Análisis del tráfico en una red troncal y en una red de acceso

Jordi Domingo i Pascual
Universitat Politècnica de Catalunya.
Departament d'Arquitectura de Computadors.
Grup de Comunicacions de Banda Ampla.
Campus Nord. Mòdul D6.
Jordi Girona 1-3. 08034 Barcelona.
jordi.domingo@ac.upc.es

Resumen
A medida que crece el número de usuarios de Internet el volumen de tráfico cursado y la complejidad de interconexión de la red aumenta con un ritmo de crecimiento muy superior al esperado. Este incremento de tráfico en la red provoca una serie de problemas a la hora de enrutar los paquetes IP de forma rápida y eficiente. Se está aumentando las capacidades de los enlaces entre los Routers y, esto, a su vez, provoca que la capacidad de los mismos sea insuficiente. Mientras se están poniendo en marcha los Routers de alta capacidad (denominados Gigarouters) la solución más frecuente es disponer de varios Routers para absorber el tráfico de los enlaces de alta velocidad. Esta situación aumenta la complejidad de la gestión de los propios Routers.
Por otro lado, las perspectivas, más o menos inminentes, de hacer negocio a través de Internet requiere un análisis profundo de las características del tráfico y del comportamiento de los usuarios desde el punto de vista de utilización de la red. A la hora de valorar económicamente los distintos servicios es necesario conocer en detalle los recursos asociados a cada uno de ellos y la utilización del mismo por parte del usuario final.
Como se presenta en las conclusiones, el análisis del tráfico se puede realizar a distintos niveles (nivel físico, paquetes IP, servicios), y presenta características completamente dispares en la red troncal, donde el tráfico corresponde a la agregación de muchos flujos, y en las redes de acceso.

Introducción
El primer problema a resolver cuando se desea analizar qué pasa en una red troncal de alta velocidad es disponer de un equipo de medida capaz de procesar adecuadamente el gran volumen de tráfico. Cuando se trata de una red troncal de alta velocidad ATM (modo de transferencia asíncrono) los equipos de monitorización y captura de tráfico ATM son muy complejos y caros. Además, debido al gran volumen de tráfico en un enlace (hasta 150 Mbps.) condiciona la metodología de análisis; no es posible capturar todo el tráfico que circula por un enlace de forma continua. De hecho, tampoco es éste el objetivo principal del estudio. Una captura basada en muestras es suficiente para determinar las características principales del tráfico, siempre y cuando se verifique la validez estadística de las muestras.
En las redes de acceso, donde el volumen de tráfico es inferior, en algunos casos se puede abordar la captura continua del tráfico, aunque normalmente se sigue la misma metodología utilizando capturas de muestras.

En este artículo se presenta una recopilación de medidas del tráfico real en Internet, tanto en las redes troncales como en las redes de acceso. Una parte de las medidas y conclusiones generales está recogida de artículos y estadísticas públicas de las redes troncales de Internet en EUA. El análisis más detallado corresponde al trabajo desarrollado en sendos proyectos de investigación financiados por la Comisión Interministerial de Ciencia y Tecnología (CICYT) en los que ha participado el grupo de Comunicaciones de Banda Ancha de la UPC en colaboración con la red académica española (RedIRIS).
Hay que destacar que tanto las características del tráfico en la red troncal de Internet en EUA como el análisis por servicios presenta una concordancia completa con las medidas y el análisis realizado en la red académica española. Las conclusiones generales son, pues, muy parecidas. Como primera conclusión se puede considerar que el tráfico que circula en la red académica representa satisfactoriamente el tráfico agregado en una red troncal de alta velocidad.
Sin embargo, esta conclusión no puede extrapolarse a las redes de acceso ni al comportamiento de los usuarios finales y el uso que estos hacen de los distintos servicios. En primer lugar no están disponibles públicamente estas medidas ya que corresponden a estudios propios de los distintos proveedores de

acceso a Internet (ISP). En segundo lugar, el perfil de los usuarios o conjunto de usuarios es muy diverso y requiere un análisis más detallado si se desea llegar a conclusiones generales.

El contenido del artículo se estructura de la forma siguiente. En primer lugar se presenta un resumen de varios estudios previos sobre el tráfico en una red troncal y que sirve como punto de partida del estudio realizado. Sigue una breve descripción de la red sobre la que se ha realizado el estudio y de la infraestructura de medidas que se ha utilizado. A continuación se describe la metodología empleada y se presentan los resultados obtenidos. A partir del análisis de estos resultados se presentan las conclusiones.

