

# Estudio sobre la Distribución de Tráfico Autosimilar en Redes Wi-Fi

Santiago Pérez, Higinio Facchini, G. Mercado, Carlos Taffernaberry, Luis Bisaro

Grupo de Investigación y Desarrollo en TICs (GRID TICs)  
Universidad Tecnológica Nacional, Facultad Regional Mendoza  
([santiagocp.higiniofac.gmercado.carlos\\_taffe.lbisaro@frm.utn.edu.ar](mailto:santiagocp.higiniofac.gmercado.carlos_taffe.lbisaro@frm.utn.edu.ar))

**Resumen.** Los modelos exactos del tráfico Ethernet y Wireless 802.11 son importantes para modelar las aplicaciones de capa superior y los buffers de memoria de switches, APs y bridges. Los análisis de cola realizados suponiendo la naturaleza Poisson del tráfico de datos 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. Desde principios 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 Wireless 802.11 de contenido diverso. Usando capturas de tráfico Wireless 802.11 realizadas en distinto momento, se han analizado y determinado los valores de los parámetros usando la distribución Pareto como modelo. Además, se ha determinado la bondad de ajuste de las muestras con Pareto, y de las muestras entre sí. Finalmente, se ha efectuado un contraste con el comportamiento y el valor de los parámetros Pareto de tráfico Ethernet de un estudio anterior. El presente trabajo desarrolla sucesivamente los siguientes apartados: 1) Introducción, 2) Modelos de Tráfico de Red, 3) Tráfico de Datos Ethernet Autosimilar con Distribución Pareto, 4) Caso de Estudio Experimental de Tráfico Wireless 802.11, 5) Modelación con Distribución Pareto, 6) Comparación de los parámetros y las curvas de Distribución Pareto Ethernet y Wireless 802.11, y 7) Conclusiones.

**Palabras claves:** tráfico, Wireless 802.11, Distribución Pareto, autosimilar

## 1 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 el modelo representa mejor el tráfico real, conocido como bondad de ajuste. Por ello, para evaluar la bondad de ajuste, deberían definirse métricas que permitan cuantificar cuan cerca del modelo están 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.

## 2 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 [1]. 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:

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

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

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 Pareto son excelentes candidatos, dado que el modelo tiene en cuenta la correlación de largo alcance en los tiempos de arribo.

### **3 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 en [2]. 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. Desde entonces, se han sucedido un gran número de trabajos haciendo aportes y dando precisiones al respecto [3-11].

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.

En [12] se describe un estudio que verifica tal comportamiento en redes Ethernet, y es usado como antecedente para el desarrollo del presente trabajo, pero sobre ambientes 802.11. Además, es utilizado para efectuar la comparación de los parámetros de la distribución Pareto en Ethernet y en Wireless 802.11

## 4 Caso de Estudio Experimental de Tráfico Wireless 802.11

Como en trabajos publicados previamente [12], para este caso de estudio se adopta la distribución Pareto, considerándola más apropiada para modelar el tráfico autosimilar wireless 802.11. Y se pretende demostrar que a partir de dos muestras Wireless 802.11 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 similares. Es decir, que el tráfico mantiene su comportamiento general más allá del tiempo.

### 4.1 Colección de trazas en red universitaria y análisis con WEKA

Los métodos de colección de tramas de red Ethernet y Wireless, 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 [13] (ex Ethereal) combinado con la herramienta de captura AIRPCAP NX: USB 802.11 a/b/g/n [14]. Y simultáneamente, con fines de comprobación y para obtener otros datos complementarios, se ejecutó el sniffer wireless AirMagnet [15]. Ambos tienen una interfaz flexible con opciones muy ricas de filtrado (Fig. 1 y 2).

