

Revisión de sistemas P2P

Alberto Los Santos Aransay, albertolsa@gmail.com

1. INTRODUCCIÓN

Desde hace unos años las redes de ordenadores distribuidas etiquetadas como “peer-to-peer” (de igual a igual) o P2P, son un foco importante de investigación. Estas arquitecturas fueron diseñadas para compartir diversos recursos informáticos (contenidos, almacenamiento, procesamiento...), mediante comunicación directa entre los propios terminales sin la intervención directa de un servidor o autoridad central.

Estas redes suponen un alto porcentaje del tráfico total de Internet, desde el 2002 la mayor parte del ancho de banda consumido era debido al intercambio P2P de ficheros, por ejemplo en 2005 representaba el 65%. A partir de 2007 la tendencia se ha invertido, siendo el principal generador de tráfico las comunicaciones HTTP (principalmente debido a la aparición de las redes sociales, y la limitación por parte de los proveedores de conexión del tráfico P2P), pero aún así supone más del 30% del tráfico total.

Basándome principalmente en la lectura recomendada [1], haré una revisión de los sistemas y aplicaciones P2P. Antes de introducir los sistemas P2P, detallaré en el capítulo 2 qué se entiende por redes P2P, y enumeraré sus principales aplicaciones.

En el capítulo 3, se indicarán las principales características de estos sistemas, y se clasificarán en función de las dos más importantes, la distribución de los nodos y la estructura de la red.

Considerando que el documento base fue publicado en 2004, existen diversas novedades en relación a nuevas aplicaciones, estudios y áreas de aplicación que serán introducidas en el capítulo 4.

Por último, en el capítulo 5 se hará una breve conclusión.



Fig. 1: Redes P2P e intercambio de información

2. DEFINICIÓN Y APLICACIONES DE LAS REDES P2P

No hay una definición totalmente acordada en relación a las redes P2P. Tal y como detalla el documento referencia de este trabajo, las definiciones más estrictas de redes peer-to-peer puras, aluden a sistemas totalmente distribuidos, en donde todos los nodos son completamente iguales en términos de funcionalidad y las tareas que llevan a cabo.

Estas definiciones no englobarían los sistemas que disponen de “supernodos” para trabajar (nodos que funcionan como mini-servidores localizados, elegidos dinámicamente) o aquellos que usan servidores centralizados para llevar una serie de tareas no fundamentales, los cuales también son aceptados como P2P. Por tanto, se podrían definir los sistemas P2P como:

“Sistemas distribuidos compuestos por nodos interconectados con dos características:

- El intercambio de recursos informáticos se realiza mediante conexión directa, sin la intermediación de un servidor central. Los servidores centralizados pueden, a veces, ser usados para otras tareas como arranque del sistema, la inserción de nuevos nodos en la red, obtener claves globales para la encriptación, etc.
- Capaces de adaptarse a la variedad e inestabilidad de la conexión, ajustándose automáticamente ante fallos y la alta variabilidad en el número de nodos.”

2.1. Aplicaciones de las redes P2P

Las arquitecturas peer-to-peer se usan para diversas aplicaciones, que van desde sencillos sistemas de intercambio de datos a más sofisticados sistemas que disponen una compleja infraestructura de soporte de servicios en Internet con multitud de opciones. Se pueden agrupar en:

- **Comunicación y colaboración.** En este grupo se incluyen los sistemas que proveen la infraestructura para realizar comunicaciones y colaboraciones directas, y en ocasiones en tiempo real, entre nodos. Algunos ejemplos son chats y aplicaciones de mensajería instantánea, IRC [2], Msn [3] y Yahoo [4] messenger, Skype [5], Jabber [6].
- **Cálculo distribuido.** Esta categoría incluye sistemas cuyo objetivo es el hacer uso del procesamiento no usado de los ordenadores disponibles en la red. Ejemplos de estos sistemas son Seti@home [7] y genome@home [8].
- **Soporte de servicios en Internet.** Diversas aplicaciones se han ido desarrollando sobre infraestructuras peer-to-peer para dar apoyo a servicios en Internet. Ejemplos comprenden sistemas P2P multicast, aplicaciones de seguridad...
- **Sistemas de bases de datos distribuidos,** por ejemplo, LRM [9] (Local Relational Model) o PIER [10], que es un motor de peticiones construido sobre una red P2P que permite realizar consultas sobre miles de ordenadores. Edutella [11] es un proyecto de código abierto que provee una infraestructura de meta-datos sobre la cuál hacer peticiones para aplicaciones P2P.
- **Distribución de contenidos.** La mayoría de los sistemas peer-to-peer actuales se encuentran en esta categoría, que incluyen sistemas e infraestructuras diseñados para compartir recursos digitales, intercambiando ficheros, publicando o almacenando contenidos, etc. Algunos ejemplos son Napster [12], Publius [13], Gnutella [14], FreeNet [15], Kazaa [16], MojoNation [17], Chord [18], eMule [19], BitTorrent [20].