Estudios previos

Desde la puesta en marcha de la NSFNET (1988-1995) como red troncal de Internet en los EUA se ha estado realizando medidas del tráfico en la red y de su evolución. En 1992 el tráfico dominante era FTP, seguido de SMTP (correo electrónico) y NNTP (distribución de noticias "news"). El tamaño medio de los paquetes IP era de 186 octetos.

En 1994 se registra un crecimiento espectacular en el volumen de tráfico en la red. Aplicaciones como Gopher y Web aparecen con un porcentaje importante del tráfico global.

En el informe correspondiente al mes de abril de 1995 la distribución del tráfico por servicios es la siguiente: WWW (21%), datos FTP (14%), NNTP (8%), TELNET (8%), SMTP (6%), IP (6%), DNS (5%), IRC (2%), Gopher (2%), control de FTP (1%), y otros (27%).

A partir de la desaparición de la NSFNET en 1995, varios operadores comerciales se hacen cargo de la explotación del conjunto de redes troncales en EUA. Actualmente son estas compañías las primeras interesadas en seguir de cerca la evolución del tráfico y detectar cuanto antes cambios significativos, si se producen.

Este mismo interés es compartido no sólo por las empresas de telecomunicaciones sino por cualquier entidad que tenga a su cargo la gestión de una red, ya sea troncal o de acceso a Internet. Es muy importante disponer de un análisis detallado del tráfico y de su evolución; sobretodo si la aparición de una nueva aplicación modifica los hábitos de trabajo, o el uso que se hace de la red, por parte de los usuarios finales.

Precisamente en este punto radica el interés del trabajo realizado y que se expone en este artículo. El primer objetivo ha sido desarrollar una metodología capaz de aportar la información necesaria para analizar correctamente el estado actual de la red y seguir su evolución día a día.

Descripción de la infraestructura de la red

La red académica española (RedIRIS) ha realizado una importante actualización de la infraestructura de comunicaciones en la red troncal que conecta los distintos puntos de acceso en las comunidades autónomas con el nodo central. Como en muchas otras redes troncales de alta velocidad la tecnología utilizada para el transporte físico es SHD (la Jerarquía Digital Síncrona) con nodos de conmutación y de multiplexación ATM. Los 17 nodos de acceso están conectados a través de la red pública ATM GigaCom de Telefónica. Un proceso de migración tecnológica similar se ha realizado en las redes académicas de otros países europeos. A su vez, las distintas redes académicas de Europa están interconectadas por una red ATM trans-europea.

En la actualidad todo el tráfico en la red es tráfico IP superpuesto sobre ATM siguiendo los estándares internacionales (RFC 1577 y RFC 1483). Los paquetes IP son encapsulados en las PDU AAL5.

Las redes de acceso en cada Comunidad Autónoma presentan distintas configuraciones y utilizan distintas tecnologías de transmisión de datos. En cada punto de acceso se dispone de un Router conectado a la red troncal. De todas formas, el tráfico sigue siendo 100% IP y la mayoría de redes locales conectadas son Ethernet (CSMA/CD).

Metodología

Para realizar la captura del tráfico ATM se ha de utilizar un equipo comercial de altas prestaciones para la monitorización y captura de los flujos ATM ya que todavía no existen equipos de bajo coste específicos para este tipo de estudios. Mediante "splitters" se obtiene una réplica del flujo de celdas ATM en cada uno de los sentidos de transmisión.

Se considera muy importante partir de la captura de los flujos de celdas ATM directamente de los enlaces troncales ya que, si bien actualmente todo el tráfico es IP sobre ATM (sobre AAL5), en un

futuro puede ser interesante poder analizar otros protocolos. Además, una análisis del tráfico ATM permite determinar si se producen pérdidas de celdas ATM (tramas AAL5 con error de ensamblaje) y estudiar el retardo medio y máximo entre paquetes de una misma conexión ATM.

Como ya se ha comentado, el primer obstáculo es que los equipos de medida están diseñados para soportar muchas funciones pero o son capaces de capturar y procesar tráfico de forma continua. Por lo tanto, la primera fase consiste en verificar que la captura no continua tiene validez estadística. Dependiendo de las opciones disponibles en el equipo de medida se captura toda la trama AAL5 o bien sólo la primera celda ATM de cada trama AAL5.

