

MODELO DE TRÁFICO BASADO EN SERIES DE TIEMPO PARA PRONOSTICAR VALORES FUTUROS DE TRÁFICO EN UNA RED DE DATOS WI-FI

Cesar A. Hernández S.
Ingeniería Electrónica, Universidad Distrital,
Bogotá D.C., Colombia
Email: cahernandezs@udistrital.edu.co

Octavio J. Salcedo P.
Ingeniería de Sistemas, Universidad Distrital
Bogotá D.C., Colombia

Luis F. Pedraza M.
Ingeniería Electrónica, Universidad Militar
Bogotá D.C., Colombia

RESUMEN

El objetivo de esta investigación es demostrar que las series de tiempo son una excelente herramienta para el modelamiento de tráfico de datos en redes Wi-Fi. Para lograr este objetivo se utilizó la metodología de Box-Jenkins, la cual se describe en este artículo.

El modelamiento de tráfico Wi-Fi a través de modelos correlacionados como las series de tiempo permiten ajustar gran parte de la dinámica del comportamiento de los datos en una ecuación y con base en esto estimar valores futuros de tráfico. Lo anterior es una ventaja para la planeación de cobertura, reservación de recursos y la realización de un control más oportuno y eficiente en forma integrada a diferentes niveles de la jerarquía funcional de la red de datos Wi-Fi.

Como resultado de la investigación se obtuvo un modelo de tráfico ARIMA de orden 18, el cual realizó pronósticos de tráfico con valores del error cuadrático medio relativamente pequeños, para un periodo de 10 días.

Palabras Claves

ARIMA, Correlación, Modelo de tráfico, Serie de tiempo, Red de comunicaciones.

1 INTRODUCCION

A lo largo del desarrollo de las redes de comunicaciones en los últimos cien años, se han propuesto diferentes modelos de tráfico, cada uno de los cuales ha resultado útil dentro del contexto particular para el que se propuso. Sin embargo hoy en día se ha demostrado que el tráfico de datos es altamente correlacionado. El fenómeno de correlación hace que la variabilidad se extienda a muchas escalas de tiempo, comprometiendo la validez de las técnicas de control diseñadas para los modelos tradicionales de tráfico. Por esta razón, ha sido necesario desarrollar modelos adicionales de tráfico más complejos, capaces de representar estas correlaciones y que tengan en cuenta las características del tráfico real, en especial las correlaciones que existen entre los tiempos entre llegadas, completamente ausentes en los modelos no correlacionados. [2].

Actualmente las redes de comunicaciones modernas no tienen una herramienta confiable que permita pronosticar tráfico hacia el futuro, lo que se plantea el siguiente interrogante: [16], [23].

¿ES POSIBLE DESARROLLAR UN MODELO ESTADÍSTICO QUE PERMITA ESTIMAR PRONOSTICOS FUTUROS DE TRAFICO WI-FI?

Las series de tiempo tienen como objetivo central desarrollar modelos estadísticos que expliquen el comportamiento de una variable aleatoria que varía con el tiempo permitiendo estimar pronósticos futuros de dicha variable aleatoria. [11].

Por tanto los modelos de tráfico a través de series de tiempo son beneficiosos para: la planeación de cobertura, reservación de recursos, monitoreo de la red, detección de anomalías, y producción de modelos de simulación más exactos, en la medida en que pueden pronosticar el tráfico en un tiempo de escala determinado. [24].

En la planeación, para futuras necesidades de cualquier sistema, la exactitud en el pronóstico de tráfico, es realmente importante para definir capacidad futura requerida y planear los cambios. Un modelo de series de tiempo bastante exacto podría predecir varios años hacia el futuro, cuya habilidad es una ventaja para la planeación de futuros requerimientos. [14].

En el presente documento se busca:

- Desarrollar un modelo estadístico que permita estimar pronósticos futuros de tráfico en redes Wi-Fi a través del modelamiento en series de tiempo.
- Evaluar los diferentes modelos actuales más utilizados para el pronóstico de tráfico, relacionados con series de tiempo.

A continuación se desglosarán cada uno de los pasos realizados durante la investigación a fin de construir el modelo de tráfico para una red de datos Wi-Fi.

2 DESARROLLO DEL MODELO DE TRÁFICO

El desarrollo de este trabajo fundamenta su metodología en la de Box-Jenkins, entonces, una vez capturados los datos se analiza la estructura de correlación que presenta la muestra, de aquí se propone y se estima un modelo basado en ecuaciones en diferencias, el cual busca capturar la dinámica de la serie, de ser correcta la formulación del modelo, este se valida y después se pronostican las futuras observaciones. Con el fin de realizar un estudio profundo del modelamiento de tráfico de datos a través

de series de tiempo, se desarrollaran varios tipos de modelamiento de la serie.