3. CLASIFICACIÓN DE LOS SISTEMAS P2P

Las redes P2P pueden ser clasificadas según diversas características, aunque en este capítulo las etiquetaremos según su centralización y estructura.

Aunque las redes P2P puras, a priori, son descentralizadas, en la práctica como se ha comentado antes, no es totalmente cierto, pudiendo definir tres categorías:

- Arquitecturas totalmente descentralizadas. Todos los nodos en la red realizan las mismas tareas, actuando a la vez como servidores y clientes. Además no existe un nodo central.
- Arquitecturas parcialmente centralizadas. La base de estas redes es similar al de las arquitecturas totalmente descentralizadas, pero alguno de los nodos asumen un rol más importante, son los *supernodos*. Actúan como índices locales para los ficheros compartidos por los usuarios locales.
- Arquitecturas descentralizadas híbridas. En estos sistemas existe un servidor central, que permite la interacción entre los nodos, pero sin entrar directamente en el intercambio de la información. Este nodo se encarga de mantener los directorios de metadatos, que describen los archivos almacenados en el sistema.



Fig. 2: Arquitectura descentralizada híbrida, parcialmente descentralizada y totalmente descentralizada [21]

Según la estructura de la red también se pueden encontrar varias categorías:

- No estructuradas: La localización del contenido no tiene relación con la topología de la red. En estas redes el contenido tiene que ser localizado, por lo que son necesarios mecanismos de búsqueda que varían de métodos de fuerza bruta al reenvío de peticiones por los nodos de la red.
- Estructuradas: Estos sistemas surgieron para intentar cubrir los problemas de escalabilidad de los sistemas no estructurados. Las redes P2P estructuradas superan las limitaciones de redes no estructuradas manteniendo una tabla de hash distribuida (DHT) y permitiendo que cada peer sea responsable de una parte específica del contenido en la red.

- Indirectamente estructuradas: Aunque la localización del contenido no está completamente especificada, está afectada por el enrutamiento.

Tabla I. Clasificación de los sistemas e infraestructuras de distribución de contenidos en base a su estructura.

	Centralización		
	Híbrida	Parcial	No centralizada
No estructurada	Napster, Publius	Kazaa, Morpheus [22], Gnutella, Edutella	Gnutella, FreeHaven [23]
Infraestructura estructurada			Chord, CAN [24], Tapestry [25], Pastry [26]
Sistemas estructurados			OceanStore [27], Mnemosyne [28], Scan [29], PAST [30], Kademia [31], Tarzan [32]

3.1. Otras características

Las redes P2P requieren de ciertas características deseables para su correcto funcionamiento, entre ellas:

- **Escalabilidad.** Las redes P2P tienen un alcance mundial con un gran número de potenciales usuarios. Idealmente, cuantos más nodos estén conectados a una red P2P mejor será su funcionamiento. Para ello, cuando los nodos llegan y comparten sus propios recursos, los recursos totales del sistema aumentan. En una arquitectura servidor-cliente con un sistema fijo de servidores, el ingreso de más clientes podría significar una transferencia de datos más lenta para todos los usuarios.
- **Robustez y replicación de contenidos.** En general, la naturaleza distribuida de las redes peer-to-peer también incrementa la robustez en caso de haber fallos en la réplica excesiva de los datos hacia múltiples destinos, y en sistemas P2P puros, permitiendo a los *peers* encontrar la información sin hacer peticiones a ningún servidor centralizado de indexado.
- **Reparto de costes entre los usuarios e incentivos.** Se comparten o donan recursos a cambio de recursos. Según la aplicación de la red, los recursos pueden ser archivos, ancho de banda, ciclos de proceso o almacenamiento de disco.
- **Anonimato.** Es deseable que en estas redes quede anónimo el autor de un contenido, el editor, el lector, el servidor que lo alberga y la petición para encontrarlo siempre que así lo necesiten los usuarios. Muchas veces el derecho al anonimato y los derechos de autor son incompatibles entre sí, y la industria propone mecanismos como el DRM (Digital Rights Management) para limitar ambos. Además hay que recordar que en estos sistemas la información puede pasar por ciertos nodos, los cuales no son ni el origen ni destino de la comunicación, almacenando estos datos de forma transparente y sin participar activamente el usuario en estas acciones.
- **Seguridad.** Es una de las características deseables de las redes P2P menos implementada y bajo mayor investigación. Los objetivos de un P2P seguro serían identificar y evitar los nodos maliciosos, evitar el contenido infectado, evitar el espionaje de las comunicaciones entre nodos, creación de grupos seguros de nodos dentro de la red, protección de los recursos de la red y almacenamiento seguro de la información.

