

Modelo de tráfico P2P basado en transacciones

F. J. Ramón Salguero, G. García de Blas, M. L. García Osma,
A. Maeso Martín-Carnerero, J. Enríquez Gabeiras

Nuevas tecnologías de red, Telefónica I+D, Madrid, España
E-mail: {fjrs, ggdb}@tid.es, maria.garciaosma@telefonica.es, {ammc, jeg}@tid.es

Resumen— En la actualidad, el intercambio de archivos P2P no es sólo uno de los componentes más importantes del mix de tráfico de Internet, sino que también se ha convertido en una forma común de distribuir comercialmente contenidos a través de las redes IP. Se hace necesario, por tanto, llegar a un nivel de comprensión del comportamiento de estas aplicaciones similar al alcanzado para otros servicios de Internet, como la navegación web. El artículo contribuye a este fin, proporcionando un modelo analítico simple que permite el estudio de los sistemas P2P de intercambio de ficheros en términos de experiencia de usuario e impacto de su tráfico en las infraestructuras de red. Además, se presentan resultados de simulaciones que demuestran la precisión del modelo analítico propuesto y se aportan resultados prácticos que permiten comprender mejor el comportamiento cualitativo de los sistemas P2P, así como un conjunto de reglas de diseño de interés tanto para desarrolladores de aplicaciones P2P como para ingenieros de tráfico.

I. INTRODUCCIÓN

Durante los últimos años, el interés por el P2P ha crecido al mismo ritmo que lo hacía este tráfico en Internet. Numerosos estudios basados en medidas ([9], [1]) muestran que el tráfico P2P representa entre un 60% y un 80% del total del tráfico transportado sobre redes IP. En consecuencia, se hace necesario alcanzar un nivel de comprensión más profundo de las características de este tráfico, y sus implicaciones en el funcionamiento de las redes.

Recientemente, el interés en la caracterización de las aplicaciones P2P ha resurgido con fuerza, pues el P2P está convirtiéndose en un modo común y eficiente de distribuir contenidos comercialmente. Es el caso de servicios de tipo VoD, como Joost [2] y, más recientemente, de plataformas de P2P *Streaming*, como SopCast [3].

La mayoría de los estudios anteriores sobre las aplicaciones P2P se centran en determinar sus características y funcionalidades, ya sea explicando el funcionamiento de los protocolos específicos de una determinada aplicación ([4]), caracterizando el tráfico generado ([5], [6]), o bien identificando tráfico P2P “escondido” ([1]). Al mismo tiempo, se ha dedicado un esfuerzo significativo a la producción de modelos analíticos que describan sistemas P2P concretos [8] o al diseño de herramientas para simular el comportamiento de estas aplicaciones ([9]).

Nuestros estudios previos ([9], [11] y [12]) han seguido estas tendencias hasta la fecha, investigando mediante simulaciones el impacto de parámetros relevantes tanto en el comportamiento de las redes de intercambio P2P como en el tráfico de interconexión. Este artículo va un paso más allá en esta dirección introduciendo un modelo analítico que describe la dinámica general de cualquier sistema de intercambio P2P de ficheros.

II. ESTADO DEL ARTE Y CONSIDERACIONES SOBRE EL MODELO

Modelar el tráfico P2P usando únicamente una tasa de tráfico por usuario predefinida, aunque es conceptualmente simple, presenta inconvenientes importantes en la práctica. En este tipo de modelado, el volumen de tráfico P2P queda fijado de antemano y no se tiene en cuenta el comportamiento dinámico del sistema. Por tanto, este tipo de modelos son inadecuados para determinar el impacto en el sistema P2P de cualquier cambio en las condiciones de contorno.

Por otra parte, se han propuesto numerosos modelos analíticos que describen y estudian detalladamente el comportamiento de determinadas aplicaciones P2P de intercambio de ficheros ([10],[11]). Sin embargo, para utilizar este tipo de modelos se requiere el conocimiento preciso de numerosos parámetros de entrada ligados a una aplicación concreta y que, normalmente, son difíciles de cuantificar.

