CLASIFICACIÓN DE LOS MODELOS DE FUENTES DE VIDEO SOBRE INTERNET

R. Hernández Cuenca: lloyihc@gmail.com, rodrigo.hernandez@javeriana.edu.co
Magíster en Ingeniería Electrónica, Pontificia Universidad Javeriana

***Resumen*— El video es en la actualidad uno de los servicios más importantes y apreciados por el ser humano. A lo largo de las últimas décadas el tráfico de video ha fluido a través de distintos tipos de redes usando diversos sistemas de codificación, con lo cual han surgido distintos modelos que buscan representar adecuadamente el comportamiento de dichas fuentes. En años recientes la masificación de Internet, los esfuerzos de algunas entidades de estandarización y la evolución propia de dispositivos y redes (con significativos aumentos en ancho de banda) han fijado las nuevas bases que han hecho que los variados tipos de fuentes de video converjan y sean cada vez más compatibles con la Internet. El presente artículo describe el estado del arte de las distintas clasificaciones de los modelos de fuentes de video, para finalizar proponiendo una clasificación que se considera cubre los actuales modelos y además establece un marco de referencia para modelos de fuentes de video futuros.**

***Palabras clave*— MPEG, H.264, ISO y JPEG.**

I. INTRODUCCIÓN

En la actualidad Internet se ha convertido en el mejor aliado del video permitiéndole llegar a todos los rincones del mundo y a gran variedad de dispositivos tales como: televisores, computadores personales, dispositivos móviles, consolas de video juegos, etc.

El tráfico de video y los requerimientos resultantes para las redes han atraído gran interés de la comunidad investigativa. Gracias a que el tráfico de video es dependiente del contenido [1], de los estándares de codificación [2], de las configuraciones del codificador [3] y de su calidad [4], los modelos de fuentes de video se han desarrollado y han evolucionado a través de iniciativas académicas que en general siguen los estándares de la industria.

Es entonces cuando la comunidad investigativa comienza a desarrollar una gran variedad de modelos de fuentes de video con propósitos específicos, que con el correr del tiempo muestran una serie de fortalezas (que se tratan de preservar) y de falencias (que se tratan de eliminar).

II. LA CODIFICACIÓN DEL VIDEO

Desde 1990 muchas tecnologías de codificación de video han sido implementadas para el almacenamiento del video o su transmisión. Estas tecnologías de codificación han sido implementadas por muchos fabricantes de dispositivos de video y por diversas industrias (principalmente la del software y la del entretenimiento), las cuales han producido los diferentes formatos de video que llegan al consumidor final.

El primer concepto que se debe tener claro en la codificación de video es el de *frame*. Un *frame* es cada uno de los "cuadros" o imágenes que conforman un video. En video es común escuchar el término "fps" o *frames* por segundo, el cual hace referencia al número de imágenes por segundo que se reproducen y que dan la percepción de movimiento.

Otros dos conceptos claves son la redundancia espacial y temporal. La redundancia espacial es la información (de pixeles) que puede ser suprimida en un *frame*, sin disminuir sustancialmente su calidad (en referencia a la percepción del ojo humano). La redundancia temporal es la información que puede ser suprimida en *frames* generalmente consecutivos, por ejemplo dentro de una misma escena, también sin disminuir ostensiblemente su calidad. Los dos conceptos anteriores son, en general, aprovechados por la codificación de video. La redundancia espacial es reducida registrando las diferencias entre las partes de un *frame*, tarea conocida como compresión intra-*frame* (la cual es muy cercana a la compresión de una imagen). Por otro lado la redundancia temporal puede ser reducida registrando las diferencias entre *frames*, lo cual es conocido como compresión inter-*frame*.

En el video codificado MPEG (*Moving Picture Experts Group*) se usan tres tipos de *frames*: los I, P y B. Un *frame* I es una imagen intra codificada sin ninguna referencia a otras imágenes. Un *frame* P, es aquel que es obtenido mediante la compresión de la información diferencial entre un *frame* original y un *frame* estimado, dónde el *frame* estimado es construido a través de previos *frames* I o P. Con respecto a los *frames* B, actualmente existen dos tipos: estructura clásica y estructura jerárquica.

Un *frame* B clásico se comprime similarmente a un *frame* P, pero éste puede ser estimado solo desde el anterior *frame* I o P y desde el siguiente *frame* I o P, lo cual se puede ver en la figura 1 (a); otros *frames* B no son referenciados debido a que no es permitido por los estándares de video que precedieron al H.264. Esta restricción es levantada en el paradigma de *frame* B generalizado (primero introducido en el estándar H.264). La figura 1 (b) muestra el paradigma de *frame* B generalizado con su estructura jerárquica la cual usa *frames* B para la predicción de otros *frames* B. El caso ilustrado es la jerarquía diádica de *frames* B, significando que el número de n *frames* B entre imágenes clave (*frames* I o P) es igual a $n = 2^k - 1, k = 1,2,3,...$ (n=3 para el ejemplo ilustrado). En este ejemplo, la secuencia es $I_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2$, donde los índices representan el número de capa temporal.

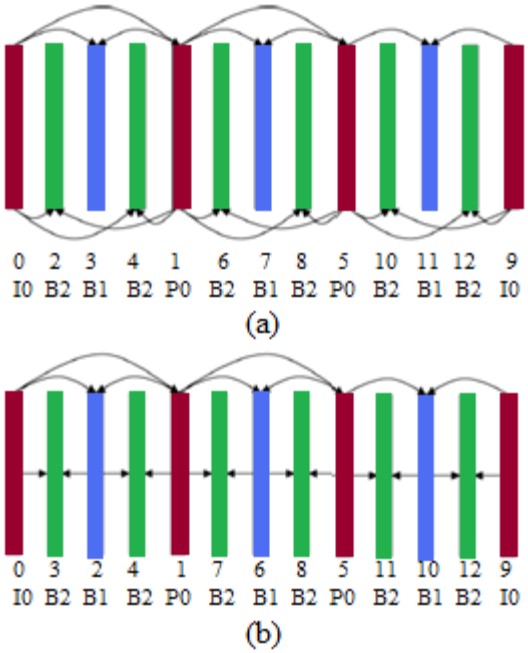

Figura 1. Estructura de predicción del *frame* B. (a) Clásica. (b) Jerárquica

En un video codificado MPEG, se organizan *frames* I, P y B en un patrón determinístico fijo llamado GOP (*Group of pictures*). Un patrón GOP en general tiene una forma (N, M) lo que quiere decir que hay N *frames* en dicho GOP y M *frames* B entre *frames* I o P, por ejemplo el GOP (12,2) sería IBBPBBPBBPBB y el GOP (12,0) se refiere a IPPPPPPPPPPP. En general un video puede ser codificado con varios GOPs, aunque en la práctica se acostumbra a codificar todo un video con un solo patrón GOP.

