

Comparación del modelo FARIMA y SFARIMA para obtener la mejor estimación del tráfico en una red Wi-Fi

Comparing FARIMA and SFARIMA models in order to obtain the best traffic estimation over a Wi-Fi network

JOEL CAROLL VARGAS

Ingeniero Electrónico. Estudiante de la maestría en Ciencias de la Información y las Comunicaciones de la Universidad Distrital Francisco José de Caldas. Bogotá, Colombia. *discomsystem@gmail.com*

CESAR HERNÁNDEZ

Ingeniero Electrónico, magister en Ciencias de la Información y las Comunicaciones. Docente e investigador de la Universidad Distrital Francisco José de Caldas. Bogotá, Colombia. *cahernandezs@udistrital.edu.co*

GABRIEL ALBERTO PUERTA APONTE

Licenciado en Electrónica. Docente de la Corporación Universitaria UNITEC. Bogotá, Colombia. *gapuerta@correo.udistrital.edu.co*

Clasificación del artículo: Investigación (Conciencias)

Fecha de recepción: Agosto 4 de 2011

Fecha de aceptación: Febrero 27 de 2012

Palabras clave: enrutador, tráfico fractal, series de tiempo, WDS, Wi-Fi.

Key words: router, fractal traffic, time series, WDS, Wi-Fi.

RESUMEN

El presente artículo se enfoca en el análisis del tráfico de la red WI-FI pública de Chía, Cundinamarca, determinando la presencia de tráfico autosimilar y aplicando posteriormente los algoritmos FARIMA y SFARIMA. Los resultados determinan cuál es el mejor predictor del tráfico de la red.

ABSTRACT

This paper focuses on the analysis of network traffic public WIFI Chia, Cundinamarca, determining the presence of self-similar traffic and apply the algorithms SFARIMA and FARIMA. The results determine what the best predictor network traffic is.

* * *

1. INTRODUCCIÓN

Poder predecir el comportamiento de un sistema resulta vital para la correcta y óptima administración de los recursos involucrados, para lo cual la ingeniería de tráfico en algunos casos se ha nutrido de avances en otros campos. Cabe citar el trabajo de Box y Jenkins en 1976, y en los años ochenta los modelos fraccionales autorregresivos de promedio móvil FARIMA (autoregressive fractionally integrated moving average ‘ARFIMA’ models), que se utilizaban para estimaciones de largo plazo (Granger (1980), Granger, Joyeux (1980) y Hosking (1981)) [1]. De esta suerte, en los años noventa se empezaron a incorporar tales modelos en diferentes campos, lo cual ha redundado positivamente en el análisis del tráfico producido en una red de paquetes de datos.

El modelo ARFIMA o FARIMA, el cual involucra la memoria de largo plazo puede considerarse como una extensión de los modelos ARIMA (memoria de corto plazo), toda vez que al estudiar el comportamiento del tráfico en una red de datos se determinó que este presenta un comportamiento autosimilar, vale decir que el todo es similar o parecido a una parte del mismo [2].

Los primeros algoritmos de predicción de series de tiempo comenzaron con los estudios sobre el comportamiento a corto plazo y a largo plazo del mercado, elaborado por el trabajo de climatología e hidrología presentado por Hurst en 1951,

en econometría está el trabajo de Box y Jenkins en 1976, ya en 1980 los modelos fraccionales autoregresivos de promedio móvil FARIMA (autoregressive fractionally integrated moving average ‘ARFIMA’ models), se utilizaban para estimaciones de largo plazo (Granger (1980), Granger, Joyeux (1980) y Hosking (1981)) [2] [3]. En los años noventa se empezaron a incorporar estos modelos en diferentes campos, en particular han tenido una especial atención en el análisis del tráfico producido en una red de paquetes, entre los primeros trabajos está el de Nancy K. Groschwitz y George C. Polyzos: *ATime Series Model of Long-Term NSFNET Backbone Traffic* adelantado entre 1988 y 1993.