Así, sería deseable un modelo analítico más simple —aunque suficientemente preciso— que permita describir apropiadamente el comportamiento de los usuarios en sistemas P2P de gran tamaño.

Partiendo de premisas comúnmente aceptadas en la literatura [5], trabajos y resultados previos ([9],[11],[12]) y medidas en la red eMule, este artículo proporciona un modelo analítico general para sistemas P2P que es significativamente más sencillo que los existentes, aunque suficientemente preciso.

Para este modelo, se ha tomado como elemento clave el propio comportamiento del usuario como consumidor del servicio, y el impacto que la QoE (*Quality of Experience*) —caracterizada por el tiempo medio de descarga— tiene en su demanda de contenidos.

III. MODELO ANALÍTICO BASADO EN TIEMPOS DE SILENCIO

A. Modelo de petición de contenidos basado en transacciones y tiempos de consumo.

El número de peticiones P2P por usuario no puede considerarse un parámetro exógeno [6]. Al contrario, este parámetro está relacionado con el tiempo medio de descarga, pues, cuanto más rápido se descargan los contenidos, más motivación tienen los usuarios para hacer nuevas peticiones al sistema. Modelar este comportamiento es especialmente interesante pues permite determinar si, por ejemplo, existe una demanda latente de tráfico P2P que puede aflorar cuando cambien las condiciones de contorno. Para modelar el comportamiento dinámico asociado a este fenómeno, se introduce a continuación un modelo de demanda basado en transacciones y tiempos de consumo de contenidos.

El usuario medio necesita un cierto tiempo para consumir un contenido descargado y, a largo plazo, este usuario no realizará una nueva petición hasta que, por lo menos, un contenido previamente descargado se haya consumido. Esta es una condición básica de estabilidad ya que, de otra forma, los contenidos permanecerían almacenados y compartidos *ad infinitum* en el PC del usuario, lo que no es una suposición realista.

Se asume también que el usuario, que es capaz de descargar numerosos contenidos simultáneamente, no excederá un número máximo de descargas pendientes simultáneas. Por consiguiente, a largo plazo, el usuario tenderá a “reemplazar” las peticiones pendientes por otras nuevas, una vez se descarguen y consuman esos contenidos.

Tenemos así un modelo basado en transacciones como el que se muestra en la siguiente figura.

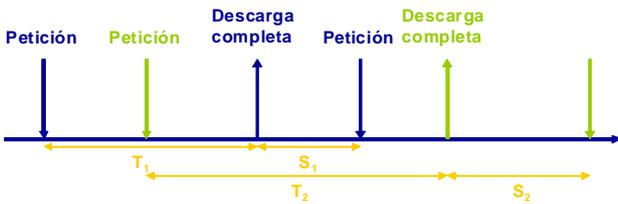


Fig. 1. Modelo de demanda P2P basado en transacciones

Podemos identificar tres parámetros en este modelo:

- T_i : Tiempo total de descarga completa del contenido i . Su media se representará como T .
- S_i : Tiempo de silencio o de consumo. Este parámetro se puede modelar por una variable aleatoria de media S .
- N_{pet} : Número medio de peticiones activas por usuario.

Este modelo basado en transacciones será nuestro punto de inicio para modelar la dinámica de los sistemas P2P.

B. Modelo general simplificado

Sin tener en cuenta los protocolos y sistemas subyacentes, una red P2P puede verse como un sistema complejo que recibe peticiones de los usuarios a una determinada tasa, λ , y las sirve con una tasa global μ , que, en la práctica, depende principalmente de tres factores: el ancho de banda de subida disponible por *peer*, el número de *peers* en el sistema y la eficiencia del protocolo P2P. Así, cada *peer* puede verse como un contribuyente a la tasa global de peticiones (λ) y como un servidor con una tasa igual a la capacidad de su *uplink*.