El tráfico de video codificado MPEG tiene características cíclicas complejas [5] de los patrones de la función de autocorrelación debido a la estructura GOP.

La dependencia de corto rango SRD (*Short Range Dependence*) y de largo rango LRD (*Long Range Dependence*) del tráfico de video es a menudo enmascarada por la función de autocorrelación distintiva a nivel de *frame*, pero puede ser vista más claramente estudiando el GOP total en vez de los *frames* individuales. En otras palabras, dado cada GOP que contiene el mismo número y orden de *frames* I, P y B, analizar los tamaños GOP puede ilustrar la presencia de SRD o LRD más claramente [6]. Otra característica muy importante, ignorada en algunos modelos, es la cros-correlación entre los diferentes tipos de *frames*.

Un concepto de desarrollado reciente es la codificación multi capa, la cual busca tener codificadores más eficientes. La codificación en capas puede ser adicionalmente clasificada como de granularidad gruesa o granularidad fina. La mayor diferencia entre la granularidad gruesa y la granularidad fina es que la primera proporciona mejoramientos en la calidad del video sólo cuando una capa completa de mejoramiento ha sido recibida, mientras la segunda mejora continuamente la calidad del video con cada palabra recibida del flujo de bits de la capa de mejoramiento. En ambos métodos de codificación (granularidad gruesa y fina), la(s) capa(s) de mejoramiento es (son) codificada(s) con el residuo entre la imagen original y la imagen reconstruida desde la capa base. Por consiguiente, la(s) capa(s) de mejoramiento tiene(n) una fuerte dependencia sobre la capa base.

En el tráfico de video MPEG se necesita considerar el hecho que la pérdida de una parte de un *frame* I afecta todos los *frames* en su GOP, pero la pérdida de un *frame* B afecta sólo a dicho *frame* (si el *frame* B es de estructura clásica) o sólo a otros *frames* B (si el *frame* B es de estructura jerárquica). Así una aparentemente pequeña tasa de pérdida de datos, en el caso de ser *frames* I puede afectar considerablemente la calidad de video percibida, siendo este no el caso de los *frames* B.

A lo largo de las dos últimas décadas se han afianzado dos importantes series de estándares de codificación de video: los ISO (*International Standars Organization*) MPEG-x y los ITU-T (*International Telecommunication Union*) H.26x. La serie MPEG-x contiene MPEG-1, MPEG-2 y MPEG-4. Por otro lado, la serie H.26x que comenzó con el H.261 en el año 1990, ha evolucionado con los H.263, H.263+ y H.26L. Adicionalmente, algunos estándares resultaron del trabajo en conjunto de estos dos grupos; debido a esto, por ejemplo el MPEG-2 es además llamado H.262.

Los estándares de codificación entonces evolucionan al ritmo de la industria con el objetivo siempre de obtener la mejor eficiencia; dicha eficiencia está directamente relacionada con menores tasas de bits y mayor calidad del video, a veces implicando un gran poder de procesamiento.

III. ITU-T H.264

La recomendación o estándar internacional H.264 (*Advanced video coding for generic audiovisual services*) es desarrollada [7] en respuesta a la necesidad de mayor compresión del video para varias aplicaciones tales como videoconferencia, almacenamiento digital de media, difusión de televisión, transmisión de video (*streaming*) en internet y video comunicaciones. Es además diseñada para habilitar el uso de la representación de video codificado de una manera flexible en una amplia variedad de ambientes de red. El uso del estándar permite al video en movimiento ser manipulado como una forma de dato computacional y ser almacenado sobre varios medios de almacenamiento, transmitido y recibido sobre las redes existentes y futuras y ser distribuido sobre los existentes y futuros canales de difusión.

La recomendación [7] está diseñada para ser genérica en el sentido de que sirva en un amplio rango de aplicaciones, tasas de bit, resoluciones, calidades y servicios. Las aplicaciones (p. ej., Windows Media Player) deben cubrir, entre otras cosas, el almacenamiento digital de la media, difusión de televisión y comunicaciones en tiempo real. En el transcurso de la creación de la especificación se consideraron varios requerimientos de aplicaciones típicas, se desarrollaron elementos de algoritmos necesarios y estos se integraron en una sintaxis sencilla con el objetivo de facilitar el intercambio de datos de video entre diferentes aplicaciones.

Sin embargo con el objetivo de ser prácticos en la implementación de la sintaxis total de la especificación, se estipula un limitado número de subconjuntos de la sintaxis por

medio de "perfiles" y "niveles". Un "perfil" es un subconjunto de la sintaxis (del flujo de bit) completa que es especificada por la recomendación [7]. Dentro de los límites impuestos por la sintaxis de un perfil dado, es aún posible que se requiera una gran variación en el rendimiento de los codificadores y decodificadores dependiendo de los valores tomados por los elementos de sintaxis en el flujo de bit, tales como el tamaño específico de las imágenes decodificadas. En muchas aplicaciones, no es óptimo ni económico implementar un decodificador capaz de manejar todos los hipotéticos usos de la sintaxis dentro de un perfil particular.

Como solución a este problema de optimización, se especifican los niveles dentro de cada perfil. Un "nivel" es un conjunto especificado de limitaciones impuestas sobre valores de elementos de sintaxis en el flujo de bit. Estas limitaciones pueden ser simples límites sobre valores. Alternativamente pueden tomar la forma de limitaciones sobre combinaciones aritméticas de valores (p. ej., el ancho de la imagen multiplicado por el alto de la imagen multiplicado por el número de imágenes decodificadas por segundo).