Extracción de la Serie

El primer paso en el desarrollo de cualquier modelo de tráfico es tomar una muestra de datos, con los se pueda caracterizar el tráfico de un tipo de red predeterminada. En el presente estudio se decidió desarrollar un modelo de tráfico para una red con tecnología Wi-Fi, en razón a que esta tecnología es la que actualmente tiene mayor auge en el ámbito de las redes de datos debido a todas sus características, igualmente debido a que esta tecnología es relativamente nueva, no existen muchos estudios acerca del comportamiento de su tráfico, sin embargo como lo han confirmado varios estudios de tráfico en redes alámbricas e inalámbricas (como WiFi) el tráfico actual como el de Internet e incluso el de video, presenta características fuertes de correlación. [9], [10], [22], [24]. Los datos de tráfico fueron extraídos a través de la herramienta Netflow Analyzer, estos datos se capturaron como una variable de paquetes por segundo y se tomaron 889 muestras que corresponden a un poco más de 9 días con un intervalo de una muestra cada 15 minutos.

Netflow Analyzer es una herramienta de software basada en Web que permite el monitoreo del ancho de banda de cualquier red. Los datos son exportados de los dispositivos de enrutamiento hacia el Analyzer Network Traffic el cual reporta el ancho de banda utilizado en tiempo real a través de la red monitoreada. Debido a su carácter de herramienta de software basada en Web, permite la monitorización remota de cualquier red a través de los dispositivos de red (como Routers) que se encuentren configurados en este software.

A pesar de haberse extraído 889 datos de tráfico los modelos se desarrollaran únicamente con los 672 primeros (correspondientes a los siete primeros días), los datos restantes se utilizaron para comparar y evaluar los pronósticos de cada una de las series. [14].

Preprocesamiento de la Serie

Es frecuente que se presenten algunos valores perdidos dentro de la serie de tráfico. Estos generalmente se deben a varias razones entre las cuales se destacan las siguientes cuatro: (1) el router puede estar caído debido a mantenimiento del mismo, o por un reinicio accidental de este; (2) el router puede estar ocupado resolviendo solicitudes SNMP; (3) la conexión inalámbrica entre el router y la estación de monitorización de tráfico puede estar temporalmente perdida o caída; y (4) la solicitud y respuesta de los paquetes SNMP (Simple Network Management Protocol) puede haberse perdido ya que ellos son transportados usando el protocolo UDP (User Datagram Protocol es un protocolo no confiable). [4], [24].

Para el tratamiento de los valores perdidos existen dos opciones: (1) ignorar los datos perdidos, o (2) estimarlos los valores perdidos. Para la presente investigación se decidió estimar los valores perdidos con el fin de obtener un modelo más completo, en el cual se puedan observar factores de estacionalidad, tendencia o ciclo, para una muestra de tráfico de una semana.

El procedimiento adoptado para la estimación de valores perdidos se realizo a través de un software especializado para esta tarea, dicho software es el: SEATS TRAMO WINDOWS (TSW).

El procedimiento está fundamentado en la interpolación de promedios de datos de tráfico previos, estas interpolaciones se realizan examinando el patrón de tráfico y tratando en lo posible de mantener dicho patrón dentro del subconjunto de valores tráfico actuales. En algunas ocasiones se decide remplazar el valor

perdido (o valores perdidos) por uno del mismo día, hora y minuto, pero de una semana anterior o una semana posterior, esto se justifica en el fuerte patrón semanal que muestra la serie de tiempo de los datos de tráfico, en esta ocasión esto no es posible en razón a que la muestra de datos de tráfico en este estudio es para una sola semana y dos días. [4].

En la figura 1 se puede observar la serie de tráfico Wi-Fi obtenida para los 899 datos capturados.

