

Métodos de modelamiento y predicción de tráfico orientados a plataformas de transmisión de video e IPTV usando series de tiempo¹

Modeling and forecasting methods of traffic oriented platform and IPTV video transmission using time series

Modelagem e métodos de previsão de tráfico plataforma orientado e IPTV transmissão de vídeo através de séries

Recibido: diciembre de 2011
Aceptado: mayo de 2012

Bayron VillanuevaOcampo
Danilo López Sarmiento
Edwin Rivas Trujillo²

Resumen

En este artículo se hace una investigación de las principales técnicas que existen para modelar y predecir el tráfico de video de forma estadística, enfocándose en los modelos que usan series de tiempo con el fin de identificar cuáles de estos se acomodan mejor al tráfico estocástico representativo de los sistemas IPTV. Para tal fin, se hace una introducción al análisis a través de series de tiempo, y una presentación del estado del arte acerca de modelamiento de tráfico de video sobre redes de datos. De la investigación se concluye que, de los modelos que permiten describir y predecir el tráfico futuro sobre redes de datos, los que se ajustan en una mayor medida a sistemas IPTV son modelos basados en series ARIMA, de estos, el modelo SARIMA podría describir de forma más precisa las tendencias periódicas del tráfico IPTV.

Palabras clave: ARIMA, IPTV, modelo de tráfico, SARIMA, series de tiempo.

Abstract

This paper, intends to review the most important techniques that allow performing statistic video traffic modeling and forecasting, focusing in time series models, so we can identify which models are better to describe the representative IPTV stochastic traffic. For this purpose, we make a short introduction to time series analysis, and a review of the state of the art on video traffic modeling over data networks. From this research we conclude that, of all the available models to describe and forecast network traffic, the more appropriate to use within IPTV systems are ARIMA time series models, from which SARIMA model are the best option.

Keywords: ARIMA, IPTV, SARIMA, Time Series, Traffic model

¹ Artículo de revisión.

² Contactos: bfvillanuevao@corre.udistrital.edu.co; dalopezs@udistrital.edu.co; erivast@udistrital.edu.co. Facultad de Ingeniería Universidad Distrital Francisco José de Caldas. Bogotá-Colombia.

Resumo

Este artigo tem como objetivo revisar as principais técnicas existentes para a modelagem e previsão de tráfego estatisticamente vídeo, com foco em modelos usando séries temporais, a fim de identificar quais destes são o tráfego estocástico mais adequado representante sistemas IPTV. Para este fim, uma breve introdução à análise por meio de séries temporais, e uma revisão do estado da arte em modelagem de tráfego de vídeo através de redes de dados. A investigação concluiu que, dos modelos para descrever e prever o futuro de tráfego em redes de dados, que são ajustados a uma maior extensão de sistemas de IPTV são baseados em modelos da série ARIMA, estes modelo SARIMA poderia descrever em mais preciso do tráfego periódico tendencias IPTV

Palavras-chave: ARIMA, IPTV, modelo de tráfego, SARIMA, Tempo Series.

Introducción

La televisión transmitida a través de redes IP, mejor conocida como IPTV, presenta un novedoso servicio que ofrece un conjunto de características adicionales —interactividad, convergencia, portabilidad, entre otros— (KwangHee, Won Mo, Jong Koo yKwang Sik, 2011), (Park, Seung, Sim Kwon yHo Young, 2008) que elevan la experiencia de los usuarios a nuevos niveles, lo que otros proveedores de televisión no pueden ofrecer por completo. Tal vez, debido a ellos, se explique el gran acogimiento y el rápido crecimiento del servicio IPTV alrededor del mundo; nuevos datos publicados por TeleGeography, compañía dedicada a la investigación de mercado en el área de telecomunicaciones, muestran que la penetración de IPTV alrededor del mundo ha alcanzado el 15% del total de suscriptores, lo que significa que actualmente existen más de 67 millones de suscriptores afiliados a este servicio. La figura 1 muestra los índices de penetración de IPTV en comparación a medios convencionales de transmisión de TV (TeleGeography, 2012).

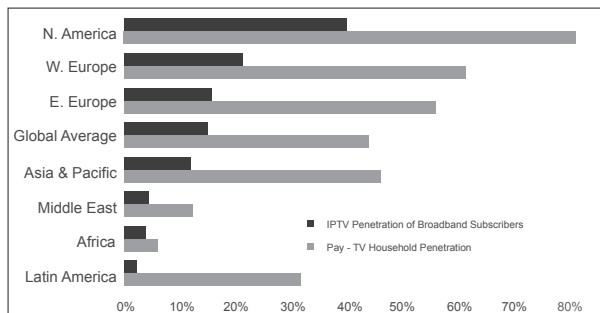


Figura 1. Penetración del servicio de IPTV alrededor del mundo.

Fuente: TeleGeography.

Este crecimiento ha sido mayor en las regiones más desarrolladas, estando Norteamérica y Europa entre las regiones en las cuales se ha extendido más este servicio, y quedando Latinoamérica bastante rezagada, por debajo incluso de continentes como África. Estas cifras muestran el rápido crecimiento de IPTV, que se espera siga creciendo de manera exponencial en todas las regiones. Por esta razón, se debe continuar realizando investigaciones acerca de este servicio, para poder satisfacer a los clientes y elevar su calidad de experiencia.

