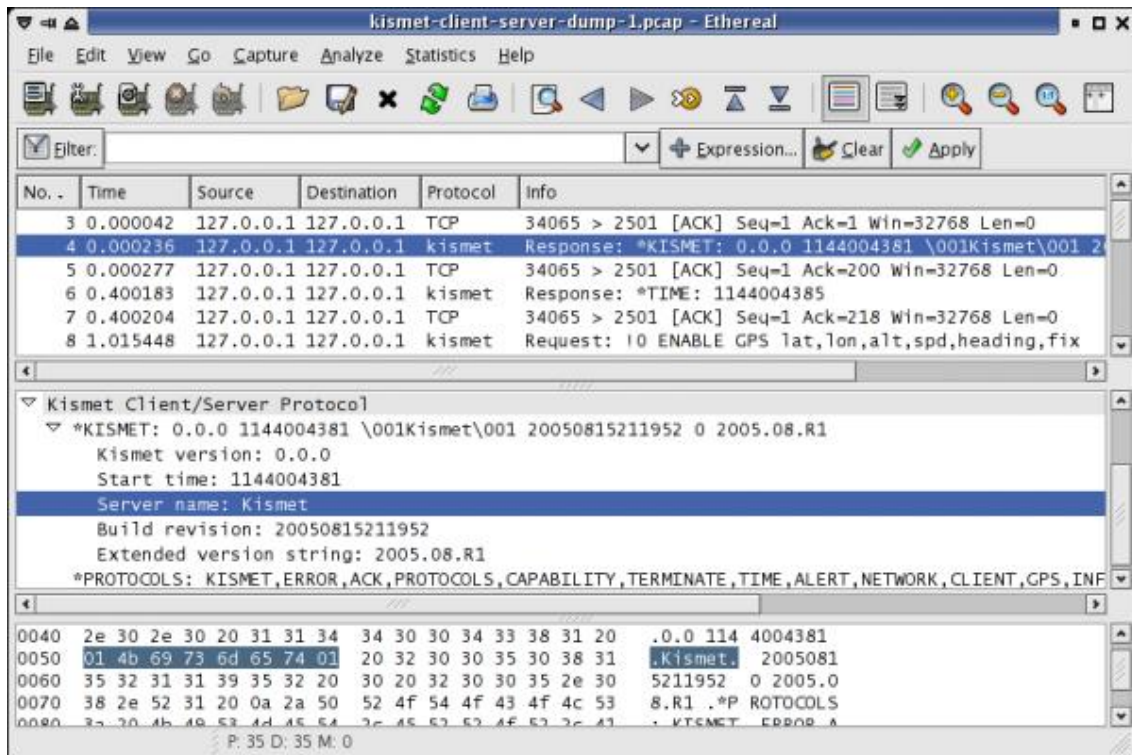


Figura n° 1 Sniffer Wireshark

Con Wireshark, se tomaron 2 muestras de 10 minutos cada una, con una diferencia de 30 minutos entre sí, sobre un switch 3COM, usando un puerto configurado como

monitor, para capturar todo el tráfico que pasaba a través de todos los puertos, con el detalle dado en la Tabla n° 1.

Muestra	Horario de toma	Numero de tramas
A	17:30 hs	116490
B	18:00 hs	74740

Tabla n° 1 Muestras de Tráfico

La Figura n° 2 muestra la representación de la cantidad de tramas en función del tiempo, en los primeros 60 segundos, con una resolución de 1 segundo para la muestra A. La figura presenta el comportamiento típico del tráfico autosimilar Ethernet, que se mantiene para distintas resoluciones de tiempo, como se ha mencionado previamente.

