Análisis del uso de los servicios Internet a partir del sistema MEHARI: Prueba de campo sobre RedIRIS

Pedro J. Lizcano, Comisión Interministerial de Ciencia y Tecnología (CICYT),
Rosario Pino, 14-16, 28020 Madrid, Spain

Arturo Azcorra, Universidad Carlos III de Madrid (UC3M),
Butarque 15, 28911 Leganés (Madrid), Spain

Josep Solé-Pareta y Jordi Domingo-Pascual, Universitat Politècnica de Catalunya (UPC),
Jordi Girona 1-3, 08034 Barcelona, Spain

Manuel Alvarez-Campana, Universidad Politécnica de Madrid (UPM),
ETSI Telecomunicación, Ciudad Universitaria, 28040 Madrid, Spain

Resumen.

Este artículo describe el sistema MEHARI cuya función es la monitorización y análisis de tráfico sobre Redes Nacional de I+D (NRN). El principal requerimiento en el diseño del sistema MEHARI fue abordar los principales desafíos, que generalmente aparecen en la mayoría de estas redes: Uso de Servicios, esquemas de tarificación, dimensionado de red y procedimientos de auditoría acorde con una Política Aceptable de Uso (AUP) dada. El sistema MEHARI consiste en un hardware de captura de tráfico de bajo coste, y varios módulos software de análisis de tráfico ejecutándose sobre esta plataforma. Estos módulos pueden generar informes sobre el uso de los servicios Internet, los principales orígenes y destinos del tráfico, etc. El sistema MEHARI ha sido probado en el marco de la Red Académica Española (RedIRIS). En este documento se presentan algunos de los datos más relevantes sobre las prestaciones y eficiencia del sistema utilizado en las pruebas de campo.

Palabras clave: Monitorización de Servicios Internet, Análisis de tráfico de Internet, Redes Nacionales de I+D, Auditoría de Políticas de Uso Aceptable

1. Introducción

El desarrollo completo del paradigma de la *Sociedad de la Información* ha liderado el desarrollo de nuevas infraestructuras IP por todo el mundo, conducido por intereses comerciales y por diferentes iniciativas gubernamentales motivadas por razones estratégicas. Incluso la UE ha jugado un papel de líder en este asunto mediante las iniciativas *TEN-34* y *TEN-155*, con el objetivo de fomentar la colaboración internacional entre la investigación realizada en las diferentes Redes Nacionales de I+D (NRN).

Pero estos desarrollos hacen frente a dos grandes problemas:

- En la mayoría de casos las NRN están financiadas con dinero publico, con presupuestos que no pueden sufragar el continuo crecimiento de la demanda de servicio de los usuarios de las NRN. Por lo tanto, se hace necesario un nuevo modelo de financiación, el cual se dirija a un acercamiento al uso responsable.

- Intereses comerciales, en los actuales escenarios totalmente competitivos, exigen que estas redes de financiación pública tengan una política de uso aceptable (AUP) para asegurarse que no influyan en el emergente mercado basado en los servicios Internet

Para cubrir estas dos líneas abiertas, herramientas específicas de auditoría de AUP son necesarias para monitorizar el uso real de las redes de financiación pública. Las posibles colaboraciones entre las redes Europeas de NRNs y sus homólogas en Norte América (p.e. Internet2 y Canarie) dependen de tener estas herramientas listas para asegurar la coherencia entre las AUP Americanas y Europeas concernientes al tráfico que se intercambia entre ellas.

Además, el antes mencionado completo desarrollo de la Sociedad de la Información también complica el dimensionado, la administración y la operación de los troncales de Internet. Estas tareas no pueden llevarse a cabo satisfactoriamente sin un conocimiento preciso y actualizado sobre lo que está sucediendo en la red. De este modo, hay dos dimensiones diferentes para este problema: la primera es la perspectiva del tráfico, y la otra es desde el punto de vista del motivo de utilización. El primero puede necesitar el análisis de contenidos en el caso de que la identificación por puerto no sea posible, y los resultados son razonablemente objetivos. El segundo indudablemente

necesita el análisis de contenidos y sus resultados están basados en heurísticos y técnicas subjetivos.