La especificación H.264 codificada en la sintaxis, está diseñada para permitir una alta capacidad de compresión para una calidad de imagen deseada. Con excepción de algunos modos de operación, los algoritmos son típicamente con pérdidas, es decir que los valores exactos de muestra de la fuente no son típicamente preservados a través de los procesos de codificación y decodificación. Se usan un buen número de técnicas para lograr compresiones altamente eficientes. Los algoritmos de codificación (no especificados en la recomendación H.264) deben seleccionar entre codificación Inter e Intra para regiones con forma de bloque en cada imagen. La Inter codificación usa vectores de movimiento para predicción (basada en bloques para explotar dependencias estadísticas temporales entre diferentes imágenes). La Intra codificación usa varios modos de predicción espacial para explotar las dependencias estadísticas espaciales en la señal fuente para una imagen sencilla. Los vectores de movimiento y modos de predicción intra pueden ser especificados para una variedad de tamaños de bloque en la imagen. La predicción residual es entonces adicionalmente comprimida usando una transformada para remover la correlación espacial al interior del bloque transformado antes de que sea cuantificado, produciendo un proceso irreversible que típicamente descarta la información visual menos importante mientras forma una aproximación cercana la muestra original.

En H.264 existe la codificación de video escalable, la cual permite la construcción de flujos de bits que contienen sub-flujos de bits [7].

La codificación de video Multivista permite la construcción de flujos de bits que representan múltiples vistas (p. ej., Video 3D). Similar a la codificación de video escalable, los flujos de bits que representan múltiples vistas pueden además contener sub-flujos de bits conformes a la especificación [7]. En la codificación Multivista también se tiene escalabilidad del flujo de bit temporal.

A. Codificación predictiva

Debido a los requerimientos conflictivos del acceso aleatorio y compresión de alta eficiencia, se especifican dos principales tipos de codificación. La codificación Intra es realizada sin referencia a otras imágenes; esta codificación puede proveer puntos de acceso a la secuencia codificada donde la decodificación puede empezar y continuar correctamente, pero típicamente además sólo muestra una moderada eficiencia de compresión. La codificación Inter (predictiva o bi-predictiva) es más eficiente usando la inter predicción para cada bloque de valores de la muestra desde alguna imagen decodificada previamente (seleccionada por el codificador). En contraste a algunos otros estándares de codificación de video, las imágenes codificadas usando inter predicción bi-predictiva pueden además ser usadas como referencias para la inter codificación de otras imágenes.

La aplicación de los tres tipos de codificación de imágenes en una secuencia es flexible, y el orden del proceso de decodificación no es generalmente el mismo que el orden del proceso de captura de imagen fuente en el codificador o el orden de salida del decodificador para desplegar. La elección es dejada al codificador y dependerá de los requerimientos de la aplicación.

B. Reducción de la redundancia espacial

Tanto las imágenes fuentes como los residuos de predicción tienen una alta redundancia espacial. El estándar está apoyado sobre el uso de un método de transformación basado en bloque, para remover la redundancia espacial. Después de la inter predicción de muestras decodificadas previamente en otras imágenes o predicción de base espacial de muestras decodificadas previamente dentro de la imagen actual, el resultante residuo de predicción es partido en bloques de 4x4 pixeles. Estos son convertidos al dominio transformado donde ellos son cuantificados. Después de la cuantificación muchos de los coeficientes de transformación son cero o tienen baja amplitud y pueden así ser representados con una pequeña cantidad de datos codificados. El proceso de transformación y cuantificación en el codificador no es especificado en el estándar [7].

IV. CLASIFICACIONES ANTERIORES DE LOS MODELOS DE FUENTES DE VIDEO

Debido a las ventajas que proporcionan los modelos de fuentes de video, a lo largo de los años se han ido creando una gran cantidad de modelos. Incluso diferentes autores han propuesto varias clasificaciones las cuales han ido perdiendo vigencia con el pasar de los años.

Es así como a mediados de los 90s se afirma que los modelos pueden ser divididos en tres principales clases [6]: procesos Auto-Regresivos, cadenas de Markov y modelos auto-similares o fractales.

Estaban primero los modelos Auto-Regresivos (AR), debido a que ellos son aproximaciones clásicas. Después de que el primer modelo AR fue aplicado al tráfico de video en el año 1988 [8], los procesos AR y sus variaciones siguieron siendo altamente populares. El modelo más sencillo es el Auto Regresivo de primer orden cuyo histograma es mostrado en la Figura 2 y el cual es generado por la relación recursiva $\lambda(n)= a\,\lambda(n-1)+bw(n)$ en donde w(n) es una secuencia de variables aleatorias Gaussianas y a y b son constantes. Se

asume que w(n) tiene media η y varianza 1. Adicionalmente, se asume que $|a| < 1$.

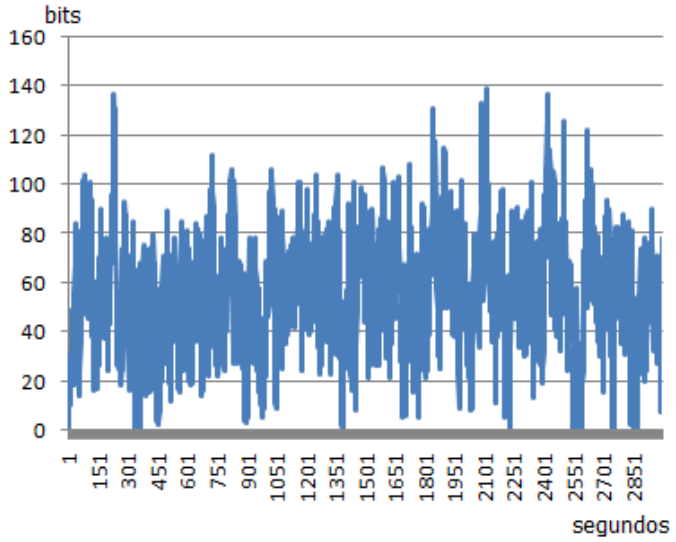

Figura 2. Histograma del modelo AR(1)

Dentro de los modelos Auto-Regresivos hay algunos que usan una combinación lineal de dos procesos AR para modelar la función de auto-correlación (ACF) del tráfico de video, en la cual un proceso AR es usado para modelar pequeños intervalos y otro para grandes intervalos. El uso de un simple proceso AR es preferido por el modelo de Krunz [9] de desviación de los tamaños de *frames* I de su media en cada escena usando un proceso AR. En [10] Liu que se basó en el trabajo de Krunz, propone el modelo *Nested* AR, el cual usa un segundo proceso AR para modelar el tamaño de *frame* medio de cada escena. En ambos casos, los cambios en escena son detectados y la longitud de la escena es modelada como una variable aleatoria con distribución geométrica. Para modelar teleconferencias (video que en general no tiene cambios de escenas), Heyman [11] propone un modelo Discreto Auto-Regresivo (DAR) y un modelo GBAR (Gamma-Beta *Auto-Regression*) [12] el cual tiene estadísticas marginales distribuidas Gamma y una auto-correlación geométrica.