C. Modelo de tráfico P2P basado en transacciones

Si asumimos que la mayoría de usuarios en un sistema P2P tienen la capacidad de su *uplink* saturada (la situación más común, pues la mayoría de los usuarios tienen conexiones con capacidad de bajada superior [11]), podemos expresar la tasa global de servicio de archivos (μ) como la capacidad global de todos los *peers* de la red para servir archivos. Por tanto, teniendo en cuenta el modelo de petición de contenidos basado en transacciones y tiempos de silencio, podemos expresar la tasa global de peticiones de contenidos (λ) y la tasa de servicio (μ) como:

$$\lambda = \frac{N_{\text{usuarios}} \cdot N_{\text{pet}}}{T + S} \quad (1)$$

$$\mu = \frac{N_{\text{usuarios}} \cdot BW_{\text{subida, usuario}}}{F} \quad (2)$$

donde:

- N_{usuarios} : número medio de usuarios en el sistema P2P durante el intervalo de observación.
- N_{pet} : número medio de peticiones (descargas pendientes paralelas) por usuario.
- T : tiempo medio de descarga.
- S : tiempo medio de silencio (o *consumo*).
- $BW_{\text{subida, usuario}}$: ancho de banda de subida medio por usuario.
- F : tamaño de contenido medio (bits).

$$\text{Y, dado que } \lambda \approx \mu, \text{ i.e., (1) } \approx (2), \text{ tenemos que } \lambda = \frac{N_{\text{usuarios}} \cdot N_{\text{pet}}}{T + S} \approx \frac{N_{\text{usuarios}} \cdot BW_{\text{subida, usuario}}}{F} = \mu \quad (3)$$

Después de simplificar los parámetros que están presentes en ambos lados de la ecuación y reordenar, obtenemos una ecuación simple para describir el comportamiento del tráfico P2P:

$$T = \frac{N_{\text{pet}} \cdot F}{BW_{\text{subida, usuario}}} - S \quad (4)$$

IV. CONSECUENCIAS PRÁCTICAS DEL MODELO ANALÍTICO PROPUESTO

Debe señalarse que:

- La expresión anterior no depende del número de usuarios en el sistema, lo cual es coherente con el modelo. Supongamos que el sistema está en un estado estacionario como el descrito por la ecuación (4) y se añade un nuevo usuario al sistema. Este usuario introduciría $\frac{N_{\text{pet}}}{T + S}$ peticiones adicionales al sistema (en media) pero también incrementaría la capacidad global de servirlos en una cantidad de $\frac{BW_{\text{subida, usuario}}}{F}$. Por tanto, la expresión global quedaría inalterada.
- Las variables de la ecuación pueden ser fácilmente agrupadas para identificar qué agentes pueden controlarlas:
 - **Tiempo medio de descarga (T)**. Cuanto más bajo sea este parámetro, mejor será la QoE.
 - **Los usuarios** son capaces de modificar el punto de equilibrio mediante el **tiempo medio de silencio (S)**, el **número medio de peticiones simultáneas (N_{pet})** y el **tamaño medio de contenido (F)**.
 - **El operador de la red** puede modificar el equilibrio cambiando el **ancho de banda medio en los enlaces de subida de la conexión de acceso ($BW_{\text{subida, usuario}}$)**
- En condiciones de trabajo normales y en estado estacionario, se puede asumir que **T + S es constante**. A corto y medio plazo, F y $BW_{\text{subida, usuario}}$ pueden considerarse constantes, ya que aumentan sólo ocasionalmente y, además N_{pet} puede asumirse también constante en el tiempo, al haber diversos factores que limitan el número máximo de peticiones activas por usuario. Este resultado implica que hay una **compensación directa entre el tiempo medio de descarga y el tiempo medio de silencio**.

V. SIMULACIONES

Para verificar la precisión del modelo analítico, se ha llevado a cabo un conjunto de simulaciones mediante una herramienta desarrollada específicamente para estudiar el comportamiento de eMule, que es uno de los sistemas P2P más populares.