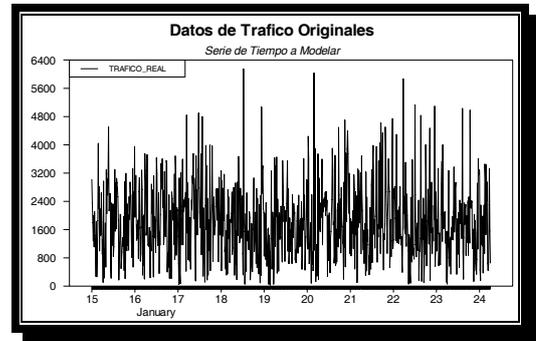


Figura 1. Serie de tiempo del tráfico Wi-Fi extraído

Identificación del Modelo

Debido que uno de los objetivos es comparar diferentes modelos correlacionados, es decir construir varios modelos de tráfico basados en diferentes tipos de series de tiempo y analizar cuál de ellos es el mejor estimador del tráfico Wi-Fi capturado. Esta etapa de identificación carece de sentido para esta investigación, ya que sin importar las conclusiones de la identificación del modelo, se desarrollaran cuatro modelos de tráfico correlacionados: (1) Modelo AR (Autoregresivo), (2) Modelo MA (Promedios Móviles), (3) Modelo ARMA (AutoRegresivo y Promedio Móviles), y (4) Modelo ARIMA (AutoRegresivo e Integrado de Promedio Móvil). Vale la pena recordar que los tres primeros modelos correlacionados implican la necesidad de que la serie sea estacionaria, mientras que para el cuarto modelo no es necesario que exista estacionariedad en la serie de tiempo. [6].

El concepto de estacionariedad es importante para el análisis de series de tiempo. En general, para caracterizar completamente un proceso estocástico es necesario conocer la función de densidad conjunta de sus variables aleatorias; sin embargo en la práctica no es realista pensar que esto pueda lograrse con una serie de tiempo. Como se mencionó anteriormente en lo que respecta a la covarianza, no existe dependencia del tiempo pero si de la separación (k) que hay entre las variables. Lo anterior conduce a pensar que la serie, mostrará el mismo comportamiento en términos generales sin importar el momento en el que se observe. Esto es, si se graficara un cierto número de observaciones contiguas de una serie, la gráfica que se obtendría sería bastante similar a la que se lograría al graficar el mismo número de observaciones contiguas pero k periodos hacia delante o hacia atrás de los considerados inicialmente. [7], [18].

De la observación de la serie de tiempo que nos atañe, se puede concluir fácilmente que esta no es estacionaria, debido a que su media no es cero (observar figura 1), aunque su varianza parece constante, dos condiciones imperiosas para que una serie de tiempo sea estacionaria (estos cálculos se desarrollaran más adelante).

De acuerdo a lo anterior los tres primeros modelos cuya condición es la estacionariedad de la serie, no tendrían mucho sentido. Por esto el único modelo que se ajusta a la serie de tiempo

es el ARIMA, con esto la etapa de identificación esta terminada. Sin embargo debido a que el objetivo es comparar los cuatro modelos se desarrollara primero el modelo ARIMA, el cual cuenta con la ventaja de ser integrado "I" permitiendo volver estacionaria una serie no estacionaria (por tendencia); una vez realizado el modelo ARIMA, se tomó la serie estacionaria y con ella se estimaron los modelos AR y MA, que para este caso se convierten en los modelos ARI e IMA en un sentido estricto de la palabra, el modelo ARMA ya carecería de sentido porque al calcularlo se llegaría al mismo modelo ARIMA. [5].

Estimación de Parámetros y Validación

Modelo ARIMA: para verificar la no estacionariedad de la serie de tiempo se realizó la prueba de Raíz Unitaria de Dickey –Fuller a través del software RATS (Regression Analysis of Time Series), estos son los resultados obtenidos: [12], [13], [15].

*Dickey-Fuller Unit Root Test, Series TRAFICO_REAL
Regression Run From 2007:01:15//62 to 2007:01:21//95
Observations 828
With intercept with 60 lags
T-test statistic -1.49785
Critical values: 1%= -3.443 5%= -2.867 10%= -2.569*

Según el criterio de Dickey – Fuller, la serie no es estacionaria, ya que el valor absoluto del test es menor al valor absoluto del valor crítico del 5%. Lo anterior era consecuencia de que la media de los datos no es cero como se mostró en la figura 1, aunque la varianza parece constante. Para volver esta serie estacionaria se diferenciará la serie y se volverá a realizar la prueba de raíz unitaria.

Entonces se tomó la serie de tiempo original con sus primeros 672 datos de tráfico y se diferenció inicialmente una sola vez, luego se diferenció dos veces y por último se diferenció una vez el logaritmo de la serie de tiempo original, para cada una de las series resultantes anteriormente luego de la diferenciación, se realizó la prueba de Dickey –Fuller para verificar su respectiva estacionariedad.

Del proceso anterior se concluyó que la mejor transformación es la diferenciación de la serie una sola vez con esto se obtuvo una serie como la que muestra la figura 2 (observe que esta vez la media de los datos si es cero).