El término IPTV proviene de Internet Protocol Television y la ITU-T (2012) la define como “un conjunto de servicios multimedia tal como televisión/video/audio/texto/ gráficos/datos, que son transportados sobre redes basadas en IP, y gestionadas para proveer ciertos niveles de calidad de servicio, seguridad, interactividad y confianza”.

Desde la aparición de las primeras redes de comunicación, el ser humano ha buscado formas de modelar de forma matemática el tráfico que transita a través de estas redes, con el fin de identificar estadísticamente las características que describen su comportamiento. Inicialmente, las primeras redes modeladas fueron las telefónicas, que usaban modelos como el de Erlang (1909), el cual utiliza distribuciones poissonianas; pero este tipo de modelos, aunque bastante precisos sobre redes telefónicas, no eran lo suficientemente complejos para describir el tráfico actual de las redes de datos IP, y mucho menos el tráfico de los complejos sistemas de IPTV.

Actualmente se han desarrollado todo tipo de modelos para describir de forma precisa el tráfico sobre redes de datos, basándose en el hecho de que este tipo de tráfico presenta un alto índice de autocorrelación, dependiendo, en cierta medida, de sus valores pasados. Entre estos se encuentran los modelos de dependencia de corto alcance (SRD), como los clásicos markovianos (Markov, 1971), y los de dependencia de largo alcance (LRD), dentro de los cuales encontramos los fractales y los basados en series de tiempo (Yule, 1927).

Las series de tiempo son una herramienta estadística interesante, que permite modelar el comportamiento de diversos tipos de variables aleatorias, permitiendo caracterizar su comportamiento y, adicionalmente, a diferencia de otros modelos, permite predecir los posibles valores futuros que tendrá una determinada variable aleatoria, basándose únicamente en los valores pasados de dicha variable (Box y Jenkins, 1970). Desde su creación, el análisis de las series de tiempo ha sido usado ampliamente en diversas áreas del conocimiento. Fueron inicialmente usadas para resolver problemas de economía, pero se han usado también en ciencias físicas, en estudios demográficos, procesos de control y últimamente han sido extensamente utilizadas para modelar y predecir el tráfico en diversas redes de datos basadas en IP.

La intención de este artículo es determinar las principales técnicas de análisis de series de tiempo, con el fin de ser capaces de identificar el modelo más apropiado a la hora de caracterizar el tráfico sobre una red IPTV, para lo cual se investigan los fundamentos teóricos de diversas series de tiempo en la segunda sección, mientras que en la sección tercera se hace una revisión de los trabajos relacionados en el análisis de tráfico sobre redes encargadas de transportar tráfico multimedia y redes IPTV, haciendo énfasis en trabajos que utilicen series de tiempo para modelar y predecir el tráfico. En la cuarta sección se realiza una discusión acerca de series de tiempo para modelar el tráfico en redes IPTV y se hace un análisis de qué tipo de serie de tiempo podría modelar de manera adecuada dicho tráfico. Finalmente, en la quinta sección se presentan las conclusiones y los posibles trabajos futuros.

Series de tiempo

El objetivo principal del análisis, a través de series de tiempo, es desarrollar modelos matemáticos que provean descripciones plausibles para datos de muestra que aparentemente fluctúan de manera aleatoria en el tiempo. Para ello, se define una serie de tiempo como una colección de variables aleatorias indexadas de acuerdo al orden en que fueron capturadas en el tiempo (Shumway y Stoffer, 2006). La mayoría de estas series de tiempo corresponden a procesos estocásticos, donde sus valores futuros son, en parte, determinados por sus valores pasados, por lo que realizar predicciones de sus valores futuros es posible, pero con cierto grado de incertidumbre (Chatfield, 1999).

Existen distintos objetivos cuando se analizan las series de tiempo (Chatfield, 2000), como se puede ver en los distintos trabajos relacionados mencionados en este artículo. Los principales son:

- Descripción: usualmente el primer paso al analizar series de tiempo es representar los datos tomados y obtener simples medidas descriptivas de las principales propiedades de la serie (Montgomery, Jennings y Kulahci, 2008). Muchos autores han dedicado su investigación teniendo este objetivo en mente, para poder caracterizar el tráfico que quieren describir a través de un modelo robusto y preciso.
- Predicción: otra gran ventaja que permiten las series de tiempo es poder predecir de forma precisa los valores futuros de la serie, usando para esto únicamente sus valores pasados (Hamilton, 1994).
- Control: cuando una serie de tiempo se realiza para medir un aspecto acerca de la calidad de un proceso, el objetivo de su análisis puede ser el controlar dicho proceso. En esta aplicación de las series del tiempo se usa la predicción, entonces, un proceso es controlado por los valores futuros predichos por la serie de tiempo hecha sobre la variable a controlar. Al-Tamimi, Jain y So-In, proponen un esquema de asignación de recursos de manera dinámica, que soporte el tráfico de video en tiempo real, para

esto utilizan análisis de series de tiempo estacional (Al-Tamimi, Jain & So-In, 2010).

Series de tiempo estacionarias

Una clase importante de procesos estocásticos son aquellos que son estacionarios. Se dice que una serie de tiempo es estacionaria si sus propiedades estadísticas, como la media, varianza y autocovarianza son todas constantes en el tiempo. Una serie de tiempo estacionaria es relativamente fácil de predecir, debido a que sus propiedades estadísticas son las mismas tanto en el futuro como en el pasado.

