

EVOLUCIÓN Y CONTENIDO DE LAS REDES P2P DE COMPARTICIÓN DE FICHEROS DE DOMINIO PÚBLICO

Jaime Lloret Mauri

Departamento de Comunicaciones
Universidad Politécnica de Valencia
jlloret@dcom.upv.es

Miguel A. Granados Tenorio

Departamento de Comunicaciones
Universidad Politécnica de Valencia
migrate@epsug.upv.es

Fernando Boronat Seguí

Departamento de Comunicaciones
Universidad Politécnica de Valencia
fboronat@dcom.upv.es

Abstract- Since the appearance of Peer-To-Peer filesharing networks a few years ago, many Internet users have chosen this technology to search programs, films, songs, documents, etc. The number of users using them is growing every day, however, the number of users and the number of files shared in some P2P filesharing networks is decreasing. In this article, six public P2P networks (Gnutella, FastTrack, Opennap, Edonkey, MP2P and Soulseek) have been analyzed. We have tracked their evolution in terms of connected users and number of shared files for two years. The results are compared and discussed with previous measurements taken by other authors. We have checked that those analyzed P2P networks seems to be specialized in different type of files.

I. INTRODUCCIÓN

Actualmente existen gran variedad de redes P2P y algunas de ellas con gran número de software cliente. La primera distinción que se debe realizar es entre las redes P2P y los clientes P2P. Las redes P2P son un conjunto de reglas e interacciones que permiten conectar a los clientes P2P. El cliente P2P es una aplicación de ordenador que permite interactuar con otros clientes de la misma red. El número de redes P2P de dominio público sigue incrementándose, cada vez tienen mejores prestaciones y sus clientes tienen más funcionalidades.

P2P de compartición de ficheros es una de las variantes de P2P que más adeptos está teniendo. Si bien existen usuarios con la intención de descargar ficheros de la red, sin intención de proporcionar alguno, existen muchos otros usuarios con la idea de compartir lo que tienen con toda la comunidad sin importarle quién es la otra persona que está descargando el fichero que tiene compartido en su disco duro.

Las principales redes P2P públicas de compartición de ficheros que existen actualmente en Internet son 6: Gnutella [1], FastTrack [2], Opennap [3], Soulseek [4], Edonkey [5] y MP2P [6], aunque existen muchas otras no tan comunes [7]. Algunas de estas redes son propietarias, sin embargo, poseen aplicaciones clientes de código abierto desarrollados por usuarios utilizando ingeniería inversa.

Las redes P2P se pueden clasificar según los parámetros que se tengan en cuenta (estructura, grado de centralización, tipo de descarga, etc.). En todos estos tipos de redes, la

transferencia de información se realiza directamente entre los clientes finales, no teniendo que intervenir ningún servidor central como intermediario de dicha transferencia.

Debido al impacto social que ha tenido y está teniendo las redes P2P de compartición de ficheros, tanto empresas como investigadores invierten tiempo y dinero en analizar diversos factores de éstas redes. La primera preocupación de la sociedad ha sido la legalidad de los ficheros que son intercambiados en dichas arquitecturas [8]. Seguidamente empezaron a estudiarse los riesgos que estas presentan tanto para los usuarios en sus hogares particulares [9], como para empresas en las que algunos de sus trabajadores que utilizan los ordenadores de su puesto de trabajo para conectarse a estas redes [10]. A esto cabe añadir que la mayoría de las redes P2P de compartición de ficheros no se han creado teniendo en cuenta la seguridad (autenticación, permisos sobre ficheros o integridad de los ficheros), incluso algunos de los programas cliente tienen “adware” o “spyware” permitiendo que aparezcan ventanas Web de advertencia o que se pueda realizar un seguimiento de los usuarios en la Web. Aun así, teniendo en cuenta lo anteriormente expuesto, sigue habiendo gran crecimiento en la utilización de las redes P2P por parte de los usuarios de Internet.

Algunos ISP comenzaron a observar que sus redes se encontraban saturadas y en ocasiones el tráfico P2P alcanzaba el 60% del tráfico total [11], debido a la cantidad de usuarios que empezaban a utilizar P2P de compartición de ficheros. Resultados no tan espectaculares, pero que si dan destellos de la magnitud de su utilización, se pueden obtener de las mediciones realizadas por los administradores de Internet2 el 16 de febrero de 2004 [12], donde el 10.46% del tráfico que atraviesa sus redes es P2P de compartición de ficheros. También se pueden observar en CAIDA (Cooperative Association for Internet Data Analysis), donde el tráfico en Internet está dominado por protocolos P2P de compartición de ficheros y por HTTP [13].