Tabla 1. Evolución cronológica de las series de Tiempo

AUTOR	AVANCE / APOORTE	AÑO
HURST	trabajo de climatología e hidrología	1951,
BOX y JENKINS	Econometría	1976
Granger, Joyeux, Hosking	FARIMA	1980-1981
Nancy K. Groschwitz y George C. Polyzos	Primeros trabajos en tráfico de redes	1988- 1993

2. PROCESOS DE MEMORIA DE LARGO PLAZO

Cuando se realiza un estudio de serie de tiempo es normal que se refleje la propiedad consistente en que la correlación decrece exponencialmente a cero, como en los modelos ARMA. Para los modelos de memoria de largo plazo se parte de la Ec. (1).

$$\lim_{n \rightarrow \infty} \sum_{i=-n}^n |\rho_i| = \infty \quad (1)$$

Donde $\rho(k)$ corresponde a la función de correlación.

Diferente de los modelos con memoria de corta duración que se enmarcan en la Ec. (2) [3].

$$\lim_{n \rightarrow \infty} \sum_{i=-n}^n |\rho_i| = k \quad (2)$$

Las series de tiempo con memoria de larga duración se distinguen porque la correlación entre dos observaciones en periodos de tiempo muy grandes es alta, y el decaimiento de ella no es exponencial como en el caso ARMA.

Los modelos FARIMA comportan el parámetro d , el cual mide el grado de intensidad de la memoria de larga duración [4].

2.1 Modelo FARIMA

El modelo de un proceso FARIMA (p,d,q), con media μ , en donde el índice p señala el orden de autorregresión, d indica el orden de la diferenciación, y q es el orden del promedio móvil, puede ser descrito como sigue [5]:

$$\Phi(L)(1-L)^d(R_t - \mu) = \Theta(L)\varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2) \quad (3)$$

Donde L es el operador de retraso,

$$\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p \quad (4)$$

$$\Theta(L) = 1 + \vartheta_1 L + \dots + \vartheta_q L^q \quad (5)$$

$(1-L)^d$ es el operador de diferenciación fraccional el cual corresponde a:

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)} \quad (6)$$

Con $\Gamma(\cdot)$ es la función gamma (generalizada factorial) [6], no es estacionario para $d \geq 0.5$ dando una varianza infinita, si se toma $d \in (-0.5, 0.5)$ (ec.4), El modelo FARIMA permite que el parámetro diferenciador d pueda tomar valores reales no enteros, R_t corresponde al proceso estocástico que es estacionario e invertible siempre que todas las raíces de $\Phi(L)$ y $\Theta(L)$ (Ec.2) se encuentren fuera del círculo unitario. El parámetro d determina el comportamiento de la memoria del modelo, de forma tal que: si d toma valores entre 0 y 0.5 ($d \in (0, 0.5)$), la función de autocorrelación r es proporcional a k^{2d-1} y decae hiperbólicamente a cero cuando $k \rightarrow \infty$, afirmándose que el modelo exhibe una memoria de largo plazo o dependencia de largo plazo. Si $d \in (-0.5, 0)$ el proceso es de memoria intermedia (antipersistencia) o dependencia negativa de largo plazo, y por último para el caso que $d = 0$, la memoria es de corto plazo y corresponde a un modelo ARMA estacionario e invertible.

3. METODOLOGÍA

En el municipio de Chía, Cundinamarca, se implementó una red inalámbrica WI-FI con una cobertura de 105 Kilómetros cuadrados aproximadamente, con una velocidad promedio de acceso a internet de 2 Mbps en cada nodo, utilizando la tecnología WDS (*Wireless Distribution System*) para la distribución de celdas y enlaces de 7 kilómetros como backhaul para el suministro de Internet en las celdas. En la Fig. 1 se puede apreciar la distribución de celdas, nodos y backhaul en la red WI-FI de Chía:

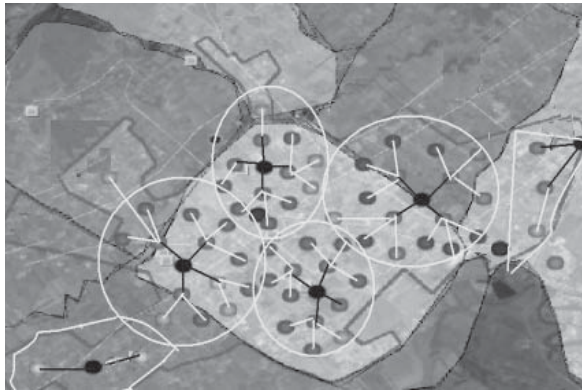


Fig. 1 Arquitectura lógica de la red WI-FI en el Municipio de Chía

Como es conocido, el primer paso consiste en tomar una muestra de datos en los que se pueda caracterizar el tráfico de la red. Los datos de tráfico fueron extraídos utilizando la herramienta de *PRTG Network Monitor*. Con esta herramienta se puede monitorear en tiempo real los valores de tráfico de las redes de datos.

Para la primera captura se utilizó una variable Kbit/s y se tomaron 24 muestras correspondientes a 2 horas con intervalos de 5 minutos. En la Fig. 3 se puede observar gráficamente el comportamiento de la red WI-FI en dos horas.