Siguiendo con la primera clasificación, la segunda categoría consiste de modelos modulados Markov, los cuales emplean cadenas de Markov para crear otros procesos (p. ej., procesos de Bernoulli [13]). Rose [14] usa cadenas de Markov anidadas para modelar el tamaño del GOP. Datos sintéticos son generados a nivel GOP con lo que el modelo, de hecho, hace tosca la escala de tiempo y de esta forma no apropiada para redes de alta velocidad. Chen [15] usa un modelo AR perforado modulado doblemente Markov, en el cual un proceso de Markov anidado describe la transición entre los diferentes estados y un proceso AR describe los tamaños de *frame* de cada estado. La complejidad computacional de este método es muy alta debido a la combinación de un modelo doblemente Markov y un proceso AR. Sarkar [1] propone dos algoritmos modulados Markov basados Gamma; en cada estado de la cadena de Markov, los tamaños de los *frames* I, P y B son generados como variables aleatorias distribuidas Gamma con diferentes conjuntos de parámetros. Aunque los modelos modulados Markov pueden capturar la LRD (*Long Range Dependence*) del tráfico de video, para dichos modelos es usualmente difícil definir y segmentar con precisión las fuentes de video en estados diferentes en el dominio del tiempo (debido a la naturaleza dinámica del tráfico de video).

Finalmente en esta primera clasificación, la tercera clase consiste de los modelos auto-similares y fractales. Garrett y Willinger [16] proponen un modelo fraccional ARIMA (FARIMA – *Fractional Autoregressive Integrated Moving Average*) para replicar las propiedades LRD de secuencias comprimidas, pero no proporciona un modelo explicito para la estructura SRD (*Short Range Dependence*) del tráfico de video. Usando los resultados del modelo FARIMA, en [17] se presenta un modelo de tráfico fractal auto-similar; sin embargo este modelo no captura las variaciones múltiples en escalas de tiempo.

A finales de los años 90s surge otra clasificación en la cual los modelos pueden ser categorizados en dos clases [18, 19]: modelos SRD y modelos LRD. Estos modelos se usan para capturar dos cantidades estadísticas: distribución marginal y función de auto-correlación del tiempo de arribo del tráfico.

Un buen modelo de tráfico debe capturar las características de secuencias de video y predecir con precisión el rendimiento de la red (p. ej., probabilidades de desbordamiento de buffer y pérdida de paquetes). Entre las varias características del tráfico de video, hay dos [20] de mayor interés: 1) la distribución de los tamaños de *frames*; y 2) la función de auto-correlación (ACF) que captura dependencias comunes entre los tamaños de *frames* en video VBR (*Variable Bit Rate*). Adicionalmente, comparada a la tarea de fijar un modelo a la distribución de los tamaños de *frames*, capturar la estructura ACF del video VBR es mas desafiante debido al hecho de que los videos exhiben a la vez propiedades LRD y SRD. La coexistencia de SRD y LRD indica que la estructura ACF del tráfico de video es similar a procesos SRD en pequeñas escalas de tiempo y LRD en grandes escalas de tiempo. Así, usando solo un modelo LRD o SRD no se tienen resultados satisfactorios.

Por lo anterior se han desarrollado otros modelos muy importantes que no se encasillan en una sola categoría. Tal es el caso del modelo M/G/∞ [21] (tiempo entre arribos/ tiempo de servicio/ número de servidores) en donde M hace referencia a una distribución exponencial y G a una distribución Pareto; este modelo se compara en rendimiento a los modelos FARIMA y DAR. Los resultados de simulación indican que el modelo M/G/∞ proporciona predicciones aceptables de la perdida de datos y de tasas de error de *frame* a varias cargas de tráfico y tamaños de buffer. El modelo M/G/∞ es muy estudiado debido a su simplicidad teórica, su flexibilidad para exhibir SRD y LRD en una forma parsimoniosa y sus ventajas para estudios de simulación, tales como el bajo costo computacional.

Otros modelos que exhiben a la vez SRD y LRD son los modelos *Wavelet*, cuyo análisis esta típicamente basado en la descomposición de una señal usando una familia ortonormal de funciones base, la cual incluye una función *wavelet* pasa alto y un filtro *scaling* pasa bajo; la primera genera los coeficientes detallados, mientras que la segunda produce los coeficientes de aproximación de la señal original. La

transformada *wavelet* reduce fuertemente la correlación temporal en la señal de entrada, lo cual significa que señales con propiedades LRD producen coeficientes *wavelet* dependientes de corto rango [22].

Ya para el año 2005 surge otra clasificación con dos categorías: los DRMs (*Data Rate Models*) y los FSMs (*Frame Size Models*) [23]. Un DRM genera tasas de llegada de datos y es bueno para predecir las probabilidades de pérdida de paquetes promedio y la probabilidad de desbordamiento de *buffer*; sin embargo, no es útil para identificar *frames* afectados debido a pérdida de datos. Por otro lado un FSM genera tamaños de *frames* individuales, por lo que puede ser usado para predecir tanto la tasa de pérdida de datos como los *frames* afectados debido a la pérdida de datos. Modelos como el AR (*AutoRegressive*), DAR (*Discrete*-AR), MRP (*Markov Renewal Process*), MRP-TES (MRP *Transform-Expand-Sample*), MC (*finite-state Markov Chain*) y GBAR (*Gamma-Beta Auto-Regression*) pertenecen a los DRMs.

Con respecto a los FSMs, varios modelos han sido propuestos para la distribución del tamaño de *frame*, incluyendo las distribuciones log-normal, Gamma y varias distribuciones híbridas (p. ej., Gamma/Pareto o Gamma/log-normal) produciendo modelos como el GOP GBAR (*Gamma-Beta Auto-Regression*) el cual considera la estructura cíclica del GOP del tráfico de video [24] (lo cual lo extiende el modelo GBAR). Existen otros modelos FSMs como el MMG (*Markov-Modulated Gamma*), el GACS (*Gaussian Auto-regressive and Chi-Square*) [25] y algunos otros más, aunque en cantidad son menores que los DRMs.