Podemos encontrar algunos artículos donde se ha analizado el número de usuarios en línea o en promedio que tienen algunas arquitecturas [14] y en algunas ocasiones el número de usuarios máximo (Napster: 20 millones de usuarios conectados con 1,5 billones de canciones compartidas en julio de 2000 [15]). También existen artículos donde se ha muestreado y analizado el tanto por ciento de usuarios, de algunas regiones del mundo, conectados a una u otra red [16]. Incluso existen estudios del coste económico que conlleva el descargar archivos,

teniendo en cuenta el tiempo necesitado para su descarga, utilizando estas arquitecturas [17].

En ninguno de los artículos y trabajos anteriormente mencionados se ha realizado el estudio desde dentro de las propias redes utilizando como fuente la información el propio protocolo de la red P2P. En todas las redes que hemos analizado, el protocolo de dicha red permite obtener el número de usuarios, número de ficheros, cantidad de información compartida y tipo de archivos compartidos dentro de cada una de las redes P2P.

El presente artículo está estructurado como se explica a continuación. En la sección 2 se muestra la evolución que han tenido las redes P2P que hemos analizado. La sección 3 muestra la evolución que ha tenido el contenido de dichas redes comparándolo con mediciones realizadas por otros autores. En la sección 4 se ha medido el porcentaje de video, audio, documentos, programas y archivos comprimidos que tienen actualmente dichas redes. Finalmente, en la sección 5, están las conclusiones.

II. EVOLUCIÓN DE LAS REDES P2P

En las figuras 3 y 4 se puede ver la evolución de usuarios y el número de archivos compartidos que han tenido las redes Gnutella, FastTrack, Opennap, Edonkey entre marzo de 2003 y marzo de 2005. Las medidas de las redes MP2P y Soulseek han sido tomadas entre marzo de 2004 y marzo de 2005.

Para realizar las medidas hemos utilizado las siguientes aplicaciones clientes P2P: Para la red Gnutella, el cliente Limewire, para la red FastTrack el cliente KaZaa Lite, para la red Opennap el cliente Xnap, para la red eDonkey el cliente eDonkey2000, para la red MP2P el cliente Piolet y para la red Soulseek el cliente Nicotine. Todos ellos pueden encontrarse en la referencia [7].

Se debe tener en cuenta que la réplica de archivos en los usuarios de las arquitecturas P2P se rige por la ley de la potencia (Power Law) y más concretamente la ley de Zipf, tal como se demuestra en los artículos [18][19][20], donde se destaca que un pequeño número de usuarios tienen la gran mayoría de archivos. Hemos podido observar, con los datos obtenidos, que todas las redes P2P analizadas han aumentado la cantidad de archivos compartidos por usuario. Esto tiene sentido en cuanto a que la cantidad total de archivos compartidos en una red P2P siempre tiende a crecer debido a que los archivos más populares tienden a replicarse. Sin embargo, se debe tener en cuenta que la capacidad de los discos duros de los usuarios de estas redes es limitada y por tanto, numerosos usuarios suelen grabar en discos ópticos los archivos descargados, borrándolos posteriormente de sus discos duros. Aún así, se ha observado que los usuarios de la red Gnutella, durante el tiempo que hemos estado tomando medidas, han tenido una media de 328 archivos compartidos en sus discos duros. Sin embargo, los usuarios de la red FastTrack han tenido la media de 224 archivos compartidos por usuario. Los usuarios de la red OpenNap han tenido la media de 579 archivos por usuario. Los usuarios de la red eDonkey tienen una media de 87 archivos por usuario. Por último, los usuarios de MP2P han tenido una media de 245 archivos por usuario. La red Soulseek no permite saber cuantos archivos existen compartidos por todos los usuarios y por tanto no hemos podido sacar la media. Se puede

encontrar información más detallada sobre las medidas realizadas del número de usuarios, el número total de ficheros compartidos y la cantidad de información compartida, durante una semana, en cada una de estas redes P2P en [21].

