Análisis y gestión del tráfico peer-to-peer en la Red CTnet

Peer-to-Peer Traffic Analysis and Management in CTnet Network

A. M. Guirado, J. Malgosa y M. Escudero

Resumen

Las aplicaciones peer-to-peer (P2P) para la compartición de ficheros están cambiando drásticamente el papel de los Proveedores de Servicios de Internet (PSI). La razón principal es el uso indiscriminado del ancho de banda produciendo una ruptura del balance entre beneficios y costes. Además, deben tenerse en cuenta aspectos legales, debido a la piratería asociada a estas aplicaciones, que son especialmente perjudiciales en redes corporativas, institucionales y gubernamentales donde no se desea asumir la responsabilidad de transportar este tipo de tráfico de datos.

Este artículo presenta un análisis del protocolo eDonkey que es uno de los protocolos P2P más populares. El análisis se enfoca en estudiar aquellas características principales que permiten entender su comportamiento. Para ello se propone una metodología que identifica eficientemente los flujos TCP asociados. Además, se han realizado pruebas con un mecanismo basado en técnicas de spoofing de paquetes TCP Reset para mitigar el tráfico eDonkey y reducir sus efectos en la carga de la red.

El protocolo eDonkey es uno de los protocolos P2P más populares

Palabras clave: Peer to peer, medidas y análisis, tecnologías de gestión de red e identificación de protocolos.

Summary

Peer-to-peer (P2P) file-sharing applications are drastically changing the role of Internet Service Providers (ISP). The key reason is an unexpected and indiscriminate network bandwidth use that breaks the balance between costs and benefits. In addition, legal issues must be considered due to piracy associated with P2P applications. P2P effects are especially harmful in corporate, institutional and government networks that do not want to assume the responsibility to carry this type of data traffic.

This paper presents an analysis of eDonkey protocol, one of the most popular P2P protocols. The analysis is focused on studying those important characteristics that allow understanding its behavior. A methodology to identify efficiently the associated TCP flows is also proposed. A mechanism based on spoofing TCP Reset packets is also tested to mitigate the eDonkey traffic effects and alleviate the network load.

Keywords: Peer to peer, measurement and analysis, network management technologies and protocols identification.

1.- Introducción

Los protocolos utilizados por las aplicaciones P2P son cada vez más sofisticados, siendo capaces de utilizar puertos TCP y/o UDP arbitrarios, inclusive puertos well-known como 21 (ftp), 80 (web), etc... Recientes estudios muestran que entre un 30% y un 70% del tráfico P2P es intercambiado a través de puertos aleatorios. Aproximadamente el 60% del tráfico cursado por la red de un proveedor de acceso a Internet se debe a aplicaciones de intercambio de ficheros. Incumplimiento de los niveles de calidad de servicio, cuellos de botella, aumentos no planificados de ancho de banda, son consecuencias del crecimiento desproporcionado del tráfico P2P con las que se tienen que enfrentar los administradores de la red.

De entre todas las aplicaciones P2P el protocolo eDonkey es el más popular y dispone de más de 1.5 millones de usuarios simultáneos. Encuestas realizadas sobre 5.000 usuarios europeos en el verano de 2003, muestran que el 15% de los usuarios descarga al menos una película al mes. Para el caso de los usuarios españoles este porcentaje se amplia a un 38%.

Los efectos del fenómeno P2P se agravan enormemente en redes corporativas, institucionales o gubernamentales donde existen servicios y aplicaciones críticas cuyo funcionamiento debe ser

garantizado. En este caso aparece la necesidad imperiosa de definir políticas de uso del ancho de banda en segmentos críticos de la red, generalmente el enlace de salida a Internet, que permitan "identificar" y "controlar" el tráfico P2P, priorizando el tráfico generado por las aplicaciones corporativas. El presente artículo pretende dar respuesta a esta necesidad a partir del estudio y análisis de capturas reales de tráfico realizadas en la red CTnet.

2.- La red Ctnet

La red de Ciencia y Tecnología de la Región de Murcia, Red CTnet, es una red de telecomunicaciones y servicios telemáticos, de ámbito regional e interinstitucional, basada en una infraestructura tecnológicamente avanzada. La red ofrece servicios de interconexión, acceso a Internet y RedIRIS, y servicios telemáticos (correo, publicación de portales, creación de intranets, congresos virtuales, etc...)