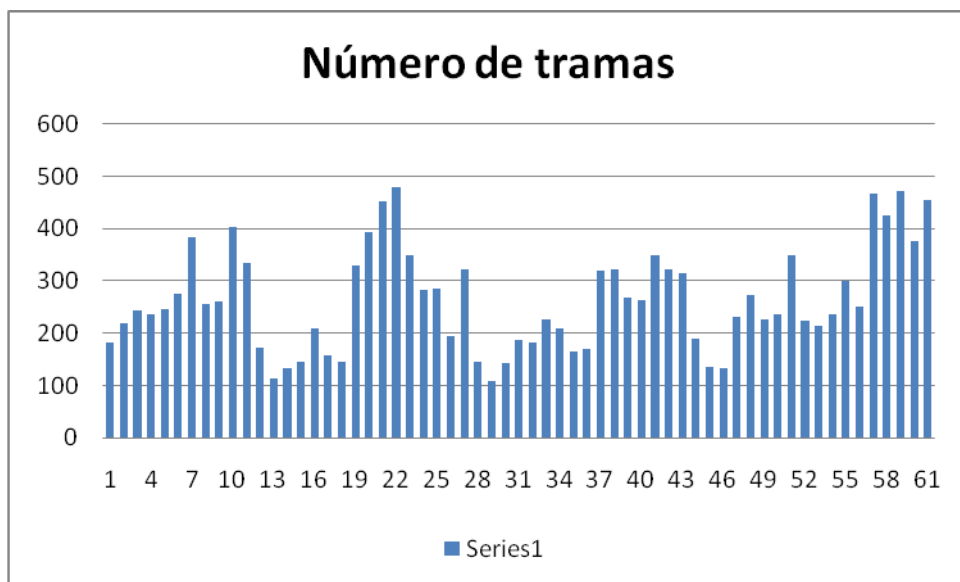


Figura n° 2 Muestra de tráfico

### *Procesamiento de las muestras*

En el procesamiento, las tramas se identificaron por su desplazamiento de tiempo entre sí, y sin los campos innecesarios de la muestra para este estudio, se exportaron como un vector al programa Matchcad, para proceder a su análisis estadístico. Las tramas se agruparon en un vector A y en un vector B según la muestra, de 56 elementos para cada caso, en intervalos de 0.9 mseg, y de 1.8 mseg respectivamente. Y se ordenaron en forma decreciente según la cantidad de tramas, como se indica a continuación:

$$A^T = \begin{array}{|c|c|c|c|c|c|c|} \hline & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 0 & 5.434 \cdot 10^4 & 1.465 \cdot 10^4 & 7.625 \cdot 10^3 & 4.627 \cdot 10^3 & 3.595 \cdot 10^3 & \dots \\ \hline \end{array} \quad \blacksquare$$

$$B^T =$$

	0	1	2	3	4	5
0	$3.507 \cdot 10^4$	$9.749 \cdot 10^3$	$5.616 \cdot 10^3$	$3.778 \cdot 10^3$	$2.5 \cdot 10^3$	...

Posteriormente, estos elementos se normalizaron dividiendo el vector A y B por el número total de tramas de cada muestra, dando origen a un nuevo vector que se llamará ObservA y ObservB (vectores de los valores observados), a los fines del resto del artículo:

$$\text{ObservA}^T =$$

	0	1	2	3	4
0	0.469	0.126	0.066	0.04	...

$$\text{ObservB}^T =$$

	0	1	2	3	4
0	0.471	0.131	0.075	0.051	...

## MODELACIÓN CON DISTRIBUCIÓN PARETO

La distribución Pareto, formulada por el sociólogo Vilfredo Pareto, es una distribución de probabilidad continua con dos parámetros  $a$  y  $b$  cuya función de densidad para valores  $x \geq b$  es:

$$f(x) = \frac{ab^a}{x^{a+1}}$$

Y su función de distribución es:

$$F(x) = 1 - \left(\frac{b}{x}\right)^a$$

El valor esperado y la varianza de una variable aleatoria X de distribución Pareto son

$$E[X] = \frac{ab}{a-1}$$

$$V[X] = \frac{ab^2}{(a-1)^2(a-2)}$$

La distribución de Pareto, puede expresarse como una función  $f(x,a,b)$ , de la siguiente forma:

$$f(x, a, b) := \frac{a \cdot b^a}{x^{a+1}}$$

Considerando la experiencia de los análisis de tráfico de trabajos previos, sobre el comportamiento del tráfico de Ethernet, se determinó que los valores apropiados de los parámetros Pareto para las muestras son:  $a=0,9$  y  $b=1$ .

Y entonces, se pudo generar el vector Esp (vector de valores esperados) para la distribución analítica de Pareto, con  $x$  variando entre 1 y 10.

$$i := 1..10$$

$$E_{i-1} := 1 \cdot \int_i^{i+1} f(x, 0.9, 1) dx$$

$$\text{Esp}^T =$$

	0	1	2	3	4	5	6	7	8
0	0.464	0.164	0.085	0.052	0.036	0.026	0.02	0.015	...

### Comparación de las muestra A y B con la distribución Pareto

La Figura n° 3 muestra la representación del vector observado ObservA y del vector esperado Esp, en función de i, variando entre 0 y 10, construida usando la herramienta Mathcad.