III. EVOLUCIÓN DEL CONTENIDO DE LAS REDES P2P

En este apartado vamos a comparar las medidas que hemos realizado con diferentes medidas tomadas por otros investigadores.

Los investigadores J. Chu, K. Labonte y B. Levine, en su artículo [22], desde el 24 de febrero hasta el 25 de marzo del 2002, tomaron medidas para la red Gnutella con el cliente Limewire. En este artículo se buscó qué archivos (avi, mp3, mpg, etc.) eran los más populares en la red Gnutella. Ahora hemos vuelto a realizar las mismas búsquedas de archivos para ver la disponibilidad actual de dichos archivos. En el artículo se crearon 3 listas que contenían: Los 50 ficheros más populares, los 50 ficheros de audio más populares y los 50 ficheros de vídeo más populares en ese momento. Tras buscar los mismos archivos en la red Gnutella, con el mismo cliente, se ha observado lo siguiente:

- Todos los archivos que estaban en el “top 50 de todo tipo de archivos” en el artículo han descendido su número de réplicas en esta red. No hay ninguno que se haya mantenido o aumentado. La media de decrecimiento está entorno al 75,6%.
- Todos los archivos que estaban en el “top 50 de audio” han descendido con una media del 74,1%.
- En el “top 50 de videos” existe gran disparidad. En más de la mitad de los videos, el número de usuarios que poseen dichos archivos ha crecido. La gran mayoría de los vídeos que han aumentado su disponibilidad son de carácter pornográfico.

Según las medidas de la sección anterior, el número de archivos de las diferentes redes P2P analizadas continúa creciendo. Sin embargo, como hemos podido comprobar, los archivos antiguos, que antes estaban muy replicados, tienden a disminuir su disponibilidad a medida que pasa el tiempo. Esto es debido a que aparecen nuevos archivos que son más descargados por los usuarios aunque tiendan a desaparecer las réplicas de los antiguos. Por lo tanto, los ficheros más descargados por los usuarios no son siempre los mismos. Estas observaciones no se aplican a los archivos de carácter pornográfico, es más, estos archivos aumentan su disponibilidad conforme pasa el tiempo.

Por otra parte, los investigadores Subhabrata Sen y Jia Wang, en su artículo [23] tomaron medidas de la red Gnutella y FastTrack entre el 10 de septiembre y el 16 de diciembre de 2001. En estas fechas, ambas redes se encontraban en pleno crecimiento, habiendo cerca de 197.445 usuarios en Gnutella y 4.450.149 usuarios en la red FastTrack. Si lo comparamos con las medidas que hemos realizado, entre dichas fechas y marzo de 2003 ambas redes llegaron a un máximo en usuarios y a partir de ésta empezaron a decrecer. Aunque, el número de usuarios de la red Gnutella está volviendo a crecer, en red FastTrack sigue todavía decreciendo.

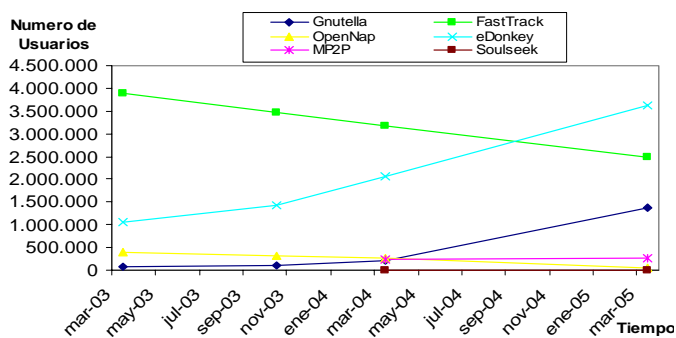


Figura 3. Número de usuarios las 6 redes analizadas

Los investigadores Beverly Yang y Hector Garcia-Molina, en su artículo [24] utilizaron los resultados experimentales que tomaron de la red OpenNap a finales del año 2000. En éstos, los usuarios de dicha red tienen una media de 168 ficheros por usuario. Como hemos podido constatar según nuestras medidas, esta media de archivos ha estado variando a lo largo de los años. En marzo de 2003, cada usuario tenía una media de 586 archivos. En octubre de 2003, esta media disminuyó a 444 archivos. En marzo de 2004 aumentó a 619 archivos por usuario. En marzo de 2005, la media es hemos obtenido es de 665 archivos por usuario. Actualmente, esta media, tiende a incrementarse.