Existen tres principales modelos de series de tiempo dedicados a estudiar y predecir los valores futuros de determinadas series de tiempo estacionarias, estos son:

- **Modelo autorregresivo:** se basa en la idea de que los valores presentes de una serie de tiempo pueden ser explicados como una función de p valores pasados, ponderados por un factor ϕ_p , conocido como el coeficiente autorregresivo, donde p determina el número de pasos necesitados en el pasado para predecir el valor actual; sumados a una función que representa las perturbaciones aleatorias del sistema y que se modela como ruido blanco; este modelo se denota como AR(p) donde p es su orden.
- **Modelo de promedios móviles:** considera que el valor de la serie estacionaria oscila o se desplaza alrededor de un valor medio μ . Además, supone que el desplazamiento de μ en el tiempo presente t es ocasionado por q perturbaciones ocurridas en el pasado, ponderados por un factor θ_j , que mide la influencia de dicha perturbación en el presente de la serie, estas perturbaciones se modelan como ruido blanco con media cero y varianza constante; este modelo se denota como MA(q), donde q es su orden.
- **Modelo ARMA:** este modelo es una combinación entre los modelos autorregresivo y de promedios móviles y se denota como ARMA(p,q) donde p es el orden de los componentes autorregresivos y q es el orden de los componentes de promedios móviles (Pellenz, 2011).

En la ecuación (1) se puede ver la representación una serie de tiempo ARMA(p,q).

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} \quad (1)$$

Donde x_t es la serie de tiempo a evaluar; los factores ϕ_j son los coeficientes autorregresivos del modelo, los factores θ_i son los coeficientes de promedio móvil, y ω_t es una secuencia de ruido blanco gaussiano. De esta ecuación se puede ver que, cuando la serie de tiempo es de orden ARMA($p,0$), la serie corresponde a un modelo autorregresivo, en cambio, cuando es de orden ARMA($0,q$), corresponde a uno de promedios móviles. La ecuación (1) se puede describir como se describe en la ecuación (2).

$$\phi(B)x_t = \theta(B)\omega_t \quad (2)$$

En donde B se conoce como el operador de desplazamiento que cumple con lo establecido en la ecuación(3), $\phi(B)$ es equivalente a la expresión de la ecuación(4) y $\theta(B)$ a la expresión de la ecuación(5).

$$B^j x_t = x_{t-j} \quad (3)$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (4)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (5)$$

ARMA Integrado – ARIMA

Se explicaron previamente las principales series de tiempo dedicadas a estudiar series estacionarias; pero muchas de las series de tiempo encontradas en la práctica no son estacionarias, por lo que los modelos AR, MA y ARMA no se pueden aplicar directamente a este tipo de series de tiempo. Un método para manipular series de tiempo no estacionarias es aplicar la diferenciación, con el fin de volverlas estacionarias. La serie se diferencia d veces hasta que ésta sea totalmente estacionaria.

Una serie de tiempo de este tipo se denomina integrada, debido a que corresponde a la integral de una serie de tiempo estacionaria. La serie de tiempo ARMA integrada o ARIMA, se denota como

ARIMA(p,d,q), donde p es el orden de los componentes autorregresivo, q el orden de los componentes de promedio móvil y d el grado de diferenciación usado. Esta diferenciación se ve representada por el operador de diferenciación, usando el operador de desplazamiento B , como muestra la ecuación(6).

$$\nabla^d x_t = (1-B)^d x_t \quad (6)$$

Entonces, la serie de tiempo usando el modelo ARIMA (p,d,q) queda representada por la ecuación (7).

$$\phi \nabla^d x_t = \theta(B) \omega_t \quad (7)$$

Ahora que se tiene descrita la serie de tiempo, lo siguiente es hallar los órdenes p , q , y d . Para lo cual se utilizan las funciones de autocorrelación y autocorrelación parcial. El siguiente paso es estimar los coeficientes ϕ y θ , para lo cual existen varios métodos que permiten esto, pero los más usados son: el método de máxima probabilidad (ML), y el método de suma de cuadrados condicional (CSS).

ARIMA Estacional - SARIMA

En la práctica, muchas series de tiempo contienen un componente periódico estacional que se repite cada s observaciones. Para tratar con este comportamiento estacional, la expresión representada por la ecuación(7) debe ser modificada para que se tenga en cuenta el efecto de la estacionalidad, a este modelo modificado se le conoce como SARIMA el cual contiene una componente autorregresiva estacional de orden P , un promedio móvil de orden Q , y una diferenciación estacional de orden D , dicho modelo se representa por la ecuación(8) y su notación es SARIMA(p,d,q)×(P,D,Q).

$$\phi(B)\phi(B^s)\nabla^d\nabla^p x_t = \theta(B)\Theta(B^s)\omega_t \quad (8)$$

Para representar una serie de tiempo SARIMA, se debe, primero, escoger valores adecuados para los dos órdenes de diferenciación D y d , con el fin de eliminar la estacionariedad y estacionalidad; para después identificar los coeficientes AR y MA adecuados para cada caso.

Existen algunas otras variantes de los modelos ARIMA anteriormente descritos. Estos son aún más complejos y se utilizan para aplicaciones más específicas y suelen requerir un mayor poder de cómputo. Uno de estos es conocido como ARIMA fraccional o FARIMA y se caracteriza por presentar un orden de diferenciación fraccional y no entero.