La Figura n° 4 muestra la misma representación, pero para el vector observado ObservB y el vector esperado Esp.

Como se observa, la modelación analítica usando la distribución Pareto es evidentemente bastante satisfactoria para ambas muestras.

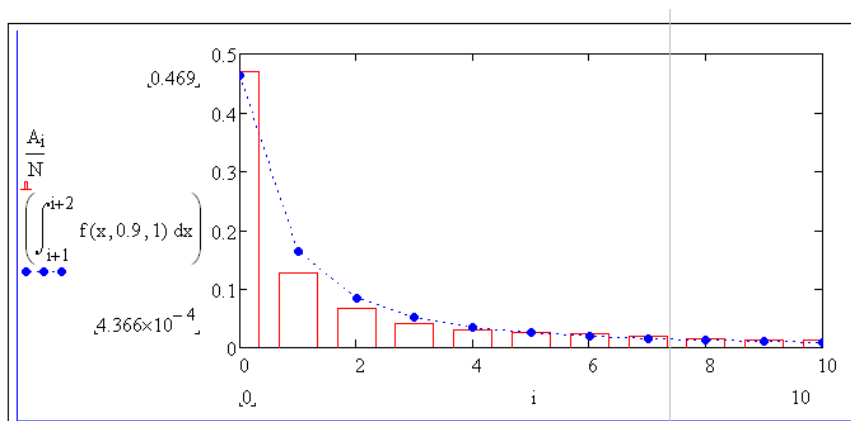


Figura n° 3 Vector ObservA

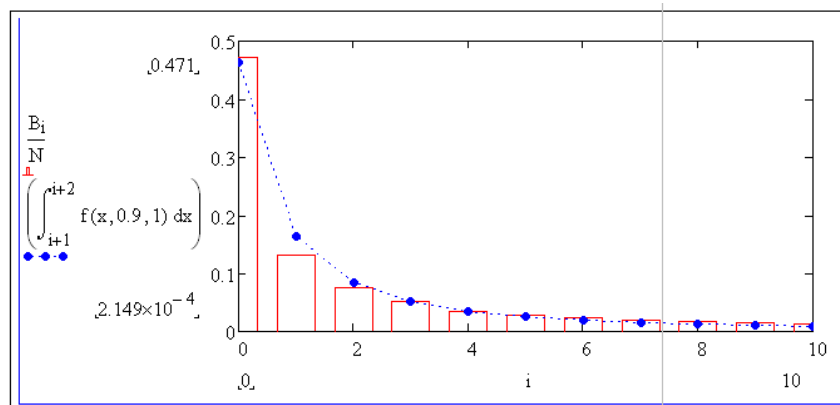


Figura n° 4 Vector ObservB

### Comparación de las muestras A y B entre si

La Figura n° 5 muestra la representación del vector observado ObservA y del vector ObservB.

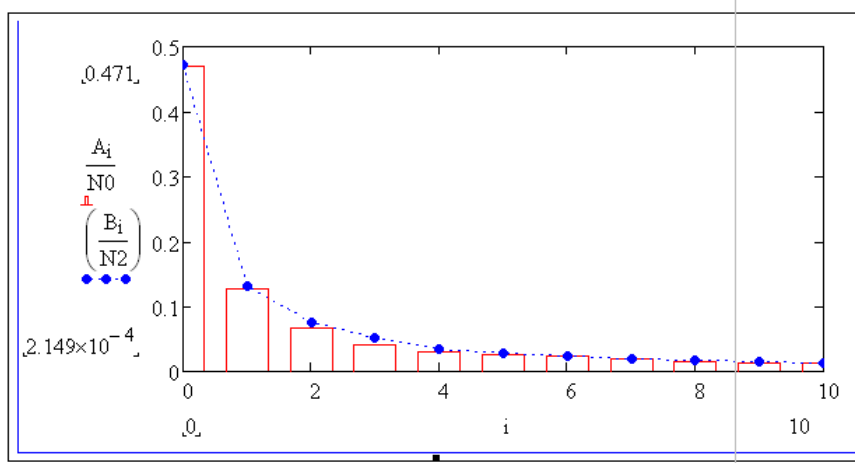


Figura n° 5 Vector ObservA y ObservB

En este caso, resulta sumamente importante observar el comportamiento prácticamente idéntico entre las muestras, tomas en distintos momentos sobre la misma red.