IV. CONTENIDO ACTUAL DE LAS REDES P2P

Con la intención de saber en qué porcentaje se encuentran diferentes tipos de archivos en las 6 redes analizadas, hemos clasificado 5 tipos de archivos:

- Video: avi, mpg, asf, mpeg, rm y wmv.
- Audio: mp3, ogg, wma, wav, aac y ac3.
- Documentos: pdf, doc, txt, rtf, wps, sxw y pps.
- Programas: exe, com, msi y sh.
- Archivos comprimidos: ace, rar, arj, zip y gz.

Esta clasificación nos permite saber qué tipos de archivos son más populares y, por tanto, replicados entre los usuarios de las redes P2P. En las figuras 5, 6, 7, 8, 9 y 10, se puede observar que en todas las redes P2P, excepto en la eDonkey, el mayor porcentaje de ficheros compartidos son de tipo audio. Esto es debido en gran medida a los orígenes de las redes P2P, pues, en sus inicios, estas redes se utilizaron exclusivamente para la compartición de ficheros de audio. Pero sin embargo el segundo tipo de ficheros compartidos en todas ellas varía notablemente. En la tabla 1 se muestra la clasificación de los tipos de archivos para cada una de las redes analizadas.

Las medidas realizadas permiten comprobar que la mayoría de archivos compartidos en estas redes son de tipo multimedia, no utilizándose tanto para otro tipo de archivos. La red cuyo porcentaje de archivos multimedia contiene (a parte de MP2P que sólo contiene archivos de audio) es, Soulseek con un 98%. A ésta le siguen seguida e Gnutella con un 77%, OpenNap con un 73%, FastTrack con un 66% y por último eDonkey con un 63%.

Por otra parte, cabe destacar que la red con mayor tamaño total de información compartida es la red eDonkey. Esto es debido a que tanto los vídeos que en ésta se comparten como los archivos comprimidos, son de gran tamaño. En ocasiones, estos archivos superan los 4 GB.

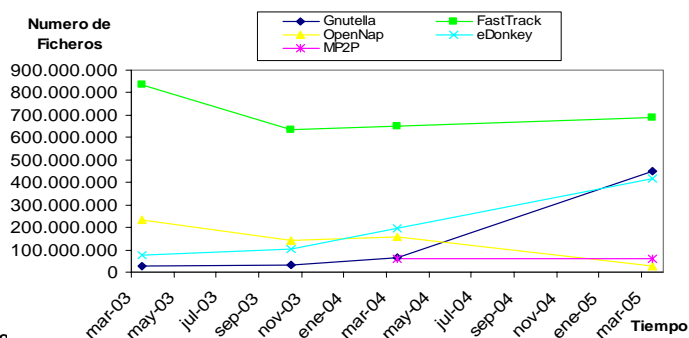


Figura 4. Número de ficheros las 6 redes analizadas

Si tenemos en cuenta que los usuarios que tienen mayor ancho de banda, están más tiempo en línea descargando archivos de la red P2P (tal como se observó en [25]), estos serán los que tienen los archivos más populares.

V. CONCLUSIONES

No existe ninguna relación entre la cantidad de usuarios que se conectan a una red P2P y la cantidad de ficheros que existen compartidos en esta red P2P. Las redes OpenNap y MP2P tenían prácticamente el mismo número de usuarios en marzo de 2004, sin embargo la red OpenNap tenía tres veces más ficheros compartidos que la red MP2P. Por otra parte, la red eDonkey que ha tenido siempre más usuarios que la red OpenNap, sin embargo, OpenNap entre marzo de 2003 y enero de 2004 tuvo más ficheros compartidos que eDonkey. Hemos podido comprobar que la red con mayor media de archivos por usuario es la red OpenNap con 579 ficheros por usuario.

El tamaño total de la información compartida en las redes P2P, no depende de la cantidad de ficheros compartidos. La red FasTrack es la red con más ficheros compartidos, sin embargo, la red con mayor cantidad total de información compartida es eDonkey debido a que la mayoría de archivos que comparte son de gran tamaño.