Otra serie de tiempo bastante estudiada se conoce como Vector ARIMA o VARIMA y es una serie de tiempo multivariada, a diferencia de las anteriormente vistas que eran univariadas. Con VARIMA se tiene un vector de diferentes series de tiempo, cada una con variables diferentes, en donde una variable se ve afectada tanto por los valores pasados de sí misma, como por los valores pasados de otras variables (Deistler, 2006). El inconveniente con este tipo de series de tiempo es que son muy complejas de analizar y son bastante complejas computacionalmente hablando, por lo que no se encuentran muchos trabajos relacionados en el área de modelamiento de tráfico que maneje dicho tipo de series de tiempo (Eichler, 2006; Katarzyna y Kamiński, 2006).

Trabajos relacionados

Se han realizado muchos estudios que se dedican a analizar las propiedades estadísticas del tráfico de video, y de contenidos multimedia en general (Deljac, Z., Kunstic, M. y Spahija, B., 2011). Los primeros trabajos desarrollados se basaron en los modelos clásicos inicialmente diseñados para estudiar el tráfico telefónico, un trabajo en particular modela el tráfico variable VBR para videos MPEG, utilizando la técnica TASI ideada por Weinstein, la cual permite modelar el tráfico como un simple modelo ON-OFF (Díez y Casares., 2005), aunque este no refleja, por completo, el comportamiento del tráfico de video.

Otros trabajos han preferido usar modelos basados en cadenas de Markov (Yang y Daigle, 2005; Lázaro, Girma y Dunlop, 2004; Adas, 1996), estos, aunque muy precisos en algunos ámbitos, se quedan cortos al tratar de describir de forma precisa el tráfico de video, debido a que presentan dependencia de corto alcance (SRD), es decir, modelan los valores presentes de tráfico teniendo en cuenta uno o más

valores inmediatamente anteriores, pero como su nombre lo dice, a muy corto alcance.

Otros estudios han concluido que el tráfico de video se puede describir mejor como un proceso auto-similar (Bojkovic, Bakmaz y Bakmaz, 2009; Nilsson, Appleby y Crabtree, 2008; Ansari, Liu y Shi, 2002), el cual presenta una dependencia de largo alcance (LRD), es decir, que se basa en todos los valores pasados de una variable para modelar las características estadísticas de la misma.

La autosimilaridad de un proceso estocástico se puede identificar utilizando el parámetro de Hurst. Si el valor de este parámetro, para una determinada serie de datos, se encuentra entre 0.5 y 1; se puede considerar que dicho proceso estocástico es auto-similar. Nilsson y otros (2008) calcularon el parámetro de Hurst para diferentes secuencias de video, codificadas usando h.264, este parámetro, para el conjunto de secuencias de video, presentó un valor medio de 0.9, y un valor mínimo de 0.8, lo que revela el alto grado de autosimilaridad de este tipo de tráfico.

Varias propuestas se han realizado con el fin de modelar las características autosimilares del tráfico de video; Ansari y otros (2002) proponen una estrategia utilizando procesos autosimilares modulados de Markov y procesos autosimilares modulados estructurados; el modelo propuesto, consistente en tres procesos autosimilares, demostró ser bastante simple, y representa de manera adecuada el tráfico de video dentro del marco propuesto.

Los modelos mencionados hasta este punto son muy útiles a la hora de caracterizar el tráfico de video sobre redes de datos, gracias a esto son extensamente utilizados para la implementación de generadores de tráfico de video sintético, muy útil en los ambientes de simulación para emular el tráfico de video MPEG. En algunos trabajos (Pellenz y Souza, 2011; Reaz, Murayama, Suzuki, Yoshimoto, Kramer, Mukherjee, 2011), se dedican exclusivamente a crear este tipo de generadores sintéticos de tráfico de video mediante el modelamiento estadístico.

Hay un tipo de modelado estadístico que, además de permitir caracterizar el tráfico de video, predice los valores futuros de este con un alto grado de precisión, estos métodos se basan en series de tiempo. Las series de tiempo son modelos de tráfico con características autosimilares y de dependencia de largo alcance, y son explicadas brevemente en la segunda sección.

Los primeros trabajos en esta área utilizaron series de tiempo no estacionarias (AR, MA, ARMA), Markov y Mehrpour (2001) realizaron una de las primeras aproximaciones en el tema, utilizando diversos modelos de series de tiempo estacionarias para caracterizar el tráfico de video MPEG-1 y se utilizaron modelos TES, AR y ARMA, los resultados son pobres en los tres casos, siendo TES el que presenta mejores resultados con el inconveniente que no identifica la característica de LRD. G. Ursun, Crovella y Matta (2011) presentan un modelo ARMA simple para predecir el número de accesos a un conjunto de videos de Youtube.

Un trabajo muy detallado en modelamiento de video VBR fue realizado por Izquierdo y Reeves (1999), donde detallan el modelamiento a través de diferentes técnicas —cadenas de Markov, procesos AR, procesos TES, procesos auto-similares, y procesos i.i.d—realizando su comparación para diferente tipo de escenas y comprando el número de parámetros necesitados. Finalizan concluyendo qué tipo de modelo es adecuado para qué tipo de servicios —videoconferencia, TV, películas—. Un estudio similar fue realizado por Tanwir y Perros (2013), utilizando trazas de video h.264/AVC concluyendo que los modelos MMG y los basados en wavelets se ajustaban adecuadamente al tráfico generado por una red IPTV.