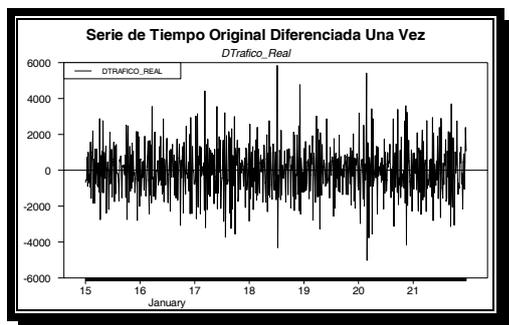


Figura 2. Serie de tiempo diferenciada una vez

Si solo se han aplicado primeras diferencias será un ARIMA(p,1,q), si requiere segundas diferencias es un ARIMA(p,2,q), en general si se aplica (1-B)d se llega a un ARIMA(p, d, q). Para el desarrollo de esta investigación se tiene entonces d=1. [7].

Se debe evitar el sobre-diferenciar la serie original y eliminar información valiosa que se manifestaría en la función de autocorrelación, ya que en un caso de sobre-diferenciación las autocorrelaciones se hacen aún más complicadas, y el modelo pierde parsimonia, se incrementa la varianza y se pierden d-observaciones. [8].

Ahora que ya se tiene una serie estacionaria se deben determinar el orden de "p" (autoregresivo) y de "q" (promedio móvil); para esto se utilizó la función de autocorrelación y la función de autocorrelación parcial. [1], [3], [12].

Para obtener estas gráficas de autocorrelación y autocorrelación parcial se utilizó el software RATS con el cual se generaron las gráficas que se muestran en la Figura 3. Estas dos gráficas (FAC y FACP) permiten estimar el valor de "p" y "q" para construir el modelo ARIMA(p,d,q) que nos interesa. Entonces de la FAC se obtiene, q=22, y de la FACP se obtiene, p=22, como la serie se diferenció finalmente solo una vez, entonces d=1. Ósea que finalmente se tiene un modelo inicial ARIMA (22,1,22).

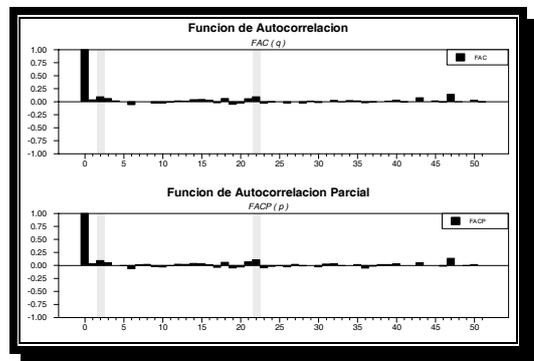


Figura 3. FAC y FACP de la serie de tráfico Wi-Fi.

Según los resultados de la FAC y la FACP se tiene el modelo presentado en la Ec. (1), para el cual aun no se conocen sus coeficientes.

$$Z_t = \theta_2 \times Z_{t-2} + \theta_{22} \times Z_{t-22} + a_t - \theta_2 \times a_{t-2} - \theta_{22} \times a_{t-22} \quad (1)$$

Ahora que se tiene un fuerte candidato, hay que estimar sus parámetros. En la práctica esta es una labor de cómputo, aquí se debe escoger el paquete a usar, para este estudio se eligió el software RATS (Por encima inclusive del Software Eviews) por la flexibilidad que ofrece y por su gran potencialidad y estimación de máxima verosimilitud.

Lo usual es, pasar de la estimación inicial al análisis de los residuos, aquí se vuelve a buscar picos pero ahora en los residuos. Estos picos revelan términos que uno debe incluir en la nueva formulación del modelo ARIMA que se volverá a estimar. Este ciclo de re-especificación dinámica termina cuando los residuos ya no presentan correlaciones (picos) y en ese momento se puede decir que son residuos son ruido blanco. [6], [20], [21].

Realizando lo anterior en el software RATS se obtuvo la primera estimación de parámetros del modelo, es decir los coeficientes del modelo ARIMA que se presentan en la Ec. (2), y cuyos valores se listan en la Ec. (2).

$$Z_t = -0.385785636 \times Z_{t-2} + 0.546189772 \times Z_{t-22} + a_t + 0.451619226 \times a_{t-2} - 0.504850197 \times a_{t-22} \quad (2)$$

En la Figura 4 se grafica el modelo estimado en la Ec. (2) en función del tiempo para el cual se extrajeron los datos de tráfico correspondientes a una semana completa. Sin embargo, de la simple inspección de dicho modelo no es posible validarlo. La validación de los modelos de series de tiempo se realiza a través de la verificación de correlación entre los residuales de dicho modelo, para lo cual es necesario aplicar la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP) a dichos residuos.