V. Características actuales de los modelos de fuentes de video

Los modelos de fuentes de video juegan un rol importante en la caracterización y análisis del tráfico de red. Además de proporcionar una mirada profunda al proceso de codificación y a la estructura de las secuencias de video, los modelos pueden ser usados para muchos propósitos prácticos incluyendo asignación de recursos de red, diseño de redes eficientes para servicios de *streaming*, y la entrega de cierta garantía de QoS (calidad de servicio) al usuario final.

Uno de los elementos más importantes que ayudan en el diseño, desempeño y operación de servicios y aplicaciones de video sobre Internet es el correcto entendimiento y apropiada caracterización de la fuente. El video en general tiene diversas características, como por ejemplo la auto-similaridad [6] la cual significa que el video tiene propiedades estadísticas similares en un rango de escalas en el tiempo, lo que estadísticamente se refiere al decaimiento lento de la función de auto correlación. Por otro lado, los videos además de los codecs utilizados, tienen características propias de su contenido tales como la duración (p. ej., videos de larga duración) o la definición (p. ej., alta definición), las cuales motivan la creación de modelos específicos tales como los modelos para video de alta definición codificado con H.264/AVC [26].

En general los modelos se han ido creando de acuerdo a la evolución de la industria del video. A inicios de los 90s, el tráfico en videoconferencias era de los más importantes por lo que fueron desarrollados modelos simples como AR/DAR y algunos modelos de Markov básicos. Posteriormente la dependencia de largo rango (LRD) fue encontrada como una propiedad del tráfico de video [27] y entonces los modelos auto similares fueron desarrollados. Los modelos FARIMA y de ruido Gaussiano fractal (FGNM) fueron propuestos para capturar la LRD. Pronto al encontrarse la coexistencia de la SRD y la LRD en el video, vinieron nuevos desafíos a los modelos simples. La SRD significa que la consideración del comportamiento de tráfico en pequeñas escalas de tiempo se hace importante. Los modelos de Markov simples no podían representar bien esta característica por lo que se propusieron nuevos modelos de Markov con la desventaja de que dichos nuevos modelos necesitaban muchos estados de Markov con el objetivo de mantener cierta precisión; entonces la complejidad de la parametrización se incrementa rápidamente por el número de estados, especialmente cuando estos son usados para evaluar muchas fuentes multiplexadas. Fue entonces cuando se encontró que los multi-fractales eran naturales a través del tráfico de red debido a que ponen más atención a los comportamientos irregulares del tráfico en pequeñas escalas de tiempo. Un modelo multi-fractal típico es un proceso multiplicativo y uno de sus modelos particulares es el modelo Wavelet Multi-fractal (MWM) que fue propuesto para tráfico de video simple. Cuando se desarrollaron los modelos multiplicativos multi-fractales para fuentes de video simple, se comprueba que ellos eran buenos en capturar la SRD y la LRD teniendo tiempos de computación menores, pero su trabajo no considera la estructura de compresión del video. De hecho la estructura de compresión puede influenciar enormemente las características del tráfico en escalas de tiempo pequeñas.

Como ya se mencionó anteriormente, dentro de la familia de codecs más importante se tiene el H.264. El video H.264/MPEG-4 AVC (*Advanced Video Coding Standard*) y SVC (*Scalable Video Coding Extension*) puede ser codificado:

- Con escalas de cuantización fijas, lo cual resulta en video de una calidad casi constante a expensas de una gran variabilidad de la tasa de bits.

- Con tasa controlada, la cual adapta las escalas de cuantización manteniendo la tasa bit casi constante a expensas de la variabilidad de la calidad del video.

Codecs como los H.264 SVC típicamente comprimen el video por debajo de la tasa de bits promedio de otros. Por otro lado, dichos codecs producen una mucha mayor variabilidad del tráfico. El coeficiente de variación (desviación estándar normalizada con la media) del tamaño de los *frames* alcanza niveles por encima de 2.5, mientras dicho coeficiente está en el rango de 0.7 a 1.4 en MPEG-4 parte 2.

Otra característica del video, es que puede ser codificado por capas. Las técnicas de codificación por capas usan una estructura jerárquica. Así mismo en el video, existe la codificación de descripción múltiple (MDC) la cual es un esquema de codificación no jerárquica que genera capas de igual importancia [20]. En las secuencias codificadas MDC, cada capa puede proporcionar calidad aceptable y varias capas juntas llevan a calidades más altas; cada capa puede ser

individualmente codificada con otras técnicas de codificación de capa.

Hay que tener en cuenta que no solo existen las técnicas de codificación de video MPEG-x y H.26x. Existen otras varias técnicas tanto estandarizadas (p. ej., *Motion* JPEG 2000) como propietarias (p. ej., RealVideo), y dentro de estas algunas ya obsoletas (p. ej., Cinepak); Las técnicas *Motion* JPEG tienen la característica de no hacer, en principio, codificación temporal (inter *frame*) y sólo hacerla a nivel intra *frame*, por lo cual no son muy eficientes a nivel de tasa de bits, pero tienen la ventaja de imponer bajos requerimientos de memoria y procesamiento a los dispositivos, lo cual las hizo bastantes útiles en las primeras cámaras digitales y en algunos modelos de cámaras IP; versiones posteriores como el video JPEG 2000, que usan la transformada Wavelet, probaron ser comparables [28] a MPEG-4 (que está basado en la *Discrete Cosine Transform* DCT).

Entonces dentro de las características de los modelos de fuentes de video, actualmente se estudian: las estadísticas de los tamaños de *frames*, las estadísticas de los tamaños GOP (*Group Of Picture*), la calidad de *frames* y GOP, las correlaciones entre los tamaños (tanto de *frames* como de GOPs), el rendimiento de la bit *rate-distortion* RD (es decir la calidad del video – PSNR – como función de la tasa de bit promedio), la bit *rate variability-distortion* VD (es decir la variabilidad de la tasa de bit como una función de la calidad del video o distorsión [29]) y otras más que dependerán del servicio o tipo de tráfico de video en particular.

VI. Clasificación propuesta

Como se describió en la sección IV en las últimas décadas se han propuesto distintas clasificaciones para los modelos de fuentes de video. A continuación se realizará un análisis con el que se evidencian las razones que hacen que dichas clasificaciones sean inadecuadas y que servirá como preámbulo a la nueva clasificación propuesta en el presente artículo.