Otro estudio bastante referenciado y uno de los primeros en usar series de tiempo no estacionarias (De La Cruz, Pallarès, Alinsy Mata, 2011), propuso un modelo ARIMA fraccional o FARIMA, para modelar el tráfico VBR de video MPEG. En dicho estudio aprovechan las características LRD de FARIMA, el inconveniente de este enfoque es el gran número de coeficientes necesarios para modelar de manera precisa el tráfico de video, en la tabla 1 se presenta un resumen del número de

coeficientes versus el parámetro de Hurst que se obtuvo después de diversas pruebas.

Número de coeficientes	Parámetro de Hurst
100	0.598
500	0.657
1000	0.714
2000	0.747
3000	0.761
4000	0.767
5000	0.771
10000	0.782
25000	0.783

Tabla 1. Parámetro de Hurst vs número de coeficientes.

Fuente: elaboración propia

Al Tamimi, Jain y So-In tal vez sean los autores que más publicaciones han hecho acerca de modelamiento de tráfico de video usando series de tiempo no estacionarias, en su primera publicación en el tema (Al Tamimi, Jain y So-In, 2008), estudian las características estadísticas de diversos videos de larga duración y proponen utilizar un modelo ARIMA estacional, o SARIMA, para describir sus características. Desarrollaron varios modelos, denominados “óptimos” para cada uno de los videos analizados utilizando las técnicas descritas en la segunda sección; a pesar de que cada uno de los videos presentaba un modelo óptimo diferente, descubrieron que existe un modelo único, muy simple, que es capaz de describir, de forma muy precisa, todas las diferentes trazas de video analizadas, con un mínimo margen de diferencia respecto a los óptimos, dicho modelo simple describe una serie de tiempo SARIMA $(1,0,1) \times (1,1,1)^s$ que fue probada extensivamente.

A este modelo se le nombró SAM (Simplified Seasonal ARIMA Model), los autores lo aplicaron extensivamente con diferentes propósitos y en distintos ambientes para modelar y predecir el tráfico futuro de video, fue probado sobre WiMAX de manera exitosa (Al Tamimi, 2008), fue probado con videos en definición estándar y HD (Al Tamimi, 2012; Al Tamimi, 2010), fue comparado

con otras técnicas de modelamiento de tráfico (Al Tamimi, 2010); fue utilizado para realizar un esquema de asignación de recursos de forma dinámica (Al Tamimi, 2010) y también se utilizó sobre videos con diferentes tipos de codificación (MPEG-4, AVC, SVC) con el fin de crear un generador sintético de trazas de video. En todas estas pruebas, el modelo SAM resultó ser muy preciso en el modelamiento de tráfico de video.

Hasta este punto se ha hablado bastante de modelos de tráfico orientados a describir las propiedades estadísticas de trazas de video basadas en MPEG; pero no se ha mencionado nada acerca de técnicas orientadas a sistemas IPTV. Melia y Meo (2010) realizan mediciones sobre un ambiente IPTV real, sobre una operadora que ofrece sus servicios en Italia, de esta investigación se desprende un análisis de las características estadísticas del tráfico IPTV, describiéndolo como un proceso estocástico con una fuerte tendencia periódica semanal.

Zhiyong, Kun, Wei, Qing, Denian y Guili (2009) monitorizaron el tráfico de la red CDN de Shanghai Telecom en el transcurso de una semana, llegando a interesantes conclusiones, como que el tráfico multicast se comporta de manera estable al haber un número grande de canales de televisión en vivo; también descubrieron que el tráfico unicast conformaba una distribución Gumbel y el tráfico RSTP una distribución Gamma; confirmaron de igual forma la alta autosimilaridad presente en las ráfagas de tráfico.

Los modelos de tráfico más básicos para IPTV se han hecho utilizando enfoques markovianos (Guo, Heng Foh, Cai, Niyato y Wong, 2011; Wan; Cai y Gulliver, 2008), aunque estos no son muy precisos para modelar el complejo tráfico de IPTV, como se mencionó previamente. En cuanto a procesos autosimilares, la información es mucho más reducida y únicamente se han desarrollado algunos modelos orientados a sistemas VoD. Los más relevantes trabajan series de tiempo estacionales, Niu, Liu, Li y Zhao (2011) desarrollan un modelo SARIMA para predecir la población futura que visualizará un determinado canal sobre un gran sistema de video bajo demanda, Niu y otros (2011) llevan esta investigación un paso más allá formulando un modelo

SARIMA más preciso para modelar el ancho de banda del sistema previamente mencionado, en la figura 2 se pueden ver los resultados obtenidos, en color azul se encuentran los datos de entrenamiento, en rojo las predicciones realizadas por el mismo y en negro la traza de tráfico original.

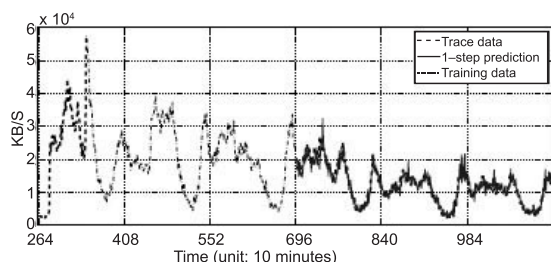


Figura 2. Predicción del ancho de banda para un canal VoD en un sistema IPTV.

Fuente: Niu (2011).