### ANÁLISIS DE BONDAD DE AJUSTE CON LA PRUEBA DE KOLMOGOROV-SMIRNOV

El uso de la Estadística es de gran importancia en la investigación científica. Casi todas las investigaciones aplicadas requieren algún tipo de análisis estadístico para que sea posible evaluar sus resultados. Por ejemplo, usando los tests o dóxicas paramétricos y no paramétricos.

Dentro de las pruebas no paramétricas, se destacan las pruebas de Kolmogorov-Smirnov para una y dos muestras. Se han propuesto diferentes métricas para describir y comparar utilizando diferencias entre distribuciones acumuladas. La prueba unimuestral de Kolmogorov-Smirnov es una prueba de Bondad de Ajuste apropiada para este caso en que se está usando la distribución Pareto. Es más eficiente que la prueba  $\chi^2$  en muestras pequeñas, y no se aplica a distribuciones discretas.

La prueba unimuestral se funda en la diferencia absoluta máxima  $D$  entre los valores de la distribución acumulada de una muestra aleatoria de tamaño  $n$ , y una distribución teórica determinada. Para decidir si esta diferencia es mayor de la razonablemente esperada con un nivel de significación  $\alpha$ , se buscan los valores críticos de  $D$  en Tablas apropiadas.

En el caso en cuestión, se comprobó con un nivel de significancia de  $\alpha=0.05$ , que los valores del vector  $O_{ac}$  (observados acumulados) y del vector  $E_{ac}$  (esperados acumulados) tuviesen un  $D_{max}$  menor a 0.410.



## Bondad de Ajuste de la muestra A con la distribución Pareto

	0		0		
$O_{ac} =$	0	0.469	$E_{ac} =$	0	0.464
	1	0.595		1	0.628
	2	0.661		2	0.713
	3	0.701		3	0.765
	4	0.732		4	0.801
	5	0.759		5	0.826
	6	0.782		6	0.846
	7	0.801		7	0.862
	8	0.816		8	0.874
	9	0.829		9	0.884

Se probó esta hipótesis nula con un nivel de significación de 0.05

1.  $H_0$  Hipótesis nula: están uniformemente distribuidos  
Ha Hipótesis alterna: no están uniformemente distribuidos
2. Nivel de significación 0.05
3. Criterio: Se rechaza  $H_0$  si  $D > 0.410$ , donde  $D$  es la diferencia máxima entre la distribución acumulada observada y la supuesta bajo la Hipótesis Nula.
4. Cálculos: Se construyó el siguiente gráfico (Figura n° 6) que muestra la distribución acumulada observada y la supuesta.

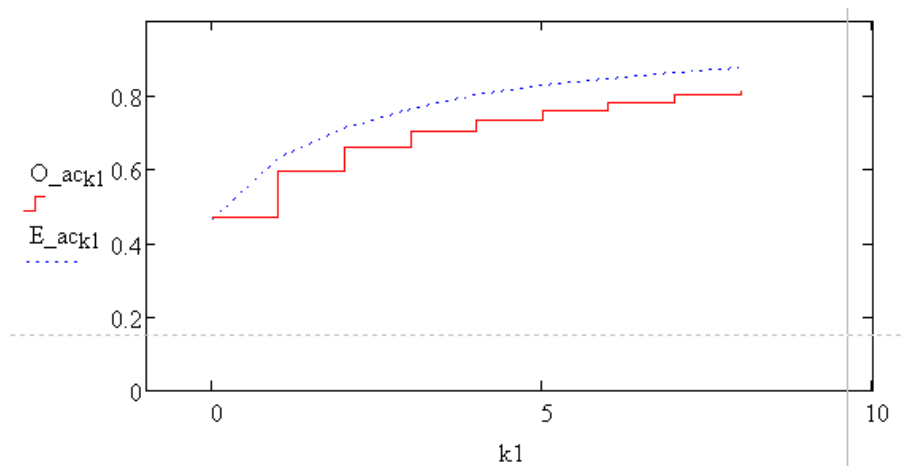


Figura n° 6 Distribución acumulada observada y supuesta muestra A

Dado que:

$$E_{ac_1} = 0.628$$

$$O_{ac_0} = 0.469$$

Se deduce que:

$$D_{max} = E_{ac_1} - O_{ac_0} = 0.159$$

Por otro lado, y de acuerdo a los datos y con un nivel de significación de 0.05, el D crítico de la Tabla 6.1 es 0.410.