Aquellas redes P2P cuyo número de usuarios es estable o tiende a decrementarse, experimentan una subida de número de archivos por usuario mayor que en las redes donde crece el número de usuarios, esto es debido a que los usuarios replican más veces los ficheros compartidos por otros usuarios. Aun así, los archivos antiguos tienden a desaparecer. Los nuevos archivos que se introducen en estas redes, aumentan rápidamente su disponibilidad, pero a medida que pasa el tiempo, empieza a disminuir. Sin embargo, los vídeos de carácter pornográfico, independientemente de su antigüedad, siguen aumentando su disponibilidad en estas redes tal como se ha demostrado en la sección 3. El número de usuarios de algunas redes P2P antiguas está decrementándose debido a la aparición de nuevas redes P2P que atraen a los usuarios de las redes P2P antiguas. Sin embargo, el número total de usuarios conectados a redes P2P de compartición de ficheros sigue aumentando.

Como se puede comprobar en [26], la popularidad de los archivos en las redes P2P tiene gran relación con su popularidad en buscadores Web. En este artículo se ha demostrado que, además, estas redes tienden a especializarse en diferentes tipos de archivos aunque en ellas se permite la compartición de todo tipo de ficheros.

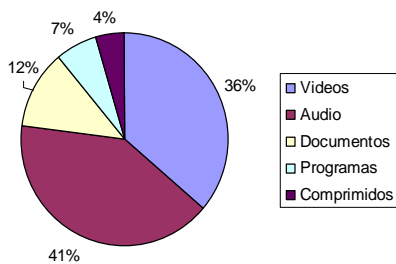


Fig. 5. Contenido de la red Gnutella

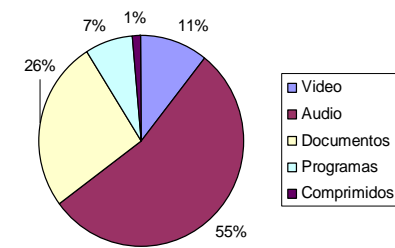


Fig. 6. Contenido de la red FastTrack

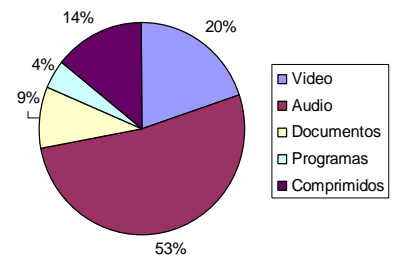


Fig.7. Contenido de la red OpenNap

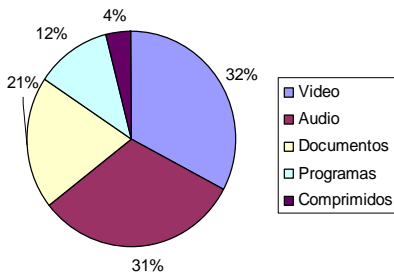


Fig. 8. Contenido de la red eDonkey

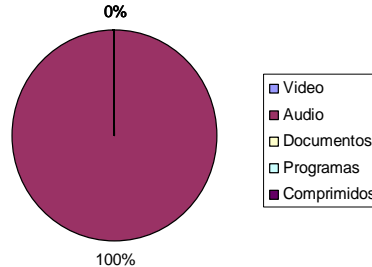


Fig. 9. Contenido de la red MP2P

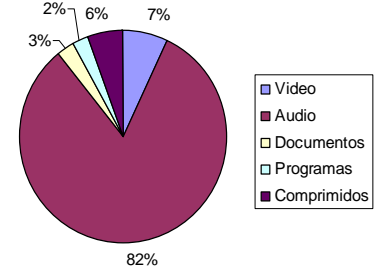


Fig. 10 Contenido de la red Soulseek

	Video	Audio	Documentos	Programas	Comprimidos
Gnutella	2°	1°	3°	4°	5°
FastTrack	3°	1°	2°	4°	5°
OpenNap	2°	1°	4°	5°	3°
eDonkey	1°	2°	3°	4°	5°
MP2P	-	1°	-	-	-
Soulseek	2°	1°	4°	5°	3°

Tabla 1. Clasificación de archivos en las redes P2P

En caso de querer realizar la búsqueda de un tipo de archivo determinado, bastaría tener en cuenta los usuarios de cada una de las arquitecturas (medidos en la sección 3) y el porcentaje de ese tipo de archivos en las arquitecturas (medidos en la sección 4) para obtener en qué arquitectura se tendrá mayor probabilidad de encontrar el fichero deseado.

REFERENCIAS

- [1] Eytan Adar and Bernardo Huberman. "ree riding on gnutella. First Monday" 5(10), October 2000.
- [2] Nathaniel Leibowitz, Matei Ripeanu, and Adam Wierzbicki, "Deconstructing the Kazaa Network" 3rd IEEE Workshop on Internet Applications (WIAPP'03), San Jose, CA. June 2003.
- [3] OpenNap Website, <http://opennap.sourceforge.net/>
- [4] Soulseek Website, <http://www.slsk.org>
- [5] Oliver Heckmann and Axel Bock. "The eDonkey 2000 Protocol." Technical Report KOM-TR-08-2002, Multimedia Communications Lab, Darmstadt University of Technology, December 2002.
- [6] MP2P Website <http://www.blubster.com/protocol1.html>
- [7] Wikipedia, <http://www.wikipedia.org/wiki/Peer-to-peer>
- [8] Palisade. Available at: http://www.palisadesys.com/news&events/press_releases/p2pstudyrelease.shtml
- [9] Staff report prepared for Rep. Tom Davis and Rep. Henry a. Waxman, Reform. "File-sharing programs and peer-to-peer networks privacy and security risks". United States House of Representatives Committee on Government. Mayo 2003
- [10] AssetMetrix Research Labs, "Corporate P2P (Peer-To-Peer) Usage and Risk Analysis", Julio 2003
- [11] Sandvine Incorporated White Paper, "Peer-to-Peer File Sharing: The impact of filesharing on service provider networks". September 2002.
- [12] Internet2 Website: <http://netflow.internet2.edu/weekly/20040216/>
- [13] CAIDA Website <http://www.caida.org>
- [14] Stefan Saroiu, P. Krishna Gummadi and Steven D. Gribble, A "Measurement Study of Peer-to-Peer File Sharing Systems," Department of Computer Science & Engineering, University of Washington, Tech Rep. UW-CSE-01-06-02.
- [15] The Pew Internet & American Life Project's Online Music, September 28, 2000. Available at http://www.pewinternet.org/pdfs/PIP_Online_Music_Report2.pdf
- [16] Sanvine Incorporated White Paper, "Regional Characteristics of P2P: File sharing as a multi-application, multi-national phenomenon", October 2003
- [17] Artur Marques at http://arturmarques.com/docs/economics/arturmarques_dot_com_freelodding.pdf
- [18] Z. Ge, D. R. Figueiredo, S. Jaiswal, J. Kurose, D. Towsley. "Modeling Peer-Peer File Sharing Systems", Proceedings IEEE Infocom 2003, April 2003.
- [19] Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steven D. Gribble, Henry M. Levy, John Zahorjan, "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload", Proceedings of the nineteenth ACM symposium on Operating systems principles, 2003, pag. 314-329.
- [20] Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker, "Search and replication in unstructured peer-to-peer networks". Proceedings of the 16th international conference on Supercomputing, ACM Press, 2002, Pag. 84-95.
- [21] J. Lloret, B. Molina, C. Palau, M. Esteve, "Public Peer-To-Peer Filesharing Networks' Evaluation". The 2nd IASTED International Conference on Communication and Computer Networks. MIT Cambridge, MA (USA). November 2004
- [22] J. Chu, K. Labonte, and B. Levine. "Availability and locality measurements of peer-to-peer file systems". In Proceeding of SPIE ITCOM: Scalability and Traffic Control in IP Networks, vol. 4868, July 2002.
- [23] Subhabrata Sen, Jia Wang. "Analyzing peer-to-peer traffic across large networks". IEEE/ACM Transactions on Networking (TON) archive. Volume 12, Issue 2. April 2004. Pag. 219-232
- [24] Yang, B., Garcia-Molina, Hector. "Comparing Hybrid Peer-to-Peer Systems". Proceedings of the 27th Intl. Conf. on Very Large Data Bases. Pag. 561-570. October 2001.
- [25] Subhabrata Sen and Jia Wang. "Analyzing peer-to-peer traffic across large networks", IEEE/ACM Transactions on Networking, Volume 12, Issue 2, pp 219-232. April 2004.
- [26] J. Lloret, J. R. Diaz, J. M. Jiménez, M. Esteve. "The Popularity Parameter in Unstructured P2P Filesharing Networks". WSEAS Transactions on Computers, Issue 6, Volume 3. December 2004.