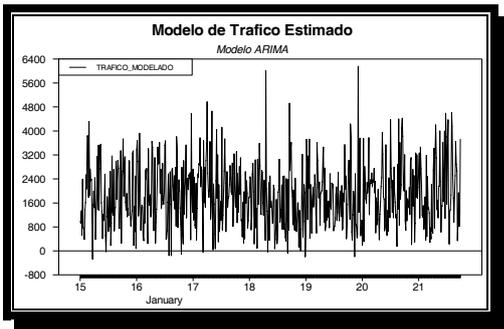


Figura 4. Serie de tráfico obtenida a partir de la Ec. (2).

Las FAC y FACP de los residuales del modelo estimado en la Ec. (2), se pueden observar en la Figura 5. De esta figura se puede observar que existe correlación entre los residuales del modelo de la Ec. (2), por lo tanto este modelo no incluye toda la dinámica de la serie de tiempo. En este punto es necesario volver a iterar y estimar nuevamente los coeficientes del modelo incluyendo ahora los nuevos valores de “p” y “q” que sugiere las funciones de autocorrelación y autocorrelación parcial, ósea para “p” los valores: 1, 2, 3, 4, 6, 7, 11, 12, 13, 17, 20, y para “q” los valores: 1, 18. De acuerdo con lo anterior se realiza una nueva estimación de los coeficientes incluyendo los nuevos parámetros de “p” y “q”.

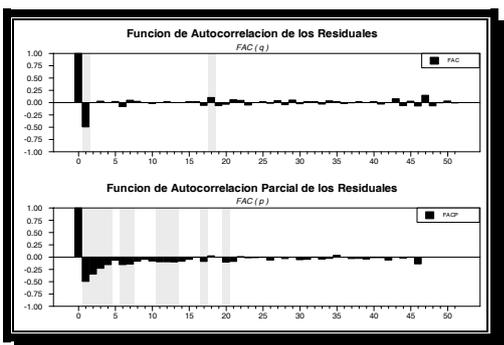


Figura 5. FAC y FACP para los residuales de la Ec. (2).

El procedimiento anterior se lleva a cabo hasta que la FAC y FACP de los residuales demuestren que no existe correlación alguna entre los residuales del modelo estimado, lo cual se logra luego de 3 iteraciones adicionales, sin embargo el número de parámetros obtenidos para el modelo correspondiente fue 18. Un modelo con un número grande de parámetros es un modelo que no presenta una buena parsimonia, por lo que se analizó el nivel de significancia de cada parámetro eliminando aquellos que presentan un valor mayor al cinco por ciento, ya que estos

no resultan significativos para el modelo. Una vez realizado lo anterior es necesario validar nuevamente el modelo y dependiendo del resultado iterar una vez más.

Finalmente se llegó a un modelo definitivo descrito por la Ec. (3) y cuyo comportamiento se muestra en la figura 6.

$$Z_t = -0.092467147 \times Z_{t-6} - 0.090197053 \times Z_{t-10} + 0.093022957 \times Z_{t-18} + a_t - 0.978862096 \times a_{t-1} - 0.009496113 \times a_{t-10} - 0.009945630 \times a_{t-18} \quad (3)$$

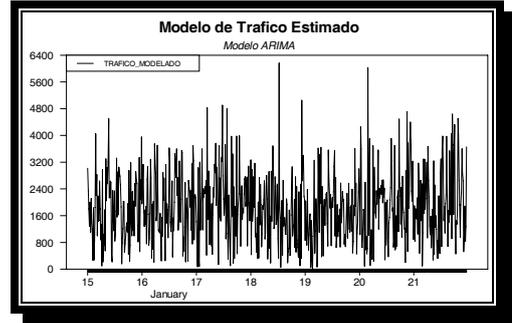


Figura 6. Serie de tráfico del modelo de la Ec. (3).

Con el fin de realizar una comparación subjetiva entre los datos estimados por el modelo desarrollado en la Ec. (3) y los datos de tráfico reales, se muestra la figura 7, donde las trazas azules indican los datos de tráfico estimados y las negras reales.

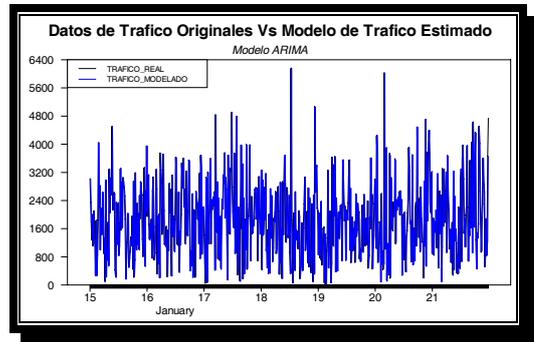


Figura 7. Datos de tráfico estimados vs datos originales.