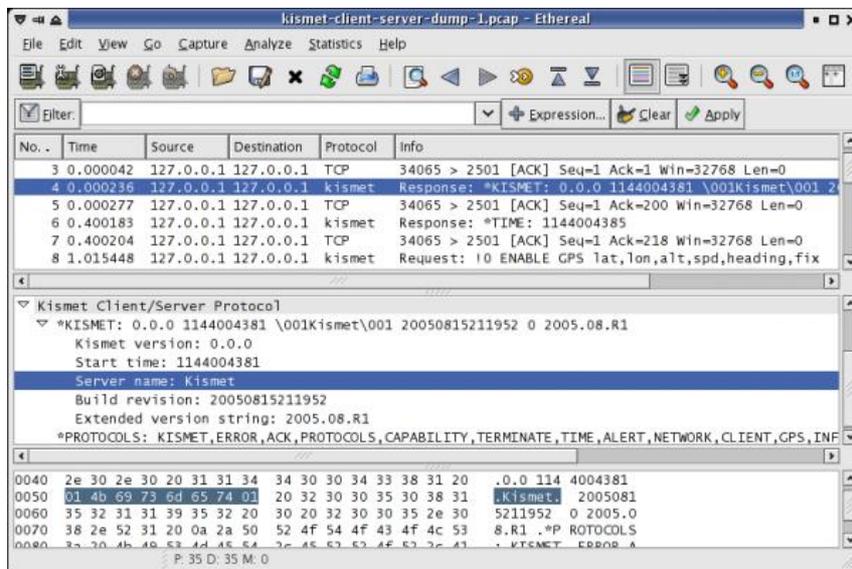


Fig. 1 Vista del ámbito de captura del Sniffer Wireshark

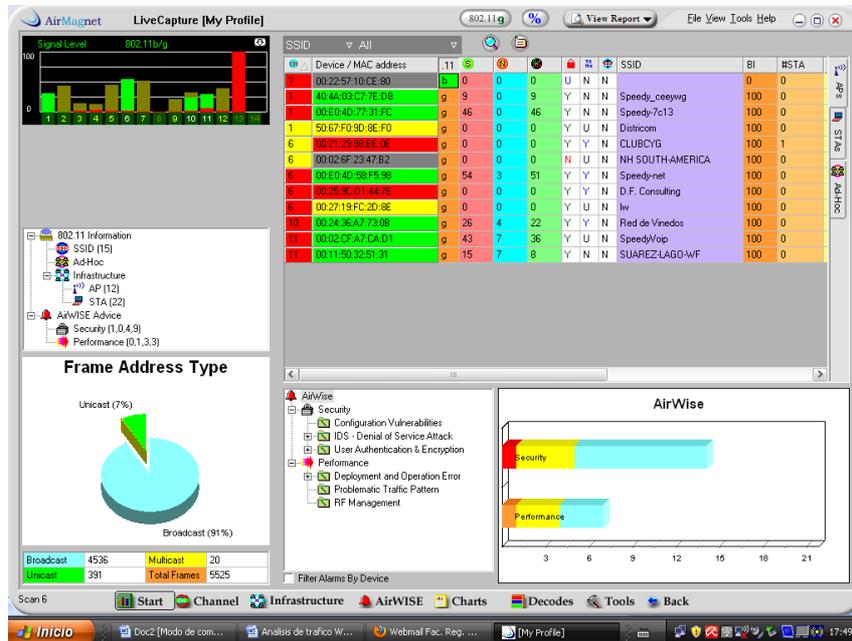


Fig. 2 Vista del ámbito de captura Sniffer AirMagnet

Con Wireshark, se tomaron 2 muestras wireless 802.11 de 10 minutos cada una, con una diferencia de 20 minutos entre sí, sobre un ambiente administrativo en producción, con varios APs, seleccionando uno de los canales habilitados, con el detalle dado en la Tabla 4.1.