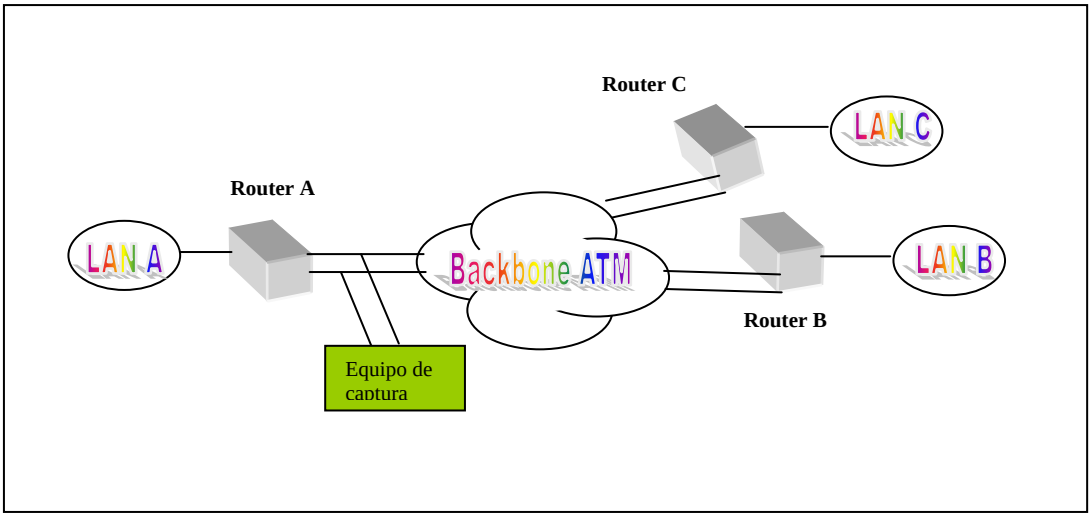

Figura 1. Captura de tráfico ATM.

La primera celda ATM de la trama AAL5 contiene la cabecera IP completa y la cabecera TCP/UDP. Esta información es suficiente para el análisis del tráfico propuesto.
Si bien es interesante analizar los protocolos de transporte asociados al propio protocolo IP, como el ICMP, el estudio se centra en los protocolos de transporte TCP y UDP. El diseño contempla la posibilidad de analizar IP v6 y los nuevos protocolos que se están definiendo.
Volviendo a la captura de los flujos de celdas ATM, los datos deben ser volcados de forma rápida a disco, en otra máquina, donde se realiza el primer filtrado de la información para reducir el volumen total de datos a almacenar. El volcado directo de la captura correspondiente a unos cuantos segundos produce ficheros de varios Goctetos; tras el primer filtrado los ficheros son de unos pocos Koctetos.
La información que se registra para cada paquete IP es la siguiente: direcciones IP fuente y destino, puertos fuente y destino, y protocolo.

El estudio realizado se ha dividido en dos partes. En primer lugar se desea caracterizar el tráfico IP sobre ATM globalmente. Para ello, se almacena la longitud de los paquetes IP, el número de paquetes, el tipo de protocolo, y se evalúa la sobrecarga correspondiente al transporte de IP sobre ATM. Estos parámetros son importantes a la hora de dimensionar la red, los routers y otros equipos de comunicaciones de la red de acceso.
La segunda parte del estudio se centra en la caracterización de las aplicaciones, siguiendo su evolución a lo largo del día, acumulando las estadísticas a lo largo de la semana, etc., y analizando el tráfico con un filtrado previo del grupo o comunidad bajo estudio. De este estudio se pueden extraer perfiles de usuarios de la red y perfiles de utilización de la red en función del día de la semana o de la zona horaria del día.
A continuación se presentan los resultados de las medidas realizadas y el análisis e interpretación de las mismas.

Tráfico IP sobre ATM

En una primera fase se identifican individualmente servicios más utilizados e identificados (http, ftp, telnet, smtp, nntp, games, irc, e IP sobre IP. El conjunto de aplicaciones con puertos menores de 1024 se identifica como "otros" y el tráfico correspondiente a aplicaciones cuyo puerto es superior a 1024 se identifica como "desconocido".