A. Análisis de clasificaciones previas

Una de las primeras clasificaciones dividió a los modelos en: procesos Auto-Regresivos, modulados Markov y procesos Auto similares o fractales. Esta clasificación, que se orienta a la forma de generación de procesos y variables, le resta importancia a las características propias del video, lo cual la hace muy rígida y además obsoleta cuando empiezan a aparecer modelos de fuentes de video con otras aproximaciones, como por ejemplo los modelos Wavelet 3DEZBC, SRP-MM (*Spatial Renewal Process - Multinomial Method*) [5] o el M/G/∞.

Posteriormente surgió una clasificación que diferenciaba a los modelos en SRDs (dependencia de corto rango) y LRDs (dependencia de largo rango). Esta clasificación a pesar de tener en cuenta dos muy importantes características de las fuentes de video, se quedó rápidamente corta cuando otras características del video empezaron a tomar fuerza como la estructura de la compresión, la correlación inter e intra GOP o la correlación cros capa.

Luego aparece una clasificación que divide a los modelos en los de tasas de datos (DRM) y en los modelos de tamaños de *frames* (FSM). Esta clasificación, aunque sigue siendo muy importante, queda desactualizada debido a que surgen nuevos modelos que buscan ser válidos en otras "facetas" del video. Cuando los codecs de video empiezan a trabajar con el concepto de codificación multi-capa, empieza a verse la necesidad de modelos de fuentes de video multi-capa que sean útiles para capturar con precisión las propiedades de tráfico de las capas de mejoramiento (o capas agregadas) las cuales tienen características propias, como por ejemplo no presentar *frames* I [23]. Por otro lado aparecen modelos de fuentes de video que buscan ser válidos para ciertas características o servicios particulares, como por ejemplo el video de larga duración [30] o para la televisión IP [31].

B. Nueva propuesta

Después de analizar una gran cantidad de modelos de fuentes de video, se encuentra que los modelos DRMs y FSMs siguen siendo importantes, aunque no cubren todas las formas de video y por otro lado carecen de precisión (o validez) en muchos escenarios. Por lo anterior se crean dos familias más, con lo cual se busca abarcar la totalidad del espectro de los modelos. Entonces la clasificación sugerida se compone de:

- Modelos de Tasas de datos (DRM).
- Modelos de tamaños de *frames* (FSM).
- Modelos multi capa.
- Modelos orientados al contenido.

Cada una de las anteriores familias no es excluyente; es decir que un modelo pertenezca a una familia no implica que dicho modelo no pueda también pertenecer a otra. Que un modelo sea clasificado en una u otra familia dependerá de muchos factores, como el objetivo principal con el que fue creado o de la forma misma en cómo se intentan encontrar los resultados.

Esta coexistencia es sana y además habitual en los modelos de fuentes de video. Por ejemplo el modelo Auto Regresivo Perforado [15] usa procesos auto regresivos y también es modulado Markov (en referencia a la histórica primera clasificación) y a la vez captura bien las propiedades SRD y LRD del tráfico de video (en referencia a la histórica segunda clasificación).

La figura 3 ilustra la relación entre la precisión y la complejidad de las cuatro familias propuestas.

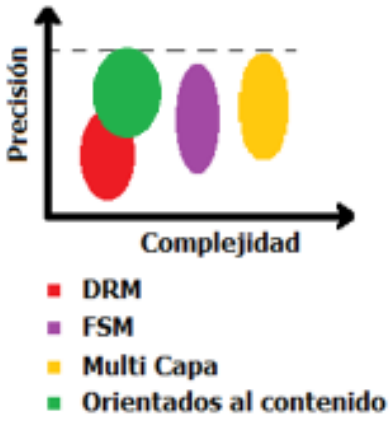

Figura 3. Relación entre precisión y complejidad

En las siguientes sub secciones se describen las características propias de cada una de las familias de modelos, las cuales permiten una mayor comprensión de la taxonomía propuesta.

Modelos de Tasas de datos (DRM)

Como se describió en la sección IV, los DRMs generan tasas de llegada de datos (análisis a nivel de bits) y son usados para encontrar características como:

- La tasa o el radio de error de bit.
- La tasa o probabilidad de pérdida de paquetes (PLR).
- La probabilidad de desbordamiento de buffer.
- El radio de error en ráfagas de bit.
- El rendimiento de la tasa de bit distorsión (RD), es decir la calidad del video como función de la tasa de bit promedio.
- La tasa de bit variabilidad-distorsión (VD), es decir la variabilidad de la tasa de bit como una función de la calidad del video o distorsión.

Estos modelos son los más numerosos y antiguos, en vista de que buscan caracterizar las fuentes de forma "clásica", aunque a veces descuiden características más complejas del video. Por lo anterior, los modelos de tasas de datos se hacen inválidos (no precisos) en muchos escenarios modernos.

Modelos de Tamaños de *frames* (FSM)

Como se describió en la sección IV, los FSMs buscan modelar las características propias de los distintos tipos de *frames* y sus correlaciones (análisis a nivel de *frames* individuales y sus inter relaciones). Estos modelos se enfocan en características como:

- La distribución de los tamaños de los *frames*.
- La tasa de error de *frame* (FER).
- La tasa de pérdida de datos y las estadísticas de los *frames* afectados debido a la pérdida de datos.
- La correlación del tamaño de *frames*.
- Los tamaños de los GOP.
- La correlación intra GOP.
- La correlación inter GOP.

Estos modelos, al enfocarse en muchas más características del video (no cubiertas por los DRMs), tienden a tener una alta complejidad tanto matemática como de los algoritmos de simulación, lo que hace que para su formulación y validación sean necesarios muchos recursos, lo cual hace que estos modelos no sean numerosos.

Modelos Multi Capa

Se propone la familia de modelos multi capa debido a que la evolución de la codificación hacia escenarios por capas, hace en consecuencia que las fuentes de video no sean modeladas de forma precisa ni por los modelos DRM ni por los modelos FSM.

Los modelos multicapa pueden ser para un número de capas en particular, como el modelo de Chandra [32] que usa una cadena de Markov de estado finito para modelar una y dos capas de tráfico de video de todos los niveles de actividad; en este modelo, enfocado a video H.261 y MPEG-2, entran nuevas características como la tasa CBR (*Constant Bit Rate*) de la capa base o la tasa VBR (*Variable Bit Rate*) máxima. El modelo [32] tiene la limitación de que sólo es válido para secuencias de video con capa base CBR.