Tabla 1 Muestras de Tráfico wireless

Muestra	Horario de toma	Número de Tramas	Porcentaje Unicast	Porcentaje Broadcast	Porcentaje Multicast
A	10:00 hs	95970	49%	49%	2%
B	10:20 hs	196400	55%	39%	4%

## 4.2 Procesamiento de las Muestras

El análisis de la distribución se realizó sobre la variable tiempos intertramas. 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 [16], para proceder a su análisis estadístico. Las tramas se agruparon en un vector A y en un vector B según la muestra, de 61 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:

$$A^T =$$

	0	1	2	3	4	5
0	$6.511 \cdot 10^4$	$8.441 \cdot 10^3$	$3.008 \cdot 10^3$	$5.025 \cdot 10^3$	$1.16 \cdot 10^3$	...

$$B^T =$$

	0	1	2	3	4	5
0	$1.309 \cdot 10^5$	$2.673 \cdot 10^4$	$9.833 \cdot 10^3$	$4.795 \cdot 10^3$	$2.469 \cdot 10^3$	...

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 llamaremos ObservA y ObservB (vectores de los valores observados), a los fines del resto del artículo:

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

	0	1	2	3	4
0	0.678	0.088	0.031	0.052	...

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

	0	1	2	3	4
0	0.666	0.136	0.05	0.024	...

## 5 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 previa, de los análisis del comportamiento del tráfico Ethernet, para el tráfico wireless 802.11 se asignaron a los parámetros Pareto los valores  $a=1,9$  y  $b=0.95$ . A partir de ello, se pudo generar el vector Esp (vector de valores esperados) para la distribución analítica de Pareto, con  $x$  variando entre 1 y 10.

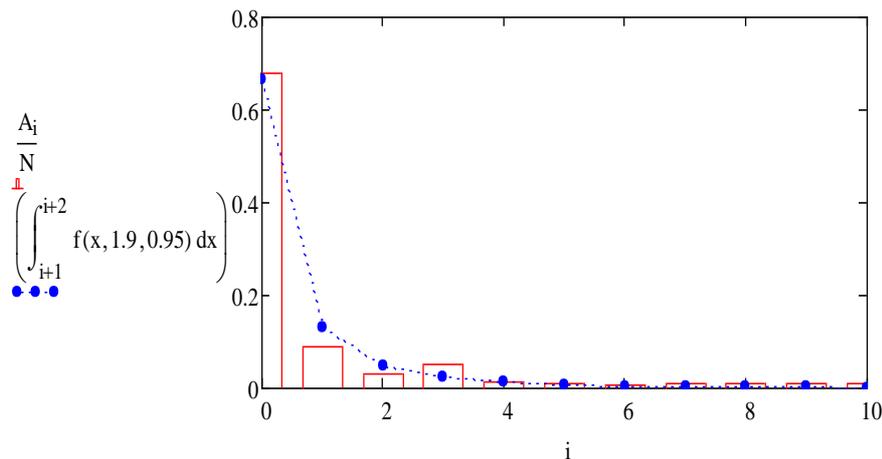
$$i1 := 1..10$$

$$E_{i1-1} := 1 \cdot \int_{i1}^{i1+1} f(x, 1.9, 0.95) dx$$

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

	0	1	2	3	4
0	0.664	0.131	0.047	0.023	...

La Fig. 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 Fig. 4 muestra la misma representación, pero para el vector observado ObservB. Se deduce que la modelación analítica con los parámetros adoptados para la distribución Pareto es bastante satisfactoria para ambas muestras.