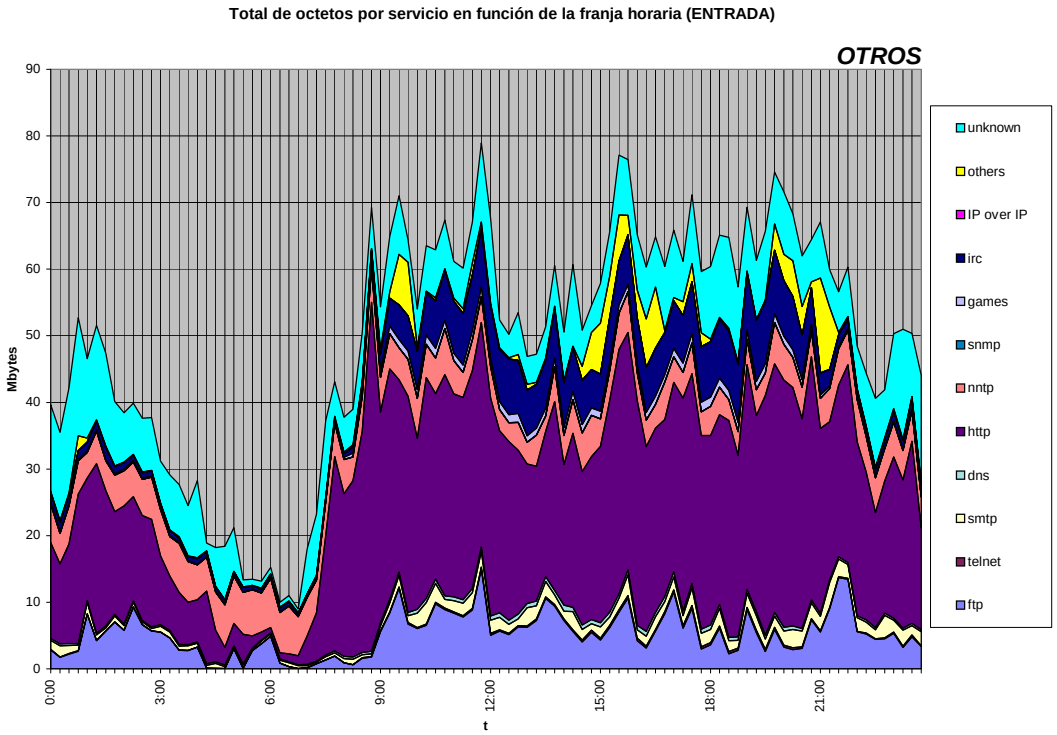

Figura 2. Flujo total de octetos clasificado por aplicaciones (flujo de entrada).