Como conclusión final se obtiene un modelo ARIMA(18,1,18) de seis parámetros el cual esta definido por la Ec. (3) y que por inspección visual de la Figura 7, modela bastante bien los 672 datos de tráfico Wi-Fi. El mismo procedimiento se lleva a cabo para el modelo ARI e IMA, presentados en las Ec. (4) y Ec. (5) respectivamente. En la sección 3 se realiza una evaluación cuantitativa del presente modelo.

$$Z_t = -0.90995170 \times Z_{t-1} - 0.74409755 \times Z_{t-2} - 0.627510 \times Z_{t-3} - 0.5740204 \times Z_{t-4} - 0.5230094 \times Z_{t-5} - 0.551627 \times Z_{t-6} - 0.4832262 \times Z_{t-7} - 0.4072993 \times Z_{t-8} - 0.380864 \times Z_{t-9} - 0.3774092 \times Z_{t-10} - 0.3385981 \times Z_{t-11} - 0.257110 \times Z_{t-12} - 0.1876921 \times Z_{t-13} - 0.0969909 \times Z_{t-14} - 0.056301 \times Z_{t-17} - 0.0624545 \times Z_{t-19} - 0.0923661 \times Z_{t-20} + 0.087331 \times Z_{t-22} \quad (4)$$

$$Z_t = +a_t - 0.91525888 \times a_{t-1} - 0.04432095 \times a_{t-6} \quad (5)$$

3 EVALUACION DEL MODELO DE TRAFICO

Es usual encontrarse con varios modelos alternativos y se debe decidir cual escoger. En primera instancia se tiene la función de autocorrelación y la función de autocorrelación parcial, las cuales se aplican a los residuos del modelo final para determinar si existe correlación entre estos, de no tener correlación se puede decir que el modelo ha sido validado con éxito. Sin embargo en este estudio no se obtuvo correlación entre los residuos para ninguno de los tres modelos desarrollados por lo que entonces la pregunta sería ¿cual escoger? [7].

Debido a lo anterior se analizaran otros criterios a parte del análisis residual para seleccionar adecuadamente un modelo:

- Criterio de calidad de ajuste.
- Criterio de parsimonia.
- Criterios estadísticos.

Calidad de Ajuste

La calidad de ajuste de un modelo esta definida como la suma de los cuadrados de los residuos dividida por el tamaño de la muestra, y su objetivo es medir la habilidad del modelo para reproducir los datos de la muestra, es decir verifica que tan parecida es la serie modelada con la real. [17].

En la Tabla 1 se muestran los valores de la calidad de ajuste para cada modelo desarrollado.

Tabla 1. Calidad de ajuste de los modelos desarrollados.

Modelo Desarrollado	Modelo ARIMA	Modelo ARI	Modelo IMA
Calidad de Ajuste	171,4137326	20489,69984	976512,583

Parsimonia

La idea de parsimonia es que un buen modelo tiene pocos parámetros ya que ha capturado las propiedades intrínsecas de la serie que se analiza, un modelo complicado con demasiados parámetros es un modelo sin parsimonia. Desde este punto de vista, el modelo ARI que se obtuvo anteriormente es un modelo sin parsimonia, debido a la gran cantidad de parámetros que presenta (en total 18 parámetros) en contraste con el modelo ARIMA (en total 6 parámetros) y el modelo IMA (en total 2 parámetros).

Se podría concluir que el modelo IMA es el que presenta mayor parsimonia, inclusive mayor a la del modelo ARIMA. Sin embargo este criterio siempre deberá ser el último que utilice para seleccionar un modelo, debido a su carácter cualitativo y no cuantitativo como si lo es el criterio de calidad de ajuste y los que se describen a continuación.

Criterios Estadísticos

A pesar de poder seleccionar un modelo adecuado a partir de los criterios de análisis de residuos, calidad de ajuste y parsimonia, se calculó también varios criterios estadísticos que permitieran realizar objetivamente un análisis comparativo entre los modelos de serie de tiempo desarrollados. Los estadísticos calculados son:

- Error cuadrático medio
- Valor absoluto de la desviación estándar
- Coeficiente de correlación

Entre estos estadísticos el más significativo es la desviación estándar en valor absoluto, para esta investigación se decidió calcular el error cuadrático medio como el promedio de las desviaciones estándares al cuadrado de los valores estimados con respecto a los originales, con el fin de obtener un valor cuantitativo de la exactitud del modelo, ya que por definición el error cuadrático medio tendría el mismo valor del criterio calidad de ajuste, el cual no nos dice que tan eficaz es el modelo, solo nos permite compararlo con otros.