**Fig. 3** Representación del vector ObservA y del vector Esp

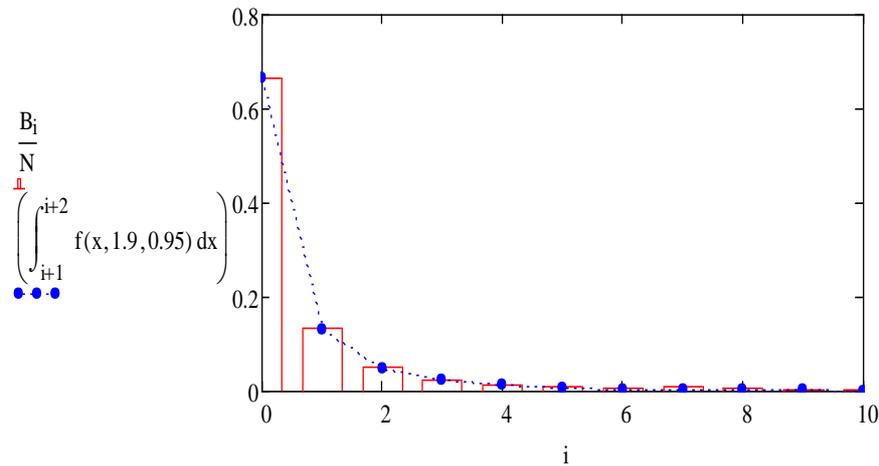


Fig. 4 Representación del vector ObservB y del vector Esp

A los efectos de evaluar estadísticamente los resultados se utilizó el test o dócima no paramétrica de Kolmogorov-Smirnov, para una y dos muestras. Esta prueba describe y compara diferencias entre distribuciones acumuladas. Es una prueba de bondad de ajuste apropiada para este caso en que se está usando la distribución Pareto [18]. La prueba 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, se buscan los valores críticos de  $D$  en Tablas apropiadas. En el caso en cuestión, se comprobó con un nivel de significancia de 0,05 que los valores acumulados del vector ObservA y ObservB, respecto del vector Esp tenían un valor menor que el  $D_{max}$  referencial de 0,410 para muestras de tamaño 10. Se hizo la misma comprobación de las muestras entre sí.

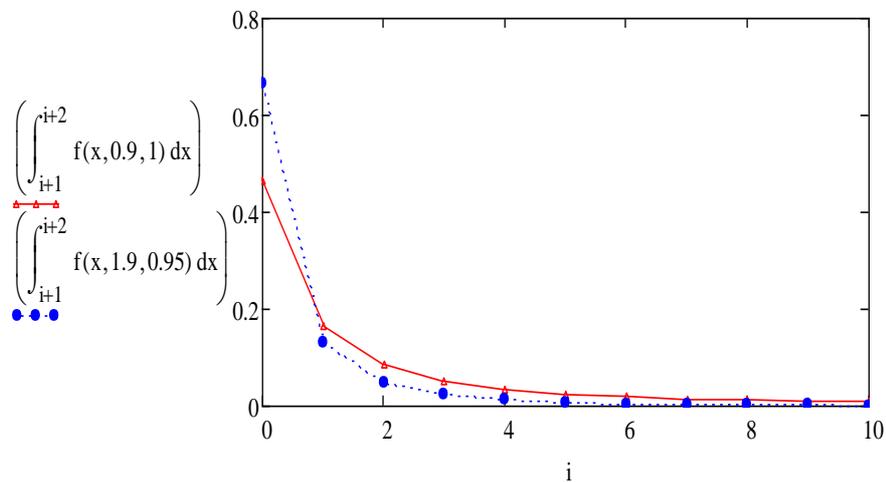
## 6 Comparación de los parámetros y las curvas de Distribución Pareto de Ethernet y Wireless 802.11.

La Fig. 5 representa en líneas punteadas la distribución Pareto de la muestra del vector ObservB del tráfico wireless 802.11e y en línea continua la distribución Pareto de la muestra del vector del tráfico Ethernet presentada en [12]. Para las muestras del tráfico wireless se usaron para la modelación Pareto los valores  $a=1.9$  y  $b=0.95$ , mientras que para la muestra del tráfico Ethernet, se modeló con los valores  $a=0.9$  y  $b=1$ . La Tabla 1 resume los valores de los parámetros  $a$  y  $b$  para estos tráficos.

	802.11	Ethernet
a	1.9	0.9
b	0.95	1

**Tabla 1** Valores para las muestras de tráfico 802.11 y Ethernet

Ambos tipos de tráficos responden al comportamiento general de autosimilitud de la distribución Pareto, con una gran cantidad de tramas con un espacio de interarribo muy chico, y pocas tramas con un espacio interarribo muy grande, característico del tráfico de ráfagas. Sin embargo, se observa que el tráfico wireless presenta un 30% más de componentes de ráfagas respecto al tráfico Ethernet, y una menor cantidad de las componentes que no son ráfagas.