Respecto al análisis del tráfico, merece la pena mencionar algunos estudios recientes para la caracterización del tráfico de Internet [1,2]. Las muestras de tráfico obtenidas en estos estudios han revelado algunos resultados interesantes sobre los patrones de tráfico de Internet sobre troncales IP/ATM [3]. La principal conclusión es, sin embargo que el tráfico Internet es altamente variable e impredecible, y por lo tanto los resultados del análisis de tráfico son sólo válidos a corto plazo.

Respecto al motivo de utilización, diferentes técnicas heurísticas pueden ser combinadas acorde con la naturaleza de la clasificación del tráfico deseada. Como el análisis de contenido es costoso en términos de CPU, se debe establecer una relación de compromiso entre la seguridad en la objetividad de los resultados y la potencia de procesado requerida. Es recomendable que los heurísticos sean comprobados manualmente periódicamente a fin de hacer el análisis más ajustado y relajado dependiendo de las desviaciones entre el tráfico real y el diagnóstico llevado a cabo por la herramienta.

Este artículo describe el diseño del sistema MEHARI, un procesador de tráfico de altas prestaciones para el análisis de las cabeceras y los contenidos de tráfico Internet. Las características del sistema son la modularidad, escalabilidad, altas prestaciones, bajo coste, flexibilidad y adaptabilidad. El hardware MEHARI está basado en componentes PC de bajo coste, y las aplicaciones de procesado pueden funcionar en cualquier máquina UNIX. La plataforma del MEHARI consiste en PCs con tarjetas STM-1 para la captura del tráfico y el pre-procesado. El tráfico procesado es suministrado desde la plataforma a los módulos software de análisis de tráfico MEHARI. La captura de MEHARI y la capacidad de procesado se pueden ampliar en módulos incrementando el número de elementos hardware en la configuración del sistema. Las funcionalidades de análisis de MEHARI pueden ser ampliadas o modificadas ya sea, cambiando la configuración o añadiendo nuevos módulos software de análisis. Tres ejemplos de tipos de información que pueden ser obtenidos con los módulos de análisis implementados actualmente son los siguientes: 1) una clasificación heurística del tráfico en las siguientes categorías: académico, comercial, ocio e indeterminado; 2) una clasificación del tráfico en función del origen y del destino; y 3) La verificación de la asignación de puerto de aplicación. El sistema que incluye estos tres módulos ha sido probado, y los resultados de las pruebas también son descritos en este articulo, en términos de funcionalidad y prestaciones.

Los datos provenientes de MEHARI pueden servir para diferentes propósitos, algunos de los cuales son:

- Ingeniería de red para ajustar la capacidad a la matriz de tráfico y los perfiles horarios.
- Caracterización y análisis del tráfico de usuario (o grupo de usuarios).
- Identificación de los servicios de información, servidores de información y clientes disponibles en la red.
- Clasificación del tráfico de la red por aplicación y también en motivo de utilización.
- Validación de la política de uso apropiada.
- Facturación ("billing") y tarificación ("charging").
- Detección y identificación de las amenazas de seguridad y ataques [4].

El resto del documento está estructurado de la siguiente forma. El apartado siguiente presenta la arquitectura funcional del sistema MEHARI. El apartado 4 describe los módulos de aplicaciones actualmente soportados por el sistema y el razonamiento que hay tras su desarrollo. El apartado 5 nos muestra algunos ejemplos de resultados obtenido en las pruebas realizados con el sistema MEHARI en un escenario real: la Red Académica Española (RedIRIS). Finalmente el apartado 6 resume las principales conclusiones de nuestro trabajo.

2. La Arquitectura Funcional de MEHARI

La figura 1 muestra la arquitectura funcional del sistema MEHARI, que consiste en los siguientes subsistemas:

- *Subsistema de Captura de Tráfico (SSCT):* Este bloque funcional es responsable de capturar las muestras de tráfico, las cuales son posteriormente analizadas por otros bloques del sistema MEHARI. EL SSCT puede ser configurado para capturar cualquier celda ATM dentro de una lista de pares VPI/VCI o en modo promiscuo (todos los pares VPI/VCI). Las celdas son reensambladas en tramas AAL5, las cuales son periódicamente volcadas a disco para procesado posterior. Dependiendo del tipo de análisis que se realice, el usuario puede escoger el volcado de toda la trama AAL5 o únicamente una parte de ella (p.e. los primeros 48 octetos).