Estas características han supuesto muchos estudios para cubrir las menos desarrolladas, y aún

existen muchas iniciativas en estos ámbitos. En el siguiente capítulo se referencian algunos estudios que intentan cubrir las carencias descritas en la referencia principal [1].

4. EVOLUCIÓN DE LOS SISTEMAS P2P

Alrededor de los sistemas P2P aparecen numerosas novedades y estudios cada año, en muchas ocasiones debido a las posibilidades que ofrecen las nuevas tecnologías. Algunas de estas novedades están centradas en mejorar las debilidades de las arquitecturas ya existentes, otras presentan nuevas aplicaciones, que pretenden cubrir áreas y aspectos en las cuales quizás no se habían aplicado este tipo de soluciones.

4.1. Revisión de estudios hechos en los últimos años sobre los sistemas P2P

Dentro de la revisión realizada sobre las tecnologías P2P existen numerosos trabajos detallando posibles implementaciones sobre las redes P2P para mejorar su funcionamiento.

- **Balaneo de carga.** El balanceo de carga es un aspecto crítico para el correcto funcionamiento de las redes peer-to-peer. En este sentido, numerosos protocolos de balanceo de carga pueden ser encontrados, por ejemplo en [33] se detallan dos soluciones, que pretenden garantizar un rendimiento óptimo sobre la estructura de datos hash sobre la que trabaja Chord.
- **Seguridad.** La seguridad en las redes P2P es algo fundamental, y numerosos estudios sobre el tema pueden ser encontrados. Como se ha indicado antes, la seguridad en redes P2P se refiere a diversos aspectos, y por tanto los estudios también. En [34] por ejemplo, se propone un modelo para evitar la distribución de código malicioso en redes peer-to-peer.
- **P4P:** Es el acrónimo *para Proactive network Provider Participation for P2P* (participación activa del proveedor de red en P2P). Es el siguiente paso en el desarrollo de servicios P2P, ya que permite minimizar el número de saltos requeridos para las transferencias de archivos, eligiendo los pares más cercanos al usuario (pertenecientes al mismo proveedor) en primer lugar. Por tanto, P4P no es un protocolo P2P, sino un medio para que los proveedores de Internet optimicen el tráfico de las redes P2P de sus usuarios.
- **Semántica:** La semántica, y en relación a Internet, la Web-Semántica es un campo actualmente muy importante. Hay numerosos trabajos que intentan traer los metadatos a las redes P2P, para facilitar la búsqueda de los recursos, o por agrupar los datos según las peticiones de los usuarios, marcar la reputación de los nodos [35].

4.2. Nuevas aplicaciones P2P

Desde la aparición de las primeras redes de intercambio de archivos P2P, los sistemas han ido evolucionando, surgiendo nuevas redes y áreas de aplicación de las redes peer-to-peer.

- **Redes de intercambio de contenidos:** En la actualidad las principales redes para la compartición de contenido en Internet son BitTorrent y eMule, dejando atrás a otras antiguas como Kazaa y Napster.
- **P2PTV:** El término P2PTV se refiere a aplicaciones diseñadas para distribuir señal de video en tiempo real sobre redes P2P. Se ha convertido en un servicio en auge, gracias al despliegue de redes cada vez con más capacidad. Sobre este concepto existen cada vez más programas P2P como por ejemplo, Tvants [36], TVUPlayer [37], PPLive [38], PPStream [39],

SopCast [40], y también diversos proyectos europeos como NAPA-WINE [41].