De entre las tecnologías de acceso soportadas, que van desde conexiones a 56Kbps, mediante línea telefónica, a STM-1 en fibra óptica, la que dispone de más usuarios es la tecnología ADSL (Asymmetric Digital Subscriber Line) con más de 650 usuarios, generando gran parte del tráfico de la red. Debe entenderse usuario como centro público (colegios, aulas de libre acceso, etc...)

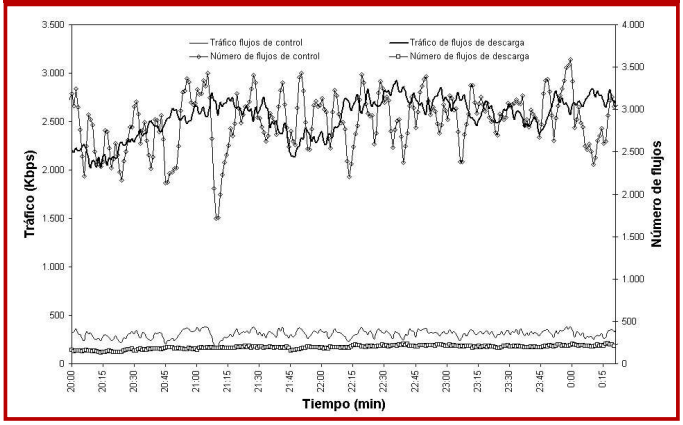
En los últimos meses, la gestión y control del tráfico es una cuestión primordial, para así poder garantizar que todos los servicios disponen del ancho de banda adecuado.

La gestión y control del tráfico es una cuestión primordial, para poder garantizar que todos los servicios disponen del ancho de banda adecuado

3.- Identificación de flujos TCP eDonkey

El algoritmo de identificación del protocolo eDonkey desarrollado, se basa en flujos TCP y en la inspección heurística de los primeros seis bytes del payload de los paquetes TCP. Estos bytes corresponden a la cabecera de los mensajes eDonkey que está formada por: 1 byte de protocolo que siempre toma el valor 0xE3, 4 bytes en formato LSB que indican la longitud del mensaje y un último byte que indica el tipo de mensaje enviado. El algoritmo consiste en la identificación de aquellos flujos, definidos por dirección y puerto origen, y dirección IP y puerto destino, cuyo primer paquete de datos (bit de control PUSH activo) tras el paquete SYN contiene el byte indicador del protocolo edonkey 0xE3 y el tipo de mensaje 0x01 (Hello-Message). A partir de ese momento cualquier paquete perteneciente al flujo es contabilizado como tráfico eDonkey sin necesidad de inspeccionar su payload. Además, el algoritmo también permite clasificar los flujos eDonkey en flujos de control, los asociados a las funciones de publicación y búsquedas de contenidos, y flujos de descarga, los asociados al intercambio de archivos.

FIG. 1: RELACIÓN ENTRE TRÁFICO Y FLUJOS EDONKEY



4.- Análisis del tráfico

Los principales resultados del análisis del tráfico capturado tras aplicar el algoritmo anterior son:

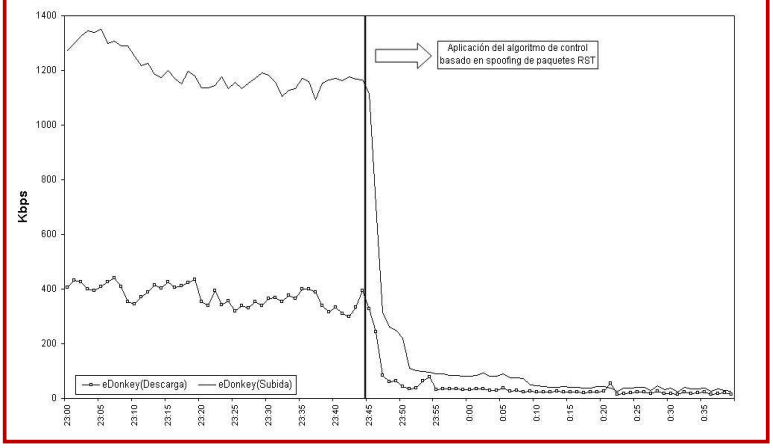
- El tráfico eDonkey produce un patrón característico asimétrico. A diferencia del tráfico normal, el tráfico en subida es mayor que el existente en descarga.
- Sin ningún tipo de restricción de ancho de banda el 80% del tráfico eDonkey, por defecto, está asociado a los puertos TCP 4661 y 4662. Sin embargo, al aplicar limitaciones de ancho de banda sobre estos puertos, el tráfico se regenera alcanzando velocidades similares, pero utilizando en este caso más de 100 puertos arbitrarios. Por tanto no es posible utilizar mecanismos basados en filtrado de puertos para limitar de forma efectiva el tráfico eDonkey.
- Los flujos de control son cortos, 62 segundos de duración media y constituyen el 90-95% del total de los flujos, superando 2.000 flujos en cualquier instante de tiempo (Figura 1).
- Los flujos de descarga son más largos, con una duración media de 664 segundos, y siguen una distribución de Heavy-Tail. Además generan el 90-95% del tráfico total aún siendo únicamente un 5-10% del total de los flujos. Así, cualquier mecanismo de control que pretenda conformar el tráfico eDonkey, limitándolo a un ancho de banda determinado, debería considerar sólo este tipo de flujos, ya que la gestión de los flujos de control, responsables de sólo el 5-10% del tráfico, no aportaría mejoras adicionales y consumiría muchos recursos en términos de CPU y memoria al ser tan numerosos.

Cualquier mecanismo de control que pretenda conformar el tráfico eDonkey, limitándolo a un ancho de banda determinado, debería considerar sólo este tipo de flujos

5.- Gestión del tráfico

A la vista de los resultados se ha propuesto un algoritmo sencillo de control del tráfico basado en el envío de paquetes RST falsos. De forma que si mediante el algoritmo de identificación anterior un flujo TCP es clasificado como eDonkey, inmediatamente se envía un paquete RST utilizando técnicas de spoofing a ambos peers abortando la sesión. Esto permite implementar fácilmente, y con un coste relativamente bajo, gestores de ancho de banda que permitan eliminar el tráfico eDonkey. En la Figura 2 se muestra el efecto sobre el tráfico en la red CTnet tras la aplicación de este algoritmo de control. En menos de diez minutos, el tráfico eDonkey se reduce a un valor residual de 50Kbps en cada dirección correspondiente a los reintentos de los peers.

FIG. 2: EFECTO DE LA APLICACIÓN DEL MECANISMO DE CONTROL DEL TRÁFICO EDONKEY



6.- Conclusiones

En este artículo se ha presentado un estudio y análisis del protocolo eDonkey, aportando un algoritmo implementable mediante herramientas sencillas para minimizar el tráfico y permitir que no sea transportado por la red. Esto es especialmente útil para redes institucionales donde se quiera filtrar el tráfico y evitar problemas legales derivados de la piratería asociada a estas aplicaciones.

Para terminar, es necesario tomar conciencia de la necesidad de realizar capturas periódicas del tráfico para análisis similares al presentado dado la evolución continua y rápida de los protocolos P2P, capaces de incorporar funcionalidades que dificulten su identificación y control.

Referencias

[1] THOMAS KARAGIANNIS, ANDRE BOIDO, Nevil Brownlee, K. Claffy, Michalis Faloutsos, Q. "A Characterization of P2P Traffic in the Backbone". Consultado en: http://www.cs.ucr.edu/~tkarag/papers/tech.pdf. 2003.

[2] Myung-Sup Kim, Hum-Jung Kang, James W.Hong. "Towards Peer-to-Peer Traffic Analysis Using Flows". Presentada en: DSOM. 2003. Páginas 55-67 Volumen 2867/2004.

[3] Kurt Tutschku. "A Measurement-Based Traffic Profile of eDonkey File sharing Service". Presentada en: Passive&Active Measurement Workshop. 2004.

A. M. Guirado-Puerta
(antonio.guirado@f-integra.org)
Fundación Integra
J. Malgosa-Sanahuja
(josem.malgosa@upct.es)
Departamento TIC
Universidad Politécnica de Cartagena
M. Escudero-Sánchez
(manuel.escudero@f-integra.org)
Fundación Integra

Se ha presentado un estudio y análisis del protocolo eDonkey, aportando un algoritmo implementable mediante herramientas sencillas para minimizar el tráfico y permitir que no sea transportado por la red