5. Decisión: Ya que  $0.159 < 0.410$  (valor de Tabla) *se acepta* la Hipótesis Nula. Luego, los valores acumulados están significativamente distribuidos conforme a una Distribución de Pareto con coeficientes  $a=0.9$  y  $b=1$ , y por lo tanto, el tráfico de datos Ethernet responden a la citada distribución con ese nivel de significancia.

TAMAÑO DE LA MUESTRA (N)	NIVEL DE SIGNIFICANCIA PARA D = MAX [ F <sub>0</sub> (X) - S <sub>n</sub> (X) ]				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490

Tabla 6.1

***Bondad de Ajuste de la muestra B con la distribución Pareto***

O<sub>ac</sub> =

	0
0	0.471
1	0.602
2	0.678
3	0.728
4	0.762
5	0.79
6	0.812
7	0.832
8	0.85
9	0.866

E<sub>ac</sub> =

	0
0	0.464
1	0.628
2	0.713
3	0.765
4	0.801
5	0.826
6	0.846
7	0.862
8	0.874
9	0.884

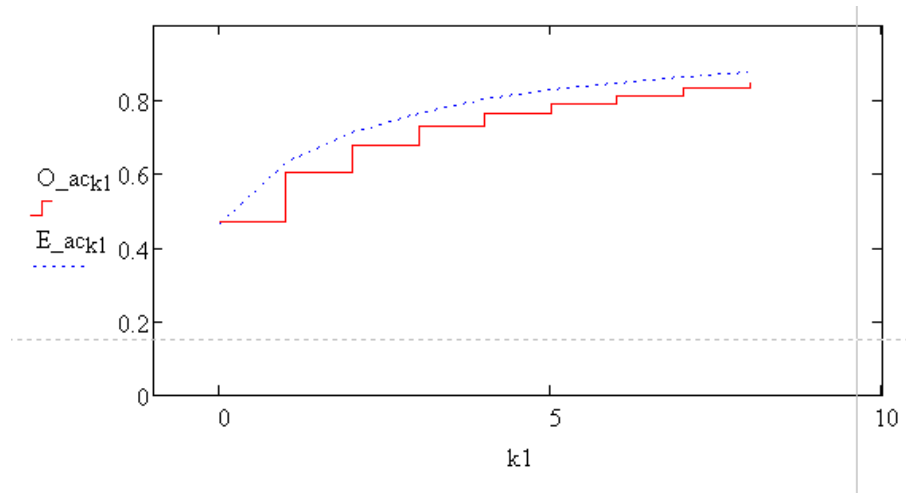


Figura n° 7 Distribución acumulada observada y supuesta muestra B

Dado que:

$$E_{ac1} = 0.628$$

$$O_{ac0} = 0.471$$

Se deduce que:

$$D_{max} = E_{ac1} - O_{ac0} = 0.157$$

6. Decisión: Ya que  $0.157 < 0.410$  (valor de Tabla) *se acepta* la Hipótesis Nula. Luego, los valores acumulados están significativamente distribuidos conforme a una Distribución de Pareto con coeficientes  $a=0.9$  y  $b=1$ , y por lo tanto, el tráfico de datos Ethernet responden a la citada distribución con ese nivel de significancia (Figura n° 7).

### ***Bondad de Ajuste entre las muestras A y B***

ObservB\_ac =

	0
0	0.471
1	0.602
2	0.678
3	0.728
4	0.762
5	0.79
6	0.812
7	0.832
8	0.85
9	0.866

ObservA\_ac =

	0
0	0.469
1	0.595
2	0.661
3	0.701
4	0.732
5	0.759
6	0.782
7	0.801
8	0.816
9	0.829