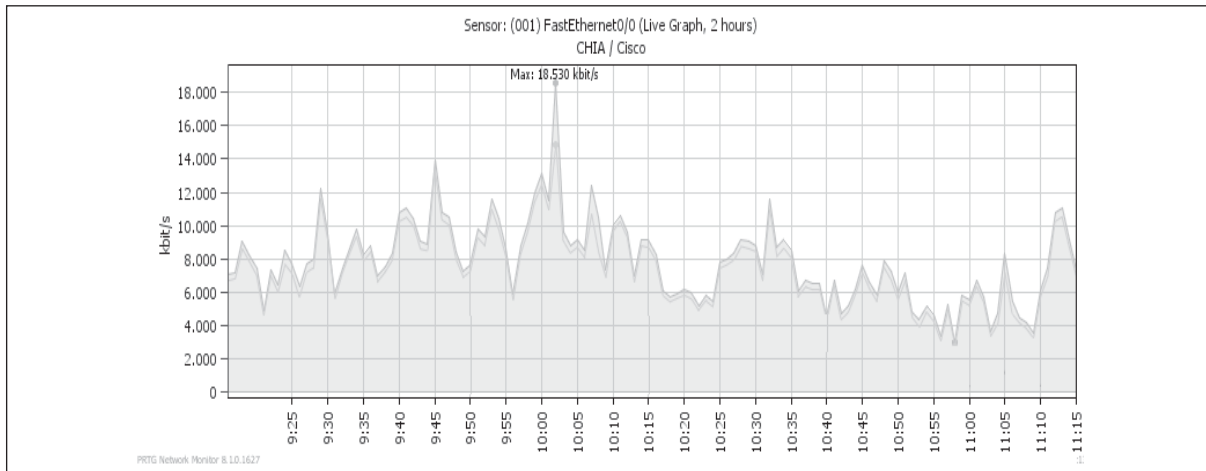


Fig. 2. Tráfico observado en la red WI-FI durante dos horas

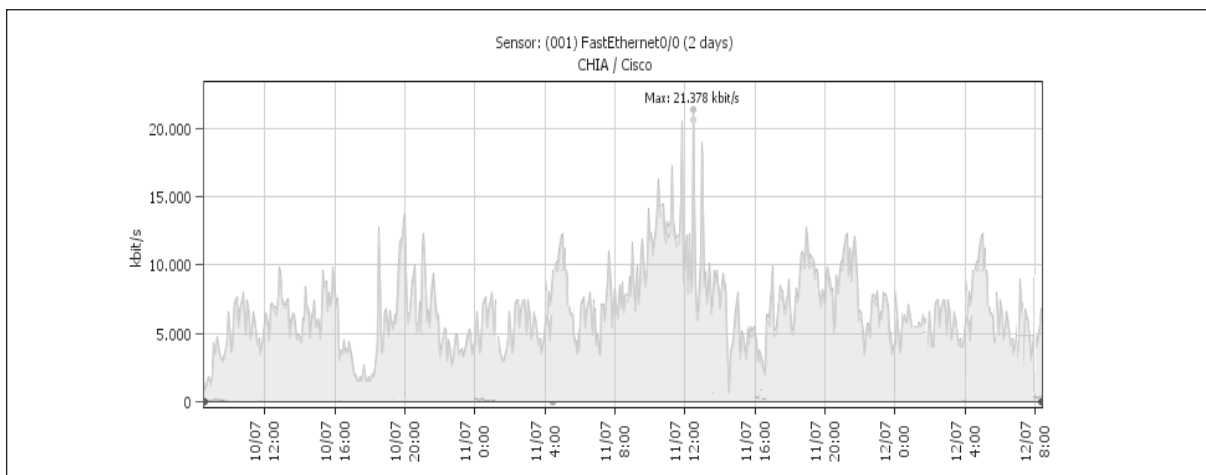


Fig. 3. Tráfico observado en la red WI-FI durante dos días

Para la segunda captura también se utilizó la variable en Kbit/s y se tomaron 12 muestras correspondientes a 2 días. En la Fig. 4 se observa gráficamente el comportamiento de la red WI-FI en dos horas.

Como se infiere la red WI-FI analizada tiene el mismo comportamiento sin importar la escala, lo que implica un comportamiento autosimilar [7].

4. RESULTADOS

Lo primero es especificar la función: function [Z] = FARIMA(N,F,O,d,stdx)

Donde las estradas de la función están dadas por:

N = es la longitud de la serie de tiempo.

F= es el valor de p que señala el orden de autorregresión.

O = es el valor de q que indica el orden del promedio móvil.

d = es el que indica el orden de la diferenciación.

stdx = parámetro para forzar la desviación estándar de la serie de tiempo.

Se define el modelo FARIMA en Matlab por

```
%%%%%%%% F(B)[(1-B)^d]Z=O(B)
```

```
%%%%%%%% F(B)Z=[(1-B)^-d]O(B)
```

Donde B es el operador de retardos

```
%%%%%%%% F(B)= 1+ B F1 + B^2 F2 ... + B^p Fp
-->Parte AR
```

```
%%%%%%%% O(B)= 1+ B O1 + B^2 O2 ... + B^q Oq
-->Parte MA
```

Se debe recordar que F(B) y O(B) están definidas dentro de la función “armax” (función que estima los parámetros del modelo ARMA (p, q)) en Matlab.

Se puede utilizar la predicción lineal para hacer la estimación sobre la base del mínimo error cuadrático de la secuencia $\hat{X}(h)$ que para redes de datos viene dado por la Ec. (7) [8]:

$$\hat{X}(h) = \sum_{j=1}^{\infty} \pi_j^{(h)} \hat{X}_{t+h-j} \quad (7)$$

Donde

$$\pi_j^{(h)} = \pi_{j+h-1} - \sum_{i=1}^{h-1} \pi_i \pi_j^{h-1} \quad \text{cuando } h > 1 \quad (8)$$

$$\pi_j^{(1)} = \pi_j$$

El cálculo del error cuadrático medio se define como:

$$\hat{\sigma}_t^2(h) = E(X_{t+h} - \hat{X}_t(h))^2 \quad (9)$$

El empleo del modelo para predecir el tráfico autosimilar en la red WI-FI de Chía, en ciertos puntos muestra una cercana relación entre los datos reales y los anticipados, como puede apreciarse en la Fig. 4 [9].