De la figura 2 se ve claramente la alta precisión de la serie de tiempo SARIMA propuesto por Niu (2011), los autores de este trabajo remarcan que el modelo no es 100% confiable como lo muestra la figura 3, donde se grafican los errores de predicción; el mayor aporte de su trabajo consiste en analizar la variabilidad de este error; a esta variabilidad la llaman volatilidad y afirman que puede ser predicha para así realizar una estimación futura del ancho de banda más precisa. Proponen un modelo usando series de tiempo GARCH —modelos generalizados auto regresivos condicionalmente heterocedásticos—, en estos, la estructura de la varianza condicional depende, además del cuadrado de los errores retrasados q periodos, de las varianzas condicionales retrasadas p periodos; de su investigación concluyen que mediante la utilización una serie SARIMA para predecir el tráfico futuro y utilizando una serie de tiempo GARCH para predecir la volatilidad sobre el error de predicción, se puede mejorar la utilización de recursos hasta tres veces.

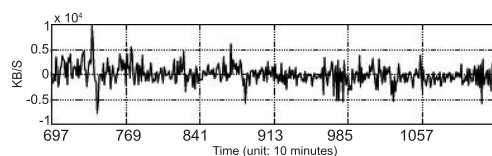


Figura 3. Errores en la predicción del modelo de la figura 2.

Fuente: Niu (2011)

Discusión

En la sección previa se realizó una revisión de los principales trabajos desarrollados en el área de modelamiento de tráfico para sistemas de transmisión de contenidos multimedia; se realizó un especial énfasis en el modelamiento a través del uso de series de tiempo de tipo ARIMA; finalmente, se mostraron algunos ejemplos de trabajos realizados sobre plataformas IPTV donde se realizó la caracterización del modelo y la predicción de sus valores futuros.

De la revisión hecha, se puede apreciar claramente que existe muy poca investigación acerca del uso de series de tiempo para realizar el modelamiento estadístico de las diferentes variables de tráfico que se pueden estudiar sobre plataformas IPTV. La mayor parte de los trabajos estudiados se enfocan en el análisis de tráfico de video, lo cual es un trabajo muy válido y que resulta de mucha ayuda para comprender el comportamiento de las trazas de video sobre las redes de datos. Pero como nuestra intención es comprender el comportamiento del tráfico generado por sistemas de transmisión de servicio IPTV, estos estudios se quedan cortos, ya que no tienen en cuenta muchos factores de vital importancia en este tipo de sistemas, como son: el manejo de parámetros de QoS, las técnicas de propagación de tráfico multicast, la convivencia con tráfico de señalización y control de las plataformas IPTV, entre otros factores.

Es cierto que se muestran algunos ejemplos de estudios del tráfico IPTV al final de la tercera sección, pero terminan siendo trabajos muy superficiales donde el mayor avance encontrado fue el modelamiento y predicción del tráfico de una plataforma IPTV orientada a proveer los servicios de video bajo demanda (Niu, 2011) usando series de tiempo SARIMA; si bien este estudio se realizó sobre una plataforma IPTV que tiene en cuenta ciertos parámetros de QoS, al prestar exclusivamente el servicio de video bajo demanda, deja por fuera el análisis del tráfico multicast.

Es por esto que este trabajo pretende ser un primer paso hacia la propuesta de un modelo que permita caracterizar y pronosticar el tráfico mediante el uso

de series de tiempo de una plataforma IPTV que se encargue de proveer servicio de televisión en tiempo real; debido a que hay muy poca investigación del modelamiento de tráfico multicast entregado en tiempo real sobre este tipo de plataformas.

Para esto, se está proyectando la construcción de un prototipo de un sistema de transmisión de servicios IPTV sobre el cual realizar la toma de datos de las diferentes variables que resultan de interés—throughput, pérdida de paquetes, jitter, retardos—. Este prototipo ha surgido como un trabajo previo desarrollado por los autores de este artículo, que ha sido soportado por Colciencias dentro de su programa de Jóvenes Investigadores y presentado ante la comunidad internacional (Villanueva y López, 2013).

Según lo investigado, el tráfico de un sistema IPTV es un tráfico con comportamiento estocástico, cuyos valores presentan una alta dependencia con sus valores pasados y que posee una fuerte tendencia periódica semanal, según Mellia y Meo (2010). Por esta razón, se considera que el modelo estadístico que podría caracterizar y predecir de forma más precisa los valores de tráfico de este tipo de sistema, es un modelo de tipo SARIMA que es muy bueno a la hora de caracterizar las tendencias periódicas de una serie de tiempo dada.

Conclusiones

En este trabajo se hace énfasis en el análisis de modelamiento de tráfico, con el fin de plantear una posible metodología que se adecúe, de manera óptima, a la hora de caracterizar y predecir el tráfico futuro en un sistema de transmisión de IPTV. Se introducen las series de tiempo como una herramienta poderosa para realizar modelos de tráfico futuro sobre redes de datos IP, entre estas se hace énfasis en las series de tiempo ARIMA. Se hace además una extensa revisión en el estado del arte acerca de modelamiento de tráfico de video. Al realizar esta investigación, no se encontró mucha información acerca de modelos de tráfico orientados a sistemas IPTV, la mayoría se enfocaba en estudiar el tráfico generado por un único video o un único canal de TV en sistemas VoD puros utilizando series de tiempo, pero no se encontraron

trabajos relacionados con modelamiento de tráfico futuro para sistemas IPTV orientados a la transmisión en tiempo real transportando la información de varios canales de televisión al mismo tiempo utilizando técnicas multicasting.