**Fig. 5** Representación del vector Wi-Fi y el vector Ethernet

## 7 Conclusiones

En este documento, se han relacionado los temas de autosimilitud, con el tráfico Wireless 802.11, la distribución de Pareto y la prueba de Kolmogorov-Smirnov, con el mismo criterio que se hizo sobre tráfico Ethernet en una publicación previa. El volumen de los trabajos y literatura sobre análisis de 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, que el tráfico de datos Wireless 802.11 responde efectivamente a la distribución Pareto, estableciéndose los valores de los parámetros de dicha distribución. Además, el estudio incluyó dos muestras de tráfico Wireless 802.11, tomadas en distintos momentos, para demostrar que el

comportamiento Pareto es independiente del tiempo, para la red bajo estudio. Comparado con el tráfico Ethernet, el tráfico wireless presenta una mayor tendencia a las ráfagas. Estas definiciones pueden ser una importante orientación en la configuración de los modelos de tráficos Ethernet y Wi-Fi en los estudios analíticos o de simulación.

## Referencias

- [1] Frost V., Frost S. y Melamed B., "Traffic Modeling for Telecommunications Networks", IEEE Communications, Mar. 1994, <http://ieeexplore.ieee.org/iel1/35/6685/00267444.pdf>
- [2] Leland W., Taqqu M, Willinger W. y Wilson D., "On the Self-Similar Nature of Ethernet Traffic", IEEE/ACM Transactions on Networking, Febrero 1994
- [3] Sarkar N. y Sowerby K., "The Effect of Traffic Distribution and Transport Protocol on WLAN Performance", Telecommunication Networks and Applications Conference (ATNAC), 2009 Australasian
- [4] Kihl M., Ödler P., Lagerstedt C. y Aurelius A., "Traffic analysis and characterization of Internet user behavior", International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) 2010
- [5] Ghani S. y Iradat F., "Loss Probability in Networks with Pareto Distributed Traffic", Second International Conference on Intelligent Systems, Modelling and Simulation 2011.
- [6] Galkin A., Simonina O., Yanovsky G., "Multiservice IP Network QoS Parameters Estimation in Presence of Self-similar Traffic", 6<sup>th</sup> Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking, NEW2AN 2006.
- [7] Ghani S., "The impact of Self Similar Traffic on Wireless LAN", International Wireless Communications and Mobile Computing Conference (IWCMC 2010).
- [8] Floyd S. y Paxson V., "Wide area traffic: the failure of poisson modeling", IEEE/ACM Transactions of Networking, vol. 3, no.3. 1995.
- [9] Liu S., Wang P. y Qu L., "Modeling and simulation of self-similar data traffic", 4<sup>th</sup> IEEE International Conference on Machine Learning and Cybernetics, 2005.
- [10] Floyd S., "Difficulties in simulating the Internet", IEEE/ACM Transactions on Networking, vol. 9, no. 4, 2001.
- [11] Ghani S., "Analysis of Pareto Distribution on Delay Performance in Networks", The 1<sup>st</sup> International Conference on Wireless Communications Frontiers (WCF 2010), 2010.
- [12] Perez S., Facchini H., Mercado G., y Bisaro L., "Análisis del Comportamiento Autosimilar del tráfico Ethernet de las Redes de Datos", CACIC 2010, 2010.
- [13] <http://www.wireshark.org>
- [14] [http://www.cacetech.com/products/airpcap\\_nx.html](http://www.cacetech.com/products/airpcap_nx.html)
- [15] <http://www.flukenetworks.com/enterprise-network/wireless-network/airmagnet-wifi-analyzer>
- [16] <http://www.ptc.com/products/mathcad>
- [17] <http://www.um.edu.ar/math/estadis/programa.htm>