Tabla 2. Valores cuantitativos del porcentaje de error del modelo FARIMA

CANTIDAD DE PUNTOS COMPARADOS 240	
Coincidencia de puntos	% de error
68	71,6

Al modelo FARIMA se le adiciona la variable M para una mejor predicción del tráfico en redes de datos [10].

$$\hat{X}(h) = \sum_{j=1}^{\infty} \pi_j^{(h)} \hat{X}_{t+h-j-M} \quad (10)$$

Donde

$$\pi_j^{(h)} = \pi_{j+h-1} - \sum_{i=1}^{h-1} \pi_i \pi_j^{h-1}, \pi_j^{(1)} = \pi_j \quad (11)$$

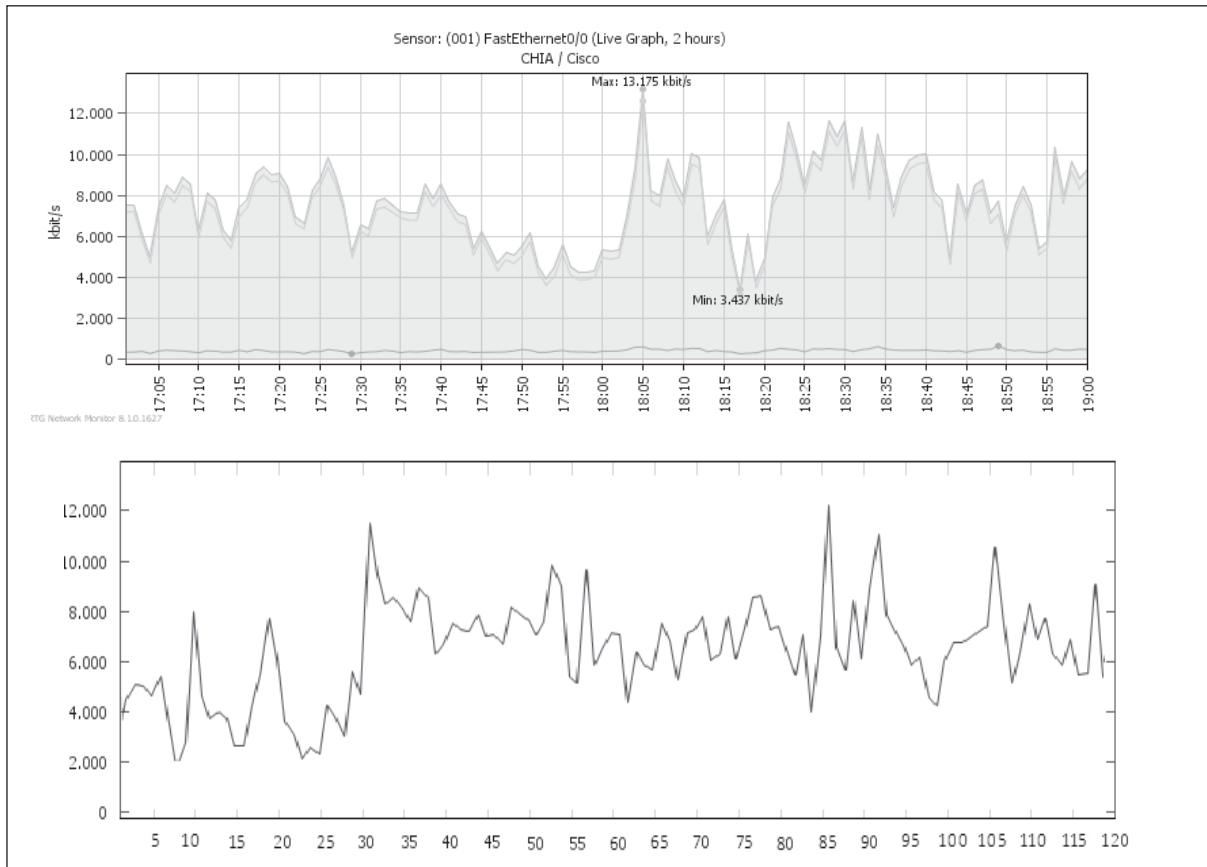


Fig. 4. Comparación del tráfico medido con el tráfico predicho utilizando S Farima.