  Se ha de hacer notar que, a pesar de que el sistema MEHARI fue inicialmente concebido para monitorizar tráfico IP/ATM, el SSCT no hace ninguna suposición acerca el contenido de la trama AAL5. De esta forma, el sistema se puede utilizar en un futuro para analizar otros protocolos encapsulados sobre ATM.

  El SSCT ha sido diseñado de manera que puede ser físicamente implementado en una o varias máquinas

de captura, permitiendo de este modo que la relación de captura aumente si es necesario. La plataforma de captura es un PC estándar con FreeBSD. La captura propiamente dicha es realizada en dos tarjetas de red con interfaz ATM Fore (una para cada sentido de transmisión) utilizando un firmware especial (OC·MON [3]). Debido a la utilización de componentes hardware estándar, el coste total del sistema es bastante bajo (aproximadamente 6,000 euros por plataforma de captura).

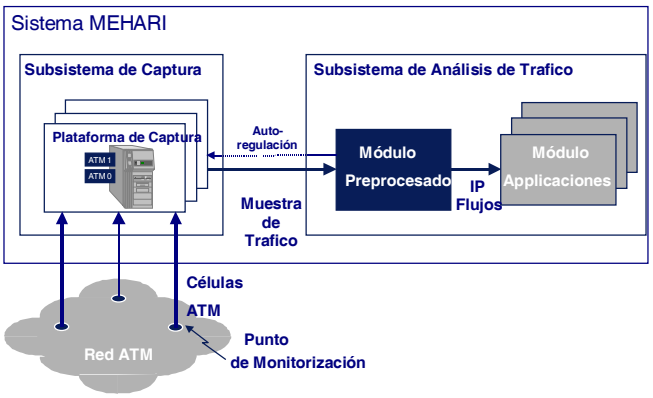

Figura 1. Arquitectura Funcional del Sistema MEHARI.

- *Subsistema de Análisis de Tráfico (SSAT):* Este bloque funcional es el responsable del análisis de las muestras generadas por el subsistema SSCT. El SSAT contiene el Módulo de Pre-procesado (MPP) y los Módulos de Aplicaciones (MAPs).

  Debido al elevado volumen de tráfico que puede ser transportado sobre los enlaces de ATM monitorizados, el SSAT debe descartar las muestras de tráfico tan pronto como sea posible. Ésta es la razón, precisamente, de la introducción del módulo MPP, el cual pre-procesa las muestras de tráfico en tiempo real, de manera que los ficheros de captura son borrados una vez se han extraído los parámetros de interés.

  Nótese que la tasa a la que las muestras de tráfico son pre-procesadas es crucial para el funcionamiento global del sistema. En el proyecto MEHARI nos concentramos en el análisis de flujos IP (ver Apartado 3.1), lo que nos dirigió a desarrollar un pre-procesado intensivo en CPU de las muestras de tráfico. Otras aplicaciones del sistema MEHARI probablemente requerirán menos capacidad de procesado. En cualquier caso, el subsistema SSAT fue diseñado para que pudiera ser implementado físicamente en una o varias máquinas. Este enfoque nos permite incrementar la capacidad de procesado tanto como se desee añadiendo más PCs o estaciones de trabajo. En cualquier caso, nótese la existencia de un mecanismo auto-regulador, el cual adapta la relación de captura a la capacidad de procesado del SSAT.

  El análisis de tráfico realmente es llevado a cabo por los MAPs. Diferentes tipos de MAPs, correspondiendo a diferentes tipos de análisis, pueden estar presentes en el SSAT. El apartado 3 describe algunos de los MAPs desarrollados para el sistema MEHARI. Cabe destacar que el sistema MEHARI ha sido diseñado para que cualquier nuevo MAP pueda ser añadido fácilmente. Además, existe la posibilidad de usar varios PCs para llevar a cabo un análisis distribuido, así que la capacidad de proceso se puede ajustar a los requerimientos de los diferentes MAPs considerados.

3. Módulos de Aplicación

En este apartado se describen los Módulos de Aplicaciones (MAPs) específicamente desarrollados para el proyecto MEHARI. Como se ha mencionado