Los modelos multicapa pueden ser propuestos también para un número de capas no específico. En [33] se describe un modelo para video MPEG-4 en donde se tienen en cuenta características como: los tamaños de los *frames* (en bits) tanto de la capa base como de las capas de mejoramiento, además la granularidad gruesa y fina, encontrando la existencia de una fuerte correlación entre la capa base y la capa de mejoramiento, por lo que también se encuentran características como los coeficientes de *cross* correlación (ccf); este modelo tiene la limitación [33] de ser válido en fuentes de video sin o con muy pocos cambios de escenas.

En modelos más avanzados como el de Dai y Loguinov [22] se encuentran características como:

- Las correlaciones inter e intra GOP.
- El radio de pérdida para cada tipo de *frame*.
- La correlación cross capa
- El radio de pérdidas de datos desbordados de las capas de mejoramiento a diferentes tasas de drenaje (D) y para distintas capacidades (c) del buffer.

Los modelos multicapa a medida que evolucionan tienden a ser más complejos, debido a las características propias de la codificación; recordemos a la codificación de descripción múltiple (MDC), en donde cada capa puede ser individualmente codificada con otras técnicas de codificación.

Al estar relacionados los modelos de fuentes video con la codificación, entonces se hace importante el conocimiento (por parte de los investigadores) de las características avanzadas de la codificación [34], como es el caso de H.264 SVC incluyendo la escalabilidad temporal (frecuencia de *frame*), escalabilidad espacial y escalabilidad de calidad SNR (*signal to noise ratio*) en múltiples capas; la escalabilidad de calidad sub-capa proporcionada por la escalabilidad de granularidad mediana (MGS), parte una capa dada (de una calidad escalable) en varias capas (sub capas) MGS.

Entonces los modelos multicapa son creados en general para ser válidos en una amplia gama de escenarios, aunque posteriormente (conforme la codificación evoluciona) prueben serlo en menor medida.

Modelos orientados al contenido

Los modelos orientados al contenido incluyen todos aquellos modelos de fuentes de video que fueron creados con el objetivo de modelar un servicio o tipo de tráfico de video en particular. Estos modelos no buscan modelar el espectro total de las fuentes sino que se especializan en características puntuales lo que los hace generalmente sencillos, tanto a nivel del modelo matemático como el de simulación, y muy precisos.

Dentro de la familia de modelos orientados al contenido, se hace la siguiente sub división:

a) **Modelos para video en formato corto**: son los que se especializan en video generado por el usuario y en video clips de corta duración (generalmente menor a 7 minutos [35]); este video tiende a ser masivo (p. ej., redes sociales o videos musicales).
b) **Modelos para video en formato largo de definición estándar (SD)**: son los que se especializan en video de larga duración (generalmente mayor a 7 minutos) y en resoluciones alrededor de 640×480 pixeles (cuando se tiene radio de aspecto 4:3); estos modelos buscan ser válidos a cargas de tráfico medianas.
c) **Modelos para video en formato largo de alta definición (HD)**: son los que se especializan en video de resoluciones iguales o mayores a 1280×720 pixeles; estas fuentes producen en general tasas de bit altas (mayores a 1 Mbps llegando a valores tan altos como 12 Mbps).
d) **Modelos para video de 3 dimensiones (3D)**: son los que se especializan en video multi vista; estas fuentes en general tienen tasas de bit iguales o mayores a las del video HD.
e) **Modelos para IPTV**: son los que se especializan en fuentes para servicios como *Live* TV (televisión peer-to-peer y transmisión de televisión en vivo sobre Internet), VoD (*Video-on-Demand*) e Internet PVR (televisión en vivo que es almacenada para ser visualizada posteriormente); este tipo de video se transporta sobre redes de operador (en su mayor parte) y a veces sobre Internet abierta.
f) **Modelos para video ambiente**: son los modelos para video que es almacenado por cámaras (de seguridad, *nannycams, petcams*, etc.) y otras formas de video persistente.
g) **Modelos para video comunicaciones**: son los modelos para video llamadas (basadas en PC), video conferencias, visualización de webcam y video monitoreo basado en la web. Estos modelos tienen la característica de ser totalmente en tiempo real.
h) **Modelos para video de juegos**: son los que se especializan en fuentes de video para juegos online, juegos de consolas en red y juegos de mundo virtual multi jugador.
i) **Modelos para video móvil**: son los que se especializan en video transmitido (o recibido) desde dispositivos móviles. Este video generalmente posee baja o mediana calidad.
j) **Modelos para video en archivos compartidos**: son los que se especializan en video *peer-to-peer* incluyendo el de los sistemas reconocidos (como BitTorrent, eDonkey, etc.) y el de los sistemas basados en la web. Estos modelos tienen la característica de que una fuente puede estar compartida en muchas "ubicaciones".

En vista de la gran complejidad que presentan los FSMs y de la creciente complejidad que presentan los modelos Multi capa (cada vez mayor, debido a la evolución propia de la codificación por capas), se hace lógico pensar que aquellos modelos que se enfoquen en un tipo de tráfico particular (los modelos orientados al contenido) y que por tanto permitan hacer muchas aproximaciones (en sus formulaciones) se convertirán en los de mayor uso a futuro.