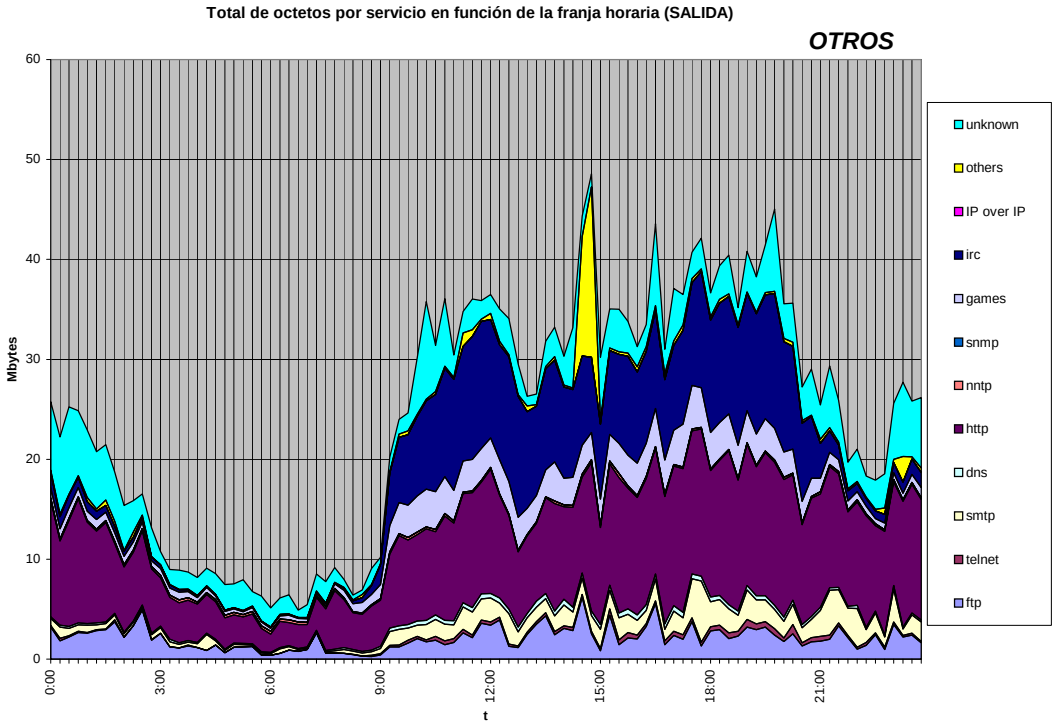

Figura 3. Flujo total de octetos clasificado por aplicaciones (flujo de salida).

Evidentemente, en un estudio posterior se analiza con más detalle tanto el conjunto de otras aplicaciones como las clasificadas como desconocidas. Las figuras 2 y 3 presentan la evolución a lo largo de las 24 horas del volumen total de paquetes IP (en octetos) en ambos sentidos y clasificados por aplicaciones.

Comparando las figuras 2 y 3 se puede observar claramente que el volumen de información de entrada es mucho mayor que el de salida. Esto es válido tanto para el tráfico global internacional (con un volumen mucho mayor) como para cada enlace de la red troncal con las comunidades autónomas.

Otra medida interesante es la longitud media de los paquetes IP por aplicación. La Figura 4 muestra de una forma más evidente el carácter "importador" de información de los servicios ftp, http y nntp.

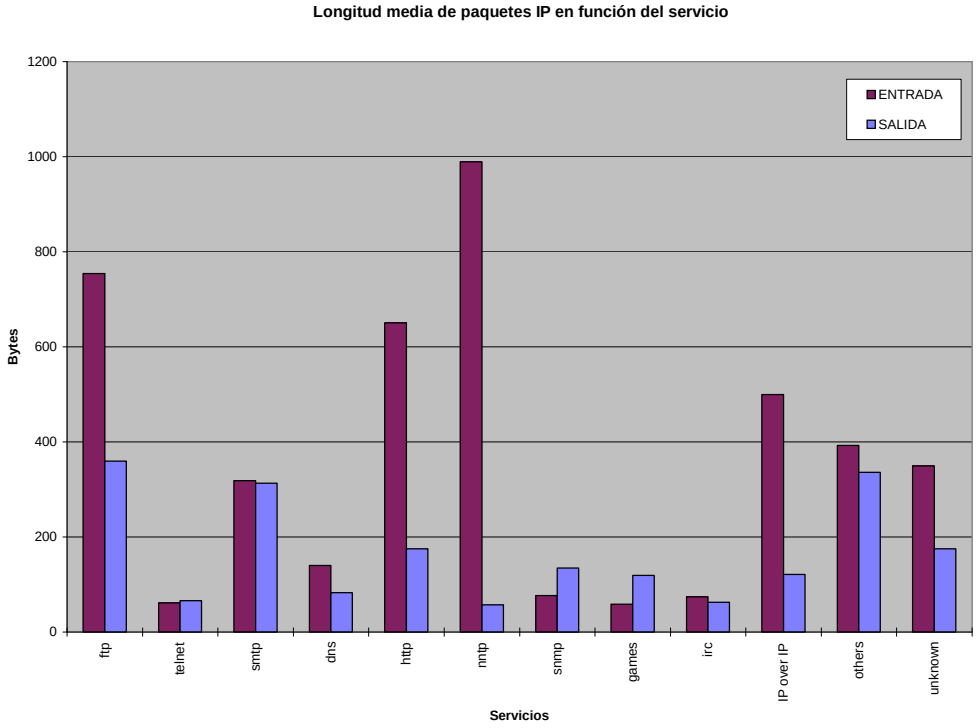

Figura 4. Longitud media de los paquete IP en función de la aplicación.

Desde el punto de vista de una red troncal ATM es importante determinar el rendimiento global de la misma cuando se transporta IP sobre ATM de forma directa. Por este motivo se analiza la distribución del tamaño de los paquetes IP en unidades equivalentes a celda ATM para ver cuantas celdas ATM corresponden, en media, a cada paquete IP. La Figura 5 muestra esta distribución. Es de destacar que más del 70% de los paquetes de salida corresponden a 1 o 2 celdas ATM, y que un 45% de los paquetes de entrada también tienen menos de tres celdas ATM. Otro punto a destacar es que prácticamente no hay ningún paquete IP con longitud superior a los 1500 octetos de la MTU de Ethernet.