La intención de este trabajo es situar las bases y los conocimientos previos para implementar un modelo de predicción de tráfico futuro usando series de tiempo, de la investigación realizada se concluye que las series de tiempo ARIMA y en especial las SARIMA son quizá las más adecuadas para cumplir con este objetivo propuesto. La idea como trabajo futuro es implementar un prototipo de un sistema de transmisión de IPTV y, sobre este, tomar datos de tráfico con el fin de establecer un modelo que sea capaz de caracterizar y predecir el tráfico futuro de IPTV ya que actualmente no existen trabajos en el tema.

Referencias

- Adas, A.M. (1996). "Supporting Real Time VBR Video Using Dynamic Reservation Based on Linear Prediction". *Proc. INFOCOM '96, San Francisco, CA*.
- Al-Tamimi, A.K.;Jain, R. and So-In, C. (2008). "SAM: A Simplified Seasonal ARIMA Model for Mobile Video over Wireless Broadband Networks". *Tenth IEEE International Symposium on Multimedia*.
- Al-Tamimi, A.K.;Jain, R. and So-In, C. (2010). "Dynamic resource allocation based on online traffic prediction for video streams". *IEEE 4th International Conference on Internet Multimedia Services Architecture and Application*.
- Al-Tamimi, A.K.;Jain, R. and So-In, C. (2010). Modeling and Prediction of High Definition Video Traffic: A Real-World Case Study. In: *Second International Conferences on Advances in Multimedia*.
- Al-Tamimi, A.K.;Jain, R. and So-In, C. (2010). "Modeling and resource allocation for mobile video over WiMAX broadband wireless networks". *IEEE Journal on Selected Areas in Communications*.28 (3):354-365.

- Al-Tamimi, A.K. (2011). *Modeling and Dynamic Resource Allocation for High Definition and Mobile Video Streams*. (Tesis inédita de doctorado). University in St. Louis: Washington
- Al-Tamimi, A.K.; Jain, R. and So-In, C. (2012). "High-Definition Video Streams Analysis, Modeling, and Prediction". *Advances in Multimedia*. ID 539396. Doi:10.1155/2012/539396
- Ansari, N.; Hai Liu and Yun Q. Shi. (2002). "On Modeling MPEG Video Traffics". *IEEE Transactions On Broadcasting*. 48(4):1-6.
- Biernacki, A. (2010). Methods of QoS Improvement for P2P IPTV Based on Traffic Modelling. In: *Complex, Intelligent and Software Intensive Systems*.
- Bojkovic, Z.; Bakmaz, B. and Bakmaz, M. (2009). Multimedia traffic in new generation networks: requirements, control and modeling. In: Mastorakis, N.; Mladenov, V.; Bojkovic, Z., Kartalopoulos, S. and Varonides, A. (Eds.). *World Scientific and Engineering Academy and Society (WSEAS), (Proceedings of the 13th WSEAS International Conference on Communications, ICCOM)*. Wisconsin: Stevens Point, USA.
- Box, G. E. P. and Jenkins G. M. (1970). *Time series analysis forecasting and control*. San Francisco: Holden Day.
- Chatfield, C. (2000). *Time Series Forecasting*. Washington D.C.: Chapman & Hall/CRC.
- Chatfield, C. (2001). *The Analysis of Time Series, an Introduction*, Washington D.C.: Chapman & Hall/CRC.
- De La Cruz, L.J.; Pallarès, E.; Alins, J.J. and Mata, J. (1998). "Self-Similar Traffic Generation Using A Fractional Arima Model. Application to the VBR Mpeg Video Traffic". *SBT/IEEE International Telecommunications Symposium*.
- Deistler, M. (2006). Linear Models for Multivariate Time Series. In: Schelter, B., Winterhalder, M., Timmer, J. (Ed.). *Handbook of Time Series Analysis*. Berlin: John Wiley and Sons.
- Deljac, Z.; Kunstic, M. and Spahija, B. (2011). A comparison of traditional forecasting methods for short-term and long-term prediction of faults in the broadband networks. In: *MIPRO, 2011 Proceedings of the 34th International Convention*.
- Díez, J.M. and Casares, V. (2005). "Modelo de Tráfico para Video MPEG VBR Escalable y no Escalable". *IEEE Latin America Transactions*. 3(3):242-247.
- Eichler, M. (2006). Graphical Modeling of Dynamic Relationships in Multivariate Time Series. In: Schelter, B., Winterhalder, M., Timmer, J. (Ed.). *Handbook of Time Series Analysis*. Berlin: John Wiley and Sons.
- Erlang, A. K. (1909). The Theory of Probabilities and Telephone Conversations. In: *Nyt Tidsskrift for Matematik B*. Vol 20.
- G`Ursun, G.; Crovella, M. and Matta, I. (2011). Describing and Forecasting Video Access Patterns. In: *30th IEEE International Conference on Computer Communications*.
- Guo T.; Foh, C.H.; Cai, J.; Niyato, D. and Eric, W.M. (2011). Performance Evaluation of IPTV Over Wireless Home Networks. In: *Transactions on Multimedia*.
- Hamilton, D. H. (1994). *Times Series Analysis*. New Jersey: Princeton University Press.
- International Telecommunication Union-Telco. Disponible en: <http://www.itu.int/ITU-T>.
- Izquierdo, M. R. and Reeves, D. S., (1999). "A survey of statistical source models for variable-bit-rate compressed video". *Multimedia Sys*. 7 (3): 199-213.
- Janevski, T. and Vanevski, Z. (2009). Statistical Analysis of Multicast versus Instant Channel Changing Unicast IPTV Provisioning. In: *Institute of Electrical & Electronic Engineer*.
- Katarzyna, J. B. and Kamiński, M. (2006). Multivariate Signal Analysis by Parametric Models. In: Schelter, B., Winterhalder, M., Timmer, J. (Ed.). *Handbook of Time Series Analysis*. Berlin: John Wiley and Sons.