REFERENCIAS

[1] U. Sarkar, S. Ramakrishnan, D. Sarkar, "Modeling Full-Length Video Using Markov-Modulated Gamma-Based Framework". IEEE/ACM Transactions on networking, Vol. 11, No. 4, pp. 638-649, agosto de 2003.
[2] J. Jiang, Y. Weng, "Video Extraction for Fast Content Access to MPEG Compressed Videos". IEEE Transactions on circuits and systems for video technology, VOL. 14, NO. 5, pp. 595-605, mayo de 2004.
[3] F. Zhai, C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, A. K. Katsaggelos, "Joint Source Coding and Packet Classification for Real-Time Video Transmission Over Differentiated Services Networks". IEEE Transactions on multimedia, VOL. 7, NO. 4, pp. 716-726, agosto de 2005.
[4] E. P. Ong, M. H. Loke, W. Lin, Z. Lu, S. Yao, "Video Quality Metrics – An Analysis for Low Bit Rate Videos", IEEE, 2007.
[5] C.H. Liew, C.K. Kodikara, A.M. Kondoz, "MPEG-encoded variable bit-rate traffic modelling". IEE Proc.-Commun, VOL. 152, NO. 5, pp. 749-756, octubre de 2005.
[6] Yellanki Lakshmi, "Performance Evaluation of VBR Video Models", Master Thesis, University of Saskatchewam, Canadá, 1999, http://www.cs.usask.ca/homepages/faculty/carey/projects/telesim.html
[7] ITU-T, "Recommendation ITU-T H.264 Advanced video coding for generic audiovisual services", ITU-T H.264 edición 5.0, marzo de 2010, http://www.itu.int/rec/T-REC-H.264-201003-I/en
[8] B. Maglaris, D. Anastasiou, P. Sen, G. Karlsson, y J. Robbins, "Performance Models of Statistical Multiplexing in Packet Video Communications", IEEE Transactions on Communications, Vol. 36, julio de 1988.
[9] M. Krunz, S. K. Tripathi, "On the Characterization of VBR MPEG Streams", Pmc. of ACM SIGMETRICS, vol. 35, junio de 1997.
[10] D. Liu, E. I. Sara, W. Sun, "Nested Auto-Regressive Processes for MPEG-Encoded Video Traffic Modeling", IEEE Trms. on CSW, vol. 1, febrero de 2001.
[11] D. F. Heyman. A. Tabatabai, T. V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM network", IEEE, vol. 2, marzo de 1992.
[12] D. P. Heyman, "The GBAR Source Model for VBR Video Conferences", IEEE Trans. on Networking. vol. 5, agosto de 1997.
[13] A. Lombardo, G. Morabito, G. Schembra, "An Accurate and Treatable Markov Model of MPEG-Video Traffic", INFOCOM, marzo de 1998.
[14] O. Rose, "Simple and Efficient Models for Variable Bit Rate Mpeg Video Traffic" Performance Evaluation, vol. 30, 1997.
[15] T. P.-C. Chen, T. Chen, "Markov Modulated Punctured Autoregressive Processes for Video Traffic and Wireless Channel Modeling", IEEE, abril de 2002.
[16] M. W. Garrett, W. Willinger, "Analysis, Modeling and Generation of Self-similar VBR Video Traffic", Proc. of ACM SIGCOMM, agosto de 1994.
[17] C. Huang, M. Devetsikiotis. I. Lambadaris, A. R. Kaye, "Modeling and Simulation of Self-Similar Variable Bit Rate Compressed Video: A Unified Approach", Pmc. of ACM SIGCOMM, agosto de 1995.
[18] H. Liu, N. Ansari, Y. Q. Shi, "Modeling VBR video traffic by Markov-Modulated self-similar processes", New Jersey Institute of Technology, IEEE, pp. 363-368, 1999.
[19] Z. Avramova, D. De Vleeschauwer, K. Laevens, S.Wittevrongel, H. Bruneel, "Modelling H.264/AVC VBR video traffic: comparison of a Markov and a self- similar source model", Springer Science+Business Media, pp. 91-102, junio de 2008.
[20] M. Dai, Y. Zhang, D. Loguinov, "A Unified Traffic Model for MPEG-4 and H.264 Video Traces", IEEE Transactions on multimedia, VOL. 11, NO. 5, pp. 1010-1023, agosto de 2009.

[21] M. Krunz, A. M. Makowski. “Modeling Video Traffic Using M/G/∞ Input Processes: A Compromise between Markovian and LRD Models” IEEE, Vol. 16 No 5, pp. 733-748, junio de 1998.
[22] M. Dai, D. Loguinov. "Analysis and Modeling of MPEG-4 and H.264 Multi-Layer Video Traffic", IEEE, 2005.
[23] W. Zhou, D. Sarkar, S. Ramakrishnan, "Traffic Models for MPEG-4 Spatial Scalable Video", IEEE GLOBECOM 2005, pp. 256-260, 2005.
[24] M. Frey, S. Nguyen-Quang, “A Gamma-Based Framework for Modeling Variable-Rate MPEG Video Sources: the GOP GBAR Model”, IEEE Transactions on networking, vol. 8, diciembre de 2000.
[25] A. Alheraish, S. A. Alshebeili, T. Alamri, "A GACS Modeling Approach for MPEG Broadcast Video", IEEE Transactions on broadcasting, VOL. 50, NO. 2, junio de 2004.
[26] Abdel Karim Al Tamimi, Raj Jain, Chakchai So-In, "Modeling and Prediction of High Definition Video Traffic: A Real-World Case Study", Washington University in St. Louis, IEEE 2010 Second International Conferences on Advances in Multimedia, pp. 168-173, 2010.
[27] X. Huang, Y. Zhou, R. Zhang, "A Multiscale Model for MPEG-4 Varied Bit Rate Video Traffic", IEEE Transactions on broadcasting, Vol. 50, No. 3, pp. 323-334, septiembre de 2004.
[28] B. Kulapala, P. Seeling, M. Reisslein, "Comparison of Traffic and Quality Characteristics of Rate-Controlled Wavelet and DCT Video", IEEE, Arizona State University, 2004.

[29] G. Van der Auwera, P. T. David, M. Reisslein, "Traffic and Quality Characterization of Single-Layer Video Streams Encoded with the H.264/MPEG-4 Advanced Video Coding Standard and Scalable Video Coding Extension", IEEE Transactions on broadcasting, VOL. 54, NO. 3, septiembre de 2008.
[30] B. Melamed, D. E. Pendarakis, "Modeling Full-Length VBR Video Using Markov-Renewal-Modulated TES Models", IEEE journal on selected areas in communications, VOL. 16, NO. 5, junio de 1998.
[31] F. Wan, L. Cai, T. A. Gulliver, "A Simple, Two-Level Markovian Traffic Model for IPTV Video Sources", Department of Electrical and Computer Engineering, University of Victoria, 2008.
[32] K. Chandra, A. R. Reibman, "Modeling One- and Two-Layer Variable Bit Rate Video", IEEE/ACM Trans. Networking, vol. 7, junio de 1999.
[33] J. A. Zhao, B. Li, I. Ahmad. “Traffic Model For Layered Video: An Approach on Markovian Arrival Process”, Packet Video, abril de 2003.
[34] P. Seeling, M. Reisslein, "Video Transport Evaluation With H.264 Video Traces", School of Electrical, Computer, and Energy Engineering, Arizona State University, 2011, disponible online http://trace.eas.asu.edu/publications/H264VidTraceTut.pdf
[35] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2009–2014", White Paper, junio 2 de 2010, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.html