Asimismo, en la Figura 6 se muestra el porcentaje de sobrecarga por el hecho de transportar paquetes IP sobre ATM. La parte inferior representa el porcentaje de tráfico útil, mientras que la superior representa la sobrecarga de las cabeceras asociadas.

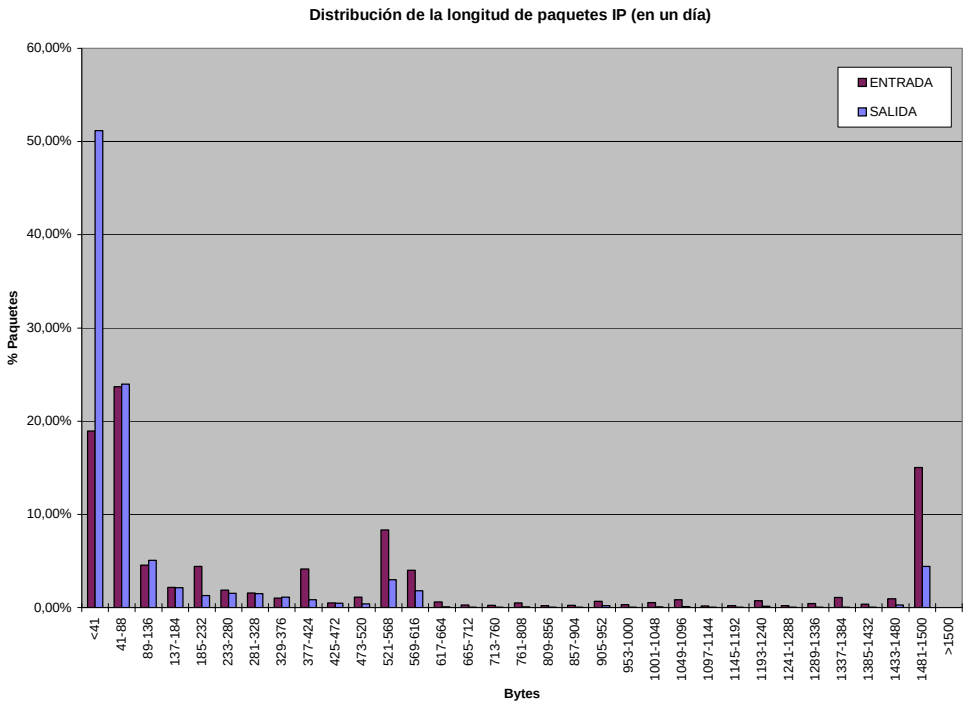

Figura 5. Distribución del tamaño de los paquetes IP.