Debido a que el promedio de las desviaciones estándares de cada dato estimado no es significativamente objetivo en razón a que puede tomar valores positivos como negativos que afectan el resultado final, se decidió tomar el promedio del valor absoluto de las desviaciones estándares de cada dato.

Y por ultimo se calculo el factor de correlación entre los datos estimados y los originales, debido a que como se explico anteriormente este estadístico da un indicativo del nivel de relación entre dos variables, lo cual no puede lograr la función de covarianza.

Los resultados de los criterios estadísticos mencionados se muestran en la tabla 2.

Tabla 2. Criterios estadísticos.

Modelo Desarrollado	Modelo ARIMA	Modelo ARI	Modelo IMA
Error Cuadrático Medio	0,1248%	15,664%	1284,12%
Desviación Estándar en Valor Absoluto	1,2133%	13,979%	122,178%
Coeficiente de Correlación	0,99992	0,99110	0,4454688

Como conclusión de los resultados obtenidos en cada criterio se decidió seleccionar el modelo ARIMA como la mejor opción para modelar los datos de tráfico Wi-Fi capturados.

Evaluación de los Pronósticos

En la realidad se puede sostener la tesis de que un modelo es realmente útil solo en la medida que anticipa la evolución de la variable. En este sentido se está aceptando que se espere a que vengan las futuras observaciones para después analizar la calidad del modelo. Esto se denomina una evaluación ex-post, y es una validación más fuerte en el sentido común, que el análisis de residuos. [18]. [19].

Tabla 3. Evaluación de Pronósticos.

Modelo Desarrollado	Modelo ARIMA	Modelo ARI	Modelo IMA
Error Cuadrático Medio	2,2665%	26,098%	1366,14%
Desviación Estándar en Valor Absoluto	5,9327%	21,842%	131,173%
Coeficiente de Correlación	0,998136	0,981707	0,382241
Calidad de Ajuste	4306,278	29074,16	1044223,

Para cada modelo se pronosticaron 217 (217=889-672) datos de tráfico que fueron respectivamente comparados con los datos de tráfico originales, para generar la Tabla 3, donde se muestran la exactitud de los pronósticos en función de estadísticos como el error cuadrático medio, el promedio de la desviación estándar en valor absoluto, el coeficiente de correlación y la calidad de ajuste, lo cual permite observar detalladamente la eficacia de cada modelo desarrollado para pronosticar los respectivos datos de tráfico.

Es realmente interesante analizar la capacidad de predicción que posee el modelo ARIMA ya que tan solo tiene un 6% de error en promedio de los pronósticos de los datos de tráfico para los dos días siguientes, según la desviación estándar en valor absoluto. Pero ¿Que tanto tiempo a futuro podría mantener tan fantásticos pronósticos?

4 CONCLUSIONES

Las series de tiempo y en especial los modelos Autorregresivos e Integrados con Promedios Móviles (ARIMA), resultan realmente apropiados para modelar el tráfico moderno en redes de datos Wi-Fi con características de correlación fuertes. La evaluación de este estudio del modelo ARIMA desarrollado y seleccionado finalmente como el más apropiado, exhibe un desempeño bastante alto con relación a la magnitud de sus residuales, los cuales no experimentaron correlación alguna.

No todos los modelos de series de tiempo caracterizan apropiadamente el tráfico Wi-Fi, en especial para la serie de tráfico Wi-Fi de este estudio, las series de tiempo como la de Promedios Móviles (MA) no logra capturar toda la dinámica de la serie; otros modelos como el AutoRegresivo (AR) proporcionan realizaciones bastante parecidas en magnitud a las originales, pero con un número grande de parámetros que reduce significativamente su parsimonia y ende aumenta el costo computacional; generalmente las series de tráfico correlacionadas como la que nos ocupa, exhibe uno o mas parámetros de una componente autorregresiva y al mismo tiempo uno o mas parámetros de una componente de promedio móviles, sugeridos por las funciones de autocorrelación y autocorrelación parcial, proponiendo como fuerte candidato un modelo ARIMA, la inclusión de ambas componentes tiene por objetivo reducir significativamente el número de parámetros necesarios para modelar la serie original, tal y como se demostró en el desarrollo de dicho modelo. Las evaluaciones realizadas a estos modelos basados en series de tiempo implicaron en su totalidad algún tipo de estadístico alrededor de los residuales, para los cuales, el modelo ARIMA obtuvo no solo los mejores resultados en comparación con los demás modelos de series de tiempo, sino además un alto grado de desempeño, confiabilidad y exactitud en sus pronósticos.