Tabla 3. Valor del error cuadrático medio para el valor del corrimiento

VALOR DEL CORRIMIENTO	ERROR CUADRÁTICO MEDIO
M = 1	1.1233
M = 2	1.2345
M = 3	1.1234
M = 4	0.6901
M = 5	1.091
M = 6	1.1486

Utilizando la Ec. (9) se calcula el error cuadrático medio variando M hasta hallar el valor mínimo del error.

En la Tabla 3 se muestra cuáles son los valores del error para cada M desde M = 1 hasta M = 6, del tráfico medido en la red WI-FI de Chía (el valor de M = 0, no se evaluó por cuanto para dicha cuantía el modelo sería FARIMA).

Como se puede apreciar en la tabla, el valor mínimo del error cuadrático medio corresponde a M = 4

Tabla 4. Valores cuantitativos del porcentaje de error del modelo S FARIMA

CANTIDAD DE PUNTOS COMPARADOS 240	
Coincidencia de puntos	% de error
97	59,5

5. CONCLUSIONES

El modelo FARIMA tiene la ventaja que de acuerdo con el parámetro d se obtendría un modelo de corta, media o larga persistencia; sin embargo, el ajuste del modelo conlleva un mayor consumo de recursos con relación a otros modelos.

Para el caso del tráfico en redes WI-FI, el modelo idóneo para predecir tráfico futuro es SFARI-

MA, toda vez que aumenta la cantidad de puntos de coincidencia entre el tráfico medido y el estimado.

El presente documento explica de forma detallada cuál es la diferencia entre el modelo SFARIMA y el modelo FARIMA. Porque controlar el error cuadrático medio en el Modelo FARIMA mejora el proceso de predicción en el caso particular del tráfico WI-FI.

REFERENCIAS

- [1] J. Kwiatkowski, "Dynamic econometric models, Bayesian analysis of long memory and persistence using ARFIMA models with an application to Polish stock market", Nicholas Copernicus University in Toruń, vol. 6, 2004.
- [2] H. Ghauch, V. Khoury, A. Sabouri, "Self Similar Traffic Packet Switching & Computer Networks" [En línea]. Disponible en: <http://www.andrew.cmu.edu/user/hgg/docs/Self%20Similar%20Traffic.pdf>
- [3] Meter Lildhold, "Long memory an AR-FIMA modelling", University of Aarhus, 2000.
- [4] F. L. Herrera, F. Venegas, F. Sánchez, "Memoria larga de la volatilidad de los rendimientos del mercado mexicano de capitales", Universidad Autónoma de México, Análisis Económico, Núm. 56, vol. XXIV, Segundo cuatrimestre de 2009.
- [5] J. Zhigang, S. Yantai, L. Jiakun, "Prediction based network bandwidth control", Tianjin University, ieeexplore.ieee.org, 2000.
- [6] C. Orellana. "Larga Memoria en Retornos de Activos: Alguna Evidencia Internacional", Universidad Santiago de Chile, 2004.
- [7] M. Álzate, "Introducción al Tráfico Autosimilar en Redes de Comunicaciones", Universidad Distrital Francisco José de Caldas, [en línea]. Disponible en: <http://www.andrew.cmu.edu/user/hgg/docs/Self%20Similar%20Traffic.pdf>
- [8] P. Brockwell, R. Davis, *Time Series: Theory and Methods*, Colorado: Springer, 2nd Edition. 1991.
- [9] F. Wang, D. Li, Y. Zhao, Prediction of self similar traffic and its application in network bandwidth allocation, *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, 2007.
- [10] W. Pan, Z. Shun-Yi, C. Xue-Jiao, "SFARIMA: A New Network Traffic Prediction Algorithm", *Information Science and Engineering (ICISE), 2009 1st International Conference on*. 2009.