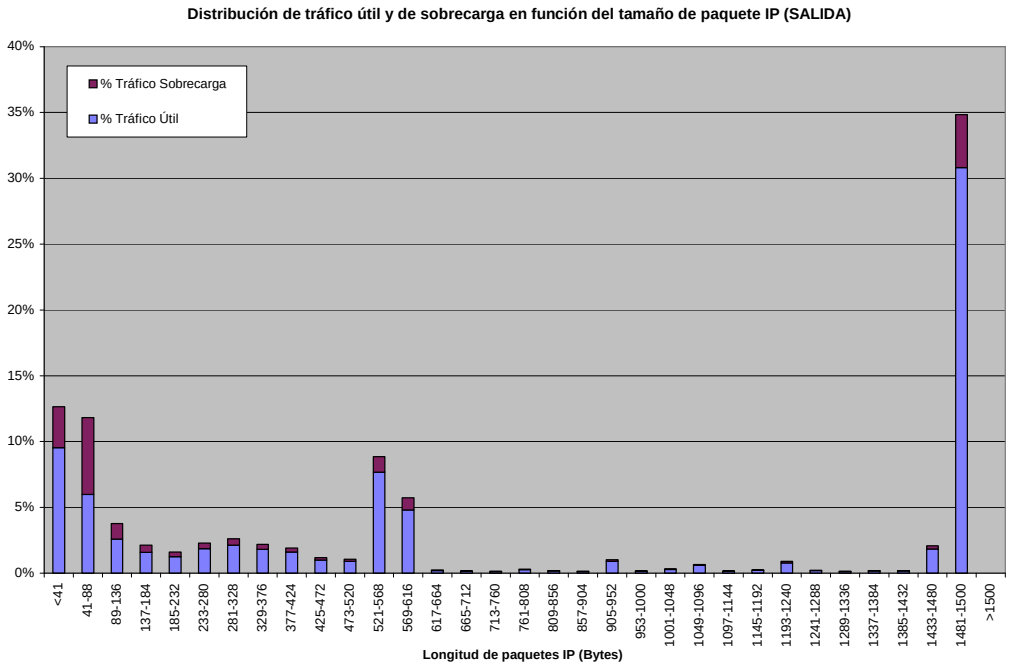

Figura 6. Sobrecarga debido a las cabeceras.

Análisis del tráfico en la red de acceso

La misma metodología de captura de tráfico y postprocesado de la información se ha aplicado a distintos grupos de usuarios. La identificación de grupos de usuarios se ha realizado en función de

grupos de direcciones IP. Evidentemente, un estudio más profundo requiere detallar con más precisión los grupos de usuarios o servidores de información de la red para poder extraer conclusiones más útiles. A modo de ejemplo, se presentan las características del flujo total de información por grupos de usuarios.

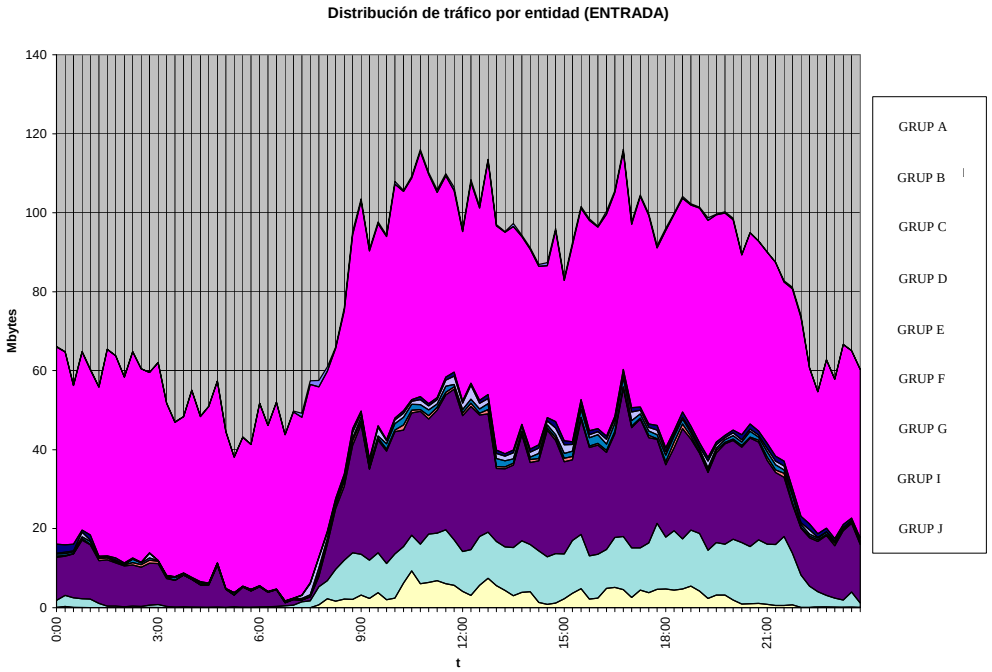

Figura 7. Evolución del tráfico global por grupos de usuarios (flujo de entrada).