REFERENCIA BIBLIOGRAFICAS

[1] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. Second international symposium on information theory. Budapest. 1973. p. 267-281.

[2] ALZATE, Marco Aurelio. Modelos de tráfico en análisis y control de redes de comunicaciones. En: Revista de ingeniería de la Universidad Distrital Francisco José de Caldas. Bogotá. Vol. 9, No. 1 (Junio 2004); p. 63-87.

[3] ANDERSON, T. W. Maximum likelihood estimation for vector autoregressive moving-average models, directions in time series. Institute of mathematical statis-

tics. 1980. p. 80-111.

[4] ANSLEY, C. F. y KOHN, R. On the estimation of ARIMA models with missing values, Time series analysis of irregularly observed data. Editorial Parzen, 1985. p. 9-37.

[5] BOX, G. E. P. y COX, D. R. An analysis of transformations. Stat. soc. 1964. p. 211-252.

[6] BOX, G. E. P. y JENKINS, Gwilym M. Time series analysis: Forecasting and control. Revised Edition. Oakland, California: Editorial Holden-Day, 1976.

[7] BRILLINGER, D.R. Time series: data analysis and theory. Universidad de California. Holden-Day. SIAM. 2001.

[8] BROCKWELL, P.J. y DAVIS, R.A. Introduction to time series and forecasting. Second edition. New York: Editorial Springer, 2002.

[9] CASILARI, E.; REYES, A.; LECUONA, A.; DIAZ ESTRELLA, A. y SANDOVAL, F. Caracterización de tráfico de video y tráfico Internet. Universidad de Málaga, Campus de Teatinos. Málaga. 2002.

[10] CASILARI, E.; REYES, A.; LECUONA, A.; DIAZ ESTRELLA, A. y SANDOVAL F. Modelado de tráfico telemático. Departamento de Tecnología electrónica, E.T.S.I. telecomunicación. Universidad de Málaga, Campus de Teatinos. Málaga. 2003.

[11] CORREA MORENO, Emilia. Series de tiempo: conceptos básicos. Medellín: Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de matemáticas, 2004.

[12] DAVIS, R. A. Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. In: Econometric theory. Vol. 12, 1996. p. 1-29

[13] DICKEY, D. A. and FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. J. Amer. stat. assoc. Vol. 74, 1979. p. 427-431.

[14] FILLATRE, Lionel; MARAKOV, Dimitry y VATON, Sandrine. Forecasting seasonal traffic flows. Computer Science Department, Brest, Paris. 2003.

[15] FULLER, W. A. Introduction to statistical time series. A survey, int. stat. Review, Vol. 53, 1976. p. 301-329.

[16] GROSSGLAUSSER, M. y BOLOT, J. C. On the relevance of long-range dependence in network traffic source. En: IEEE/ACM Trans. Networking 7. 1999.

[17] GUERRERO GUZMAN, Víctor Manuel. Análisis estadístico de series de tiempo económicas. Segunda edición. México: Editorial Thomson, 2003.

[18] HAMILTON, James D. Time series analysis. New Jersey: Princeton university press, 1994. p. 25-152.

[19] HARMANTZIS, F. C. y HATZINAKOS, D. Heavy network traffic modeling and simulation using stable FARIMA processes. Stevens institute of technology, Castle point on the Hudson. Hoboken. 2005.

[20] JONES, R. H. Multivariate autoregression estimation using residuals, applied time series analysis. New York: Academic Press, 1978. p. 139-162.

[21] MAKRIDAKIS, Spyros G.; WHEELWRIGHT, Steven C. y HYNDMAN, Rob J. Forecasting: methods and applications. Tercera edición. Editorial Wiley, 1997.

[22] OLEXA, Ron. Implementing 802.11, 802.16, and 802.20 Wireless Networks: Planning, Troubleshooting, and Operations. Editorial Newnes. 2004.

[23] PAJOUH, Danech. Methodology for traffic forecasting. The French National Institute for Transport and Safety Research (INRETS). Arcueil. 2002.

[24] PAPAPOPOULI, María; SHENG, Haipeng; RAFTOPOULOS, Elias; PLOUMIDIS, Manolis y HERNANDEZ, Felix. Short-term traffic forecasting in a campus-wide wireless network. 2004.