- Kwang Hee, C.; Won Mo, K.; Jong Koo, K. and Kwang Sik, K. (2011). Design and implementation of IMS service continuity between IPTV and mobile. In: *13th International Conference on Advanced Communication Technology*.
- Lazaro, O.; Girma, D. and Dunlop, J. (2004). "H.263 Video Traffic Modeling for Low Bit Rate Wireless Communication". *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*.3 (5):2124 – 2128.
- Markov, A.A. (1971). Extension of the limit theorems of probability theory to a sum of variables connected in a chain. In: *Appendix B of: R. Howard. Dynamic Probabilistic Systems*. Vol. 1. Markov Chains. John Wiley and Sons.
- Markov, P. and Mehrpour, H. (2001). Statistical Modeling of MPEG Coded Video. In: *Proceedings of the 9th IEEE International Conference on Networks (ICON '01)*.
- Mellia, M. and Meo, M. (2010). Communication Networks Measurement of IPTV traffic from an operative network. In: *European Transactions on Telecommunications*.
- Montgomery D.C.; Jennings, C.L. and Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. New Jersey: John Wiley & Sons.
- Nilsson, M.E.; Appleby, S. and Crabtree, I.B. (2008). Measuring the Hurst parameter of compressed video sequences. In: *Visual Communications and Image Processing*.
- Niu, D.; Liu, Z.; Li, B. and Zhao, S. (2011). Demand forecast and performance prediction in peer-assisted on-demand streaming systems. In: *The 30th IEEE International Conference on Computer Communications*.
- Niu, D.; Liu, Z.; Li, B. and Zhao, S. (2011). Understanding demand volatility in large VoD systems. In: *Proceedings of the 21st international workshop on Network and operating systems support for digital audio and video*.
- Park, I. K.; Seung, H. O.; Sim K. Y. and Ho Young, S. (2008). An implementation of user-participated interactive IPTV service system. In: *IEEE International Symposium on Consumer Electronics*.
- Pellenz, M.E. and Souza, R.D. (2011). "A conceptually simple framework for simulating hierarchical MPEG video traffic". *International Journal of Electronics and Communications*. 65 (4):296-304.
- Reaz, A.; Murayama, D.; Suzuki, K.; Yoshimoto, N.; Kramer G. and Mukherjee, B. (2011). Synthetic traffic generation for streaming video to model IPTV. In: *IEEE 5th International Conference on Advanced Networks and Telecommunication Systems (ANTS)*.
- Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications With R Examples*. New York: Springer.
- Tanwir, S. and Perros, H. (2013). "A Survey of VBR Video Traffic Models". *Communications Surveys & Tutorials, IEEE*. (99):1-25.
- TeleGeography. (2012). *Press Release*. Available: <http://www.telegeography.com/press/marketing-emails/2012/06/20/iptv-broadband-penetration-reaches-15-percent-growth-prospects-are-patchy/index.html>
- Tongqing, Q.; Zihui, G.; Seungjoon, L.; Jia, W.; Jun, X. and Qi, Z. (2009). Modeling user activities in a large IPTV system. In: *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*.
- Villanueva, B. F. and López, D. A. (2013). A Proposal for Implementing an IMS-based IPTV system over an IPv6/MPLS network Using Open Source Software. In: *The Seventh International Conference on Digital Society*.
- Wan, F. (2005). *Traffic Modeling and Performance Analysis for IPTV Systems*. (Tesis inédita de maestría). University of Victoria.

- Wan, F.; Cai, L. and Gulliver, T.A. (2008). A Simple, Two-Level Markovian Traffic Model for IPTV Video Sources. In: *Global Telecommunications Conference*.
- Won, Y.J.; Hong, J.W.K.; Mi-Jung C.; Chan-kyu H. and Jae-Hyoung Y. (2008). "Measurement of download and play and streaming IPTV traffic". *Communications Magazine*. 46 (10):154-161.
- Won, Y.J.; Mi-Jung, C.; Byung-Chul P.; Hong, J.W.; Hee-Won, L., Chan-Kyu, H. and Jae-Hyoung Y. (2008). End-User IPTV Traffic Measurement of Residential Broadband Access Networks. In: *Network Operations and Management Symposium Workshops*.
- Yang, S. and Daigle, J.N.(2005). A Source Model of Video Traffic Based on Full-Length VBR MPEG4 Video Traces. In: *IEEE Global Telecommunications Conference*. Vol. 2.
- Yule, G. U.(1927). On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*.
- Zhiyong, L.; Kun, W.; Wei, L.; Qing, X.; Denian, S. and Guili, H. (2009). Measurement and modeling study of IPTV CDN network. In: *IEEE International Conference on Network Infrastructure and Digital Content*.