- **Nodos móviles:** Debido a las limitaciones de los teléfonos móviles, éstos no podían acceder a la mayoría de los servicios P2P sin hacerlo a través de servidores proxy. En [42] describen un protocolo para dispositivos móviles que trabaja sobre el interfaz Bluetooth, además de una aplicación de mensajería instantánea entre los dispositivos. Estas redes móviles con comunicaciones sin infraestructura se denominan *MANETs* (Mobile Ad-hoc Networks) y junto a las *VANETs* (Vehicular Ad-hoc Networks) son un gran campo de estudio que está siendo abordado por muchos investigadores en la actualidad [43] y [44].

5. CONCLUSIONES

En la actualidad hay multitud de sistemas peer-to-peer, sistemas distribuidos compuestos de nodos interconectados, que pueden comunicarse directamente para realizar muy diversas aplicaciones. El uso más extendido de estas aplicaciones es el intercambio de ficheros multimedia, pero con el transcurso de los años y la aparición de nuevas tecnologías y nuevas características de los dispositivos y redes sobre las que se sustentan las comunicaciones, nuevas aplicaciones están surgiendo en otros ámbitos.

En el futuro, es de esperar que sobre estas redes aparezcan novedosas aplicaciones, en cuanto se solventen diversos de los problemas expuestos (y principalmente la seguridad). Además, la combinación de estas arquitecturas con otras áreas como GRID [45] y la semántica, y nuevos escenarios como los sistemas vehiculares, abrirán muchas posibilidades para la investigación.

REFERENCIAS

- [1] Stephanos Androutsellis-Theotokis, Diomidis Spinellis, "A survey of peer-to-peer content distribution technologies", ACM Computing Surveys (CSUR), v.36 n.4, p.335-371, December 2004
- [2] Protocolo IRC: <http://www.rfc-es.org/rfc/rfc1459-es.txt>
- [3] Sitio oficial Msn messenger: <http://webmessenger.msn.com/>
- [4] Sitio oficial Yahoo messenger: <http://messenger.yahoo.com/>
- [5] Skype: <http://www.skype.com/>
- [6] Jabber: <http://www.jabber.org/>
- [7] The seti@home project web site. <http://setiathome.ssl.berkeley.edu>.
- [8] Modern Methods in Computational Biology. (Chapter Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology) Horizon Press.
- [9] Bernstein, P.,Giunchiglia, F.,Kementsietsidis, A.,Mylopoulos, J.,Serafini, L., y Zaihayeu, I. 2002. "Data management for peer-to-peer computing: A vision". In Proceedings of the Workshop on the Web and Databases (WebDB'02).
- [10] Huebsch, R., Hellerstein, J., Lanham, N., y Thau Loo, B. "Querying the internet with pier". In Proceedings of the 29th VLDB Conference. Berlin, Germany. 2003.
- [11] Nejd, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmer, M., y Risch, T. "Edutella: A p2p networking infrastructure based on rdf." In Proceedings of the 12th International Conference on World Wide Web. Budapest, Hungary. 2003.
- [12] Sitio web de Napster: <http://www.napster.com/>
- [13] Waldman, M., Ad, R., y Lf, C. "Publius: A robust, tamper-evident, censorship-resistant web publishing system". In Proceedings of the 9th USENIX Security Symposium.2000.
- [14] Gnutella: <http://www.gnutella.com/>
- [15] FreeNet: <http://freenetproject.org/>
- [16] Kazaa: <http://www.kazaa.com/>
- [17] The MojoNation web site. <http://www.mojonation.net>.