Aunque no se supuso ninguna distribución estadística para el modelo analítico, en las simulaciones se modeló el tiempo de silencio (S) mediante una distribución exponencial negativa de media igual a 1 día, por simplicidad. En las simulaciones en las que no se especifica algo distinto, se ha asumido $N_{\text{pet}} = 7$ peticiones (distribuidas de forma binomial); $F = 700$ MB; $BW_{\text{subida, usuario}} = 128$ kbps; $BW_{\text{bajada, usuario}} = 512$ kbps y $N_{\text{usuarios}} = 1000$. Todos los resultados se presentan con un intervalo de confianza del 95%.

La figura 2 muestra cómo el tiempo medio de descarga permanece constante conforme varía el número de usuarios del sistema y la figura 3 muestra que tiempos medios de silencio más grandes producen tiempos de descarga más cortos y que, como predecía el modelo, hay una relación lineal entre ellos. Así, también tenemos que permanece constante la suma de ambos.

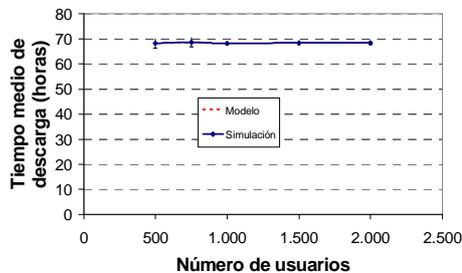


Fig. 2. Tiempo medio de descarga vs. número de usuarios en el sistema

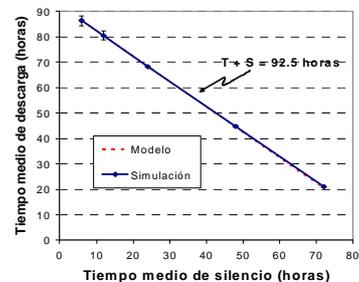


Fig. 3. Tiempo medio de descarga vs. tiempo medio de silencio

En la figura 4, se muestra que, a menos peticiones simultáneas, menor es el tiempo de descarga y, por consiguiente, hay una mejor QoE. En la figura 5, se muestra que cuanto mayor es el ancho de banda medio de los *uplinks*, menor es el tiempo medio de

descarga. Sin embargo, el tiempo medio de descarga es el mismo para anchos de banda de subida de 512 kbps y 1024 kbps. Esto se debe a que, en estas dos situaciones, el ancho de banda de bajada (512 kbps) se convierte en el cuello de botella del sistema, en vez del ancho de banda de subida.