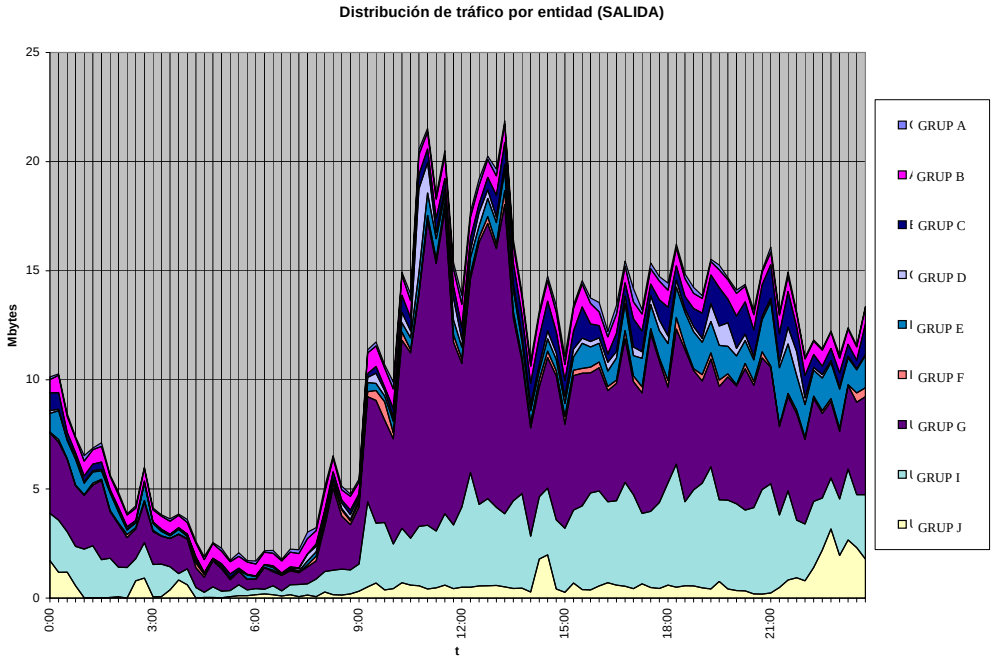

Figura 8. Evolución del tráfico global por grupos de usuarios (flujo de salida).

Conclusiones

Si bien por motivos de espacio no se han incluido gráficas correspondientes a la red troncal de Internet en los EUA (que se pueden encontrar en el artículo mencionado en la bibliografía), el comportamiento global en la red troncal es semejante. Es de destacar la gran cantidad de paquetes IP de longitud inferior a los 100 octetos y el carácter importador de información de la red académica española.

Por los que se refiere al análisis de las características de tráfico de grupos de usuarios se constata una gran variabilidad en el comportamiento de la carga de la red de acceso. El comportamiento aislado de un solo usuario durante un periodo de tiempo considerable puede modificar significativamente las estadísticas del grupo. Este efecto no es apreciable en la red troncal debido a la agregación de gran cantidad de flujos. De hecho, a medida que se intenta localizar un grupo de usuarios más reducido esta variabilidad en el comportamiento es mucho más acusada.

De cara al futuro inmediato es preciso disponer de equipos de análisis de tráfico a precios asequibles para poder hacer un seguimiento de la evolución del tráfico, especialmente en las redes de acceso, donde la influencia de un usuario o de un grupo de usuarios puede modificar sensiblemente la utilización de los recursos de la red.

Agradecimientos

El trabajo presentado se ha desarrollado dentro de los proyectos CASTBA-C (TEL96-2509-E) y MEHARI-C (TEL97-1897-E) financiados por la CICYT. El trabajo ha sido realizado conjuntamente con Josep Solé Pareta, Xavier Martínez Álvarez y Carlos Veciana Nogués. Ha colaborado activamente en ambos proyectos el grupo del Departamento de Ingeniería de Sistemas Telemáticos (DIT) de la Universidad Politécnica de Madrid (UPM) y RedIRIS, como organismo gestor de la Red Académica Española.

Referencias

"Classical IP and ARP over ATM". RFC 1577. January 1994.

"Multiprotocol Encapsulation over ATM Adaptation Layer 5". RFC 1483. July 1983.

"Wide-Area Internet Traffic Patterns and Characteristics". K. Thompson, G. J. Miller, R. Wilder. IEEE Network Magazine, November / December 1997, Vol. 11, No. 6, pp. 10-23.
http://www.vbns.net/presentations/papers/MCI/traffic.ps.gz

"Growth Trends in Wide-Area TCP Connections". V. Paxon. IEEE Network, July / August 1994, Vol. 8, No 4, pp. 8-17.

"NSFNET Backbone Statistics". April 1995.
http://www.cc.gatech.edu/gvu/stats/NSF/merit.html.

Measurements at the Federal Interchange Point FIX-West. The National Laboratory for Applied Network Research (NLANR).
http://www.nlanr.net/NA/FIX/Stats/West/Index.html

RedIRIS. http://www.rediris.es

"CASTBA: Medidas de tráfico sobre la Red Académica Española de Banda Ancha". M. Alvarez-Campana, A. Azcorra, J. Berrocal, J. Domingo, D. Larrabeiti, X. Martínez, J.I. Moreno, J. R. Pérez, J. Solé. Pendiente de publicación.