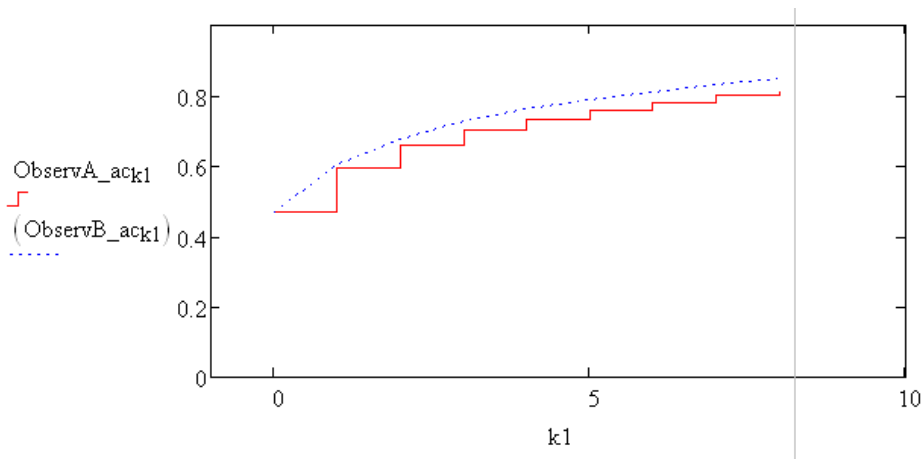


Figura n° 8 Distribución acumulada observada A y B

Dado que:

$$\text{ObservA\_ac}_1 = 0.595$$

$$\text{ObservB\_ac}_0 = 0.471$$

Se deduce que:

$$D_{\max} = \text{ObservA\_ac}_1 - \text{ObservB\_ac}_0 = 0.124$$

7. Decisión: Ya que  $0.124 < 0.410$  (valor de Tabla) *se acepta* la Hipótesis Nula.

Luego, los valores acumulados están significativamente distribuidos de forma similar, y conforme a ese nivel de significancia (Figura n° 8).

## CONCLUSIÓN

En este documento, se han relacionado los temas de autosimilitud, con el tráfico Ethernet, la distribución de Pareto y la prueba de Kolmogorov-Smirnov. El volumen de los trabajos y literatura sobre tráfico de datos es creciente, y el tema de la autosimilitud ha significado el principio de un nuevo examen del rendimiento del tráfico de datos, las técnicas de modelado, control de tráfico, dimensionamiento de buffers en dispositivos activos, entre otros. En el trabajo se ha verificado a través de un estudio experimental y usando la prueba de Kolmogorov-Smirnov, que el tráfico de datos Ethernet responde efectivamente a la distribución Pareto, estableciéndose los valores de los parámetros de dicha distribución. Y por lo tanto, se ha demostrado que dos muestras de tráfico Ethernet, tomadas en distintos momentos, tienen distribución Pareto con parámetros prácticamente similares.

## REFERENCIAS Y BIBLIOGRAFÍA

- Frost, V. y Melamed B. (1994). Traffic Modeling for Telecommunications Networks en IEEE Communications, <http://ieeexplore.ieee.org/iel1/35/6685/00267444.pdf>
- Leland, W; Taqu, M; Willinger, W. y Wilson, D (1994). On the Self-Similar Nature of Ethernet Traffic en IEEE/ACM Transaccions on Networking.

- Perez, S; Facchini, H; Mercado, G. y Bisaro, L. (2010). Análisis del Comportamiento Autosimilar del tráfico Ethernet de las Redes de Datos en CACIC 2010.
- <http://www.wireshark.org>
- <http://www.ptc.com/products/mathcad/>
- <http://www.um.edu.ar/math/estadis/programa.htm>

## **ABSTRAC**

The queuing analysis has been and is of enormous usefulness for the designers of networks and analysts of traffic, to effects of planning the capacities of the elements of network and of predicting his performance. In addition, the exact models of the Ethernet traffic are important to model the applications upper layer and the buffers of memory of switchs. These analyses depend on the nature Poisson of the traffic of data. Nevertheless, many results predicted from the analysis of queuings differ significantly from the performance observed in the reality. Diverse studies have demonstrated that for some environments the pattern of traffic is self-similar, instead of Poisson. This concept is related to other acquaintances since it are the fractals and the theory of the chaos. From beginning of the 90s were begun to publish documents referred to the self-similarity of the traffic of Ethernet. In this document the interframe distribution of sequences of Ethernet traffic of diverse content is analyzed. Using captures of Ethernet traffic realized in different moment, has analyzed the distribution Pareto as possible model, and there has been compared the goodness of fit of this possible model of distribution, and of the samples between if using the statistician Kolmogorov-Smirnov. The present work develops successively the following paragraphs: 1) Introduction, 2) Network traffic Models, 3) Self-similarity Ethernet data traffic, 4) Case of experimental Study of Ethernet traffic, 5) Analysis with distribution Pareto, and 6) Analysis of goodness of fit with Kolmogorov-Smirnov's test.

Key words: traffic, self-similarity, Ethernet, Pareto, goodness of fit