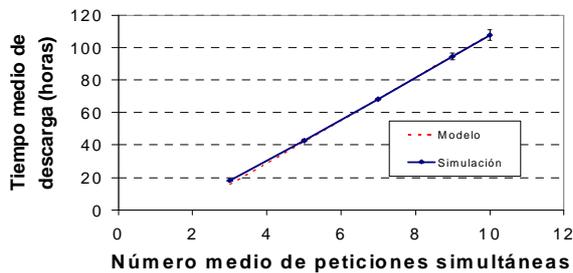


Fig. 4. Tiempo medio de descarga vs. número medio de peticiones simultáneas

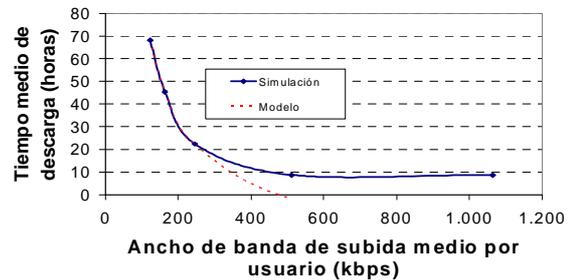


Fig. 5. Tiempo medio de descarga vs. ancho de banda de subida medio por usuario

VI. CONCLUSIONES

- Proporciona una expresión simple y compacta para evaluar los efectos en la dinámica de una red P2P ante un cambio en las condiciones de contorno, y es válida para cualquier aplicación de este tipo.
- No presupone ninguna distribución estadística concreta para las variables principales.
- Proporciona resultados bastante precisos cuando los parámetros están correctamente caracterizados.
- El punto de equilibrio no depende del número de usuarios en el sistema.
- El modelo se convierte en una herramienta útil para los agentes del servicio, ya que permite identificar fácilmente los efectos en la QoE y en la carga de la red tras cualquier cambio en alguno de los parámetros clave.
- Es fácil de entender el efecto que tiene el comportamiento de los usuarios en el punto de equilibrio del sistema.
- Consigue modelar cómo afecta el ancho de banda medio de los *uplinks* de los usuarios al punto de equilibrio. En particular, se tiene que un incremento del ancho de banda de subida medio conduce a un aumento en el tráfico ofrecido.
- Permite determinar indirectamente los tiempos de silencio a partir de medidas de tráfico y de los tiempos de descarga. En condiciones normales (asumiendo F , $BW_{subida,usuario}$ y N_{pet} constantes a corto plazo), se puede asumir que $T + S$ es constante, lo que implica que hay una compensación entre el tiempo medio de descarga y el tiempo medio de silencio.

En trabajos futuros, se tratará de refinar el modelo para que facilite el estudio de las aplicaciones P2P en escenarios con conexiones de acceso simétricas o heterogéneas, así como su ampliación a otros tipos de aplicaciones P2P, como las de *streaming* o las de VoIP.

REFERENCIAS

- [1] T. Karagiannis, A. Broido, N. Brownlee, K.C. Claffy y M.Faloutsos, "P2P dying or just hiding?", en Proceeding of IEEE Globecom 2004, Dallas, USA, Noviembre/Diciembre 2004.
- [2] Página de inicio de Joost, <http://www.joost.com/> (Verificada en Septiembre, 2008)
- [3] Página de inicio de SopCast, <http://www.sopcast.org/> (Verificada en Septiembre, 2008)
- [4] Y. Kulbak y D. Bickson, "The eMule Protocol Specification", Enero 2005.
- [5] K. Tutschku. "A Measurement-based Traffic Profile of the eDonkey Filesharing Service", en Proceedings of the 5th annual Passive & Active Measurement Workshop, Antibes Juan-les-Pins, Francia, Marzo 2004.
- [6] S. Sen y J. Wang, "Analyzing peer-to-peer traffic across large networks", IEEE/ACM Transactions on Networking, Vol. 12, No. 2, Abril 2004.
- [7] X. Hei, C. Liang, J. Liang, Y. Liu y K.W. Ross, "Insights into PPLive: A measurement study of a large-scale P2P IPTV system", en Proceedings of Internet Protocol TV (IPTV) Services over World Wide Web Workshop in WWW2006, Edimburgo, Escocia, Mayo, 2006.
- [8] S. Tewari y L. Kleinrock, "Analytical Model for BitTorrent-based Live Video Streaming", en Proceedings of IEEE NIME 2007 Workshop, Las Vegas, EEUU, Enero, 2007.
- [9] M.L. García-Osma, F.J. Ramón Salguero, G. García de Blas, J. Andrés Colás, J. Enriquez Gabeiras, S. Pérez Sanchez, y R. Trueba Fernandez, "Burrito: A simulation tool for P2P networks", en Proceedings 3rd International Workshop on Internet Performance, Simulation, Monitoring and Measurement IPS-MoMe 2005 "Internet Study 2007" por Ipoque, disponible en <http://www.ipoque.com/resources/internet-studies/internet-study-2007> (Verificada en Septiembre, 2008)
- [10] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, y J. Zahorjan, "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload". En Proceedings of the SOSP '03 (Bolton Landing, NY, Octubre 2003).
- [11] M.L. García Osma, F.J. Ramón Salguero, G. García de Blas, J. Andrés Colás, J. Enriquez Gabeiras, S. Pérez Sanchez, y R. Trueba Fernandez, "Enabling local preference in peer-to-peer traffic", COST 279 TD(04)017 technical report, 2004.
- [12] J. Enríquez Gabeiras, F. J. Ramón Salguero, G. García de Blas "Ancho de banda y P2P: análisis del tráfico de la red", disponible en <http://sociedaddelainformacion.telefonica.es/jsp/articulos/detallehemeroteca.jsp?elem=1773&origen=3> (Verificada en Septiembre, 2008)