- [18] The Chord project: <http://www.pdos.lcs.mit.edu/chord>
- [19] eMule: <http://www.emule-project.net/>
- [20] BitTorrent: http://www.bittorrent.org/beps/bep_0003.html
- [21] Imagen extraída de: http://exploradoreselectronicos.net/wiki/index.php?title=Imagen:Topolog%C3%ADas_de_red.gif
- [22] Morpheus: <http://www.morpheus.com/>
- [23] Dingledine, R., Freedman, M., Andmolar, D. "The FreeHaven project: Distributed anonymous storage service." In Workshop on Design Issues in Anonymity and Unobservability. 67–95. 2000.
- [24] Ratnasamy, S., Francis, P., Handley, M., y Karp, R. "A scalable content-addressable network." In Proceedings of SIGCOMM 2001.
- [25] Zhao, B., Kubiawicz, J., y Joseph, A. "Tapestry: An infrastructure for fault-tolerant wide-area location and routing." Tech. Rep. UCB/CSD-01-1141, Computer Science Division, University. 2001.
- [26] Rowstron, A. y Druschel, P. "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems". In Proceedings of IFIP/ACM Middleware. Heidelberg, Germany. 2001.
- [27] Kubiawicz, J., Bindel, D., Chen, Y., Eaton, P., Geels, D., Gummadi, S., Weatherspoon, H., Weimer, W., Wells, C., y Zhao, B. "Oceanstore: An architecture for global-scale persistent storage." In Proceedings of ACM ASPLOS. 2000
- [28] Hand, S. y Roscoe, T. "Mnemosyne: Peerto-peer steganographic storage." In Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02). MIT Faculty Club, Cambridge, MA. 2002.
- [29] Chen, Y., Katz, R., y Kubiawicz, J. "Scan: A dynamic, scalable and efficient content distribution network." In Proceedings of International. Conference on Pervasive Computing. 2000.
- [30] DRUSCHEL, P. AND ROWSTRON, A. 2001. Past: A largescale, persistent peer-to-peer storage utility. In Proceedings of the Eighth Workshop on Hot Topics in Operating Systems.
- [31] Mayamounkov, P. y MAazieres, D. "Kademlia: A peer-to-peer information system based on the xor metric." In Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02). MIT Faculty Club, Cambridge, MA. 2002.
- [32] Freedman, M., Sit, E., Cates, J., Andmorris, R. "Introducing tarzan, a peer-to-peer anonymizing network layer". In Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02). MIT Faculty Club, Cambridge, MA. 2002.
- [33] David R.Karger y Matthias Ruhl. "Simple Efficient Load-Balancing Algorithms for Peer-to-Peer Systems" Theory of Computing Systems Springer New York. Volumen 39, Número 6 / noviembre de 2006. Páginas 787-804.
- [34] Rice, D.O., " A Proposal for the Security of Peer-to-Peer (P2P) Networks: a pricing model inspired by the theory of complex networks". Information Sciences and Systems, 2007. CISS '07. 41st Annual Conference. 14-16 March 2007. On page(s): 812-813.
- [35] Sergio Marti , and Hector Garcia-Molina. "Taxonomy of trust: Categorizing P2P reputation systems". Computer Networks. Volúmen 50, Issue 4, 15 Marzo 2006, Página 472-484.
- [36] Tvants: <http://www.tvants.com/>
- [37] TVUPlayer: <http://tvunetworks.com/>
- [38] PPLive: <http://www.pplive.com/en/index.html>
- [39] PPStream: <http://www.ppstream.com/>
- [40] SopCast: <http://www.sopcast.org/>
- [41] E. Leonardi, M. Mellia, A. Horvath, L. Muscariello, S. Niccolini, y D. Rossi. "BUILDING A COOPERATIVE P2P-TV APPLICATION OVER A WISE NETWORK: THE APPROACH OF THE EUROPEAN FP-7 STREP NAPA-WINE" IEEE Communications Magazine. April 2008. Pgs. 20 y 22. <http://www.telematica.polito.it/~emilio/emi/ieee/riviste/commag3.pdf>
- [42] Sumino, H. Ishikawa, N. Kato, T. "Design and implementation of P2P protocol for mobile phones" Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006. Fourth Annual IEEE International Conference. 13-17 March 2006
On page(s): 6 pp.-162
- [43] Jen-Wen Ding, Fa-Hung Meng, Yueh-Min Huang. "A real-time vehicle guidance system using P2P communication". Ubi-Media Computing, 2008 First IEEE International Conference. July 31 2008-Aug. 1 2008.

On page(s): 225-230.

[44] V. González, A. Los Santos, C. Pinart, F. Milagro, "Experimental demonstration of the viability of IEEE 802.11b based inter-vehicle communications", in Proc. of 4th Int'l Conference on Testbeds & Research Infrastructures for the Development of Networks & Communities (TRIDENTCOM 2008). Innsbruck (Austria), March 2008.

[47] Computación GRID: <http://www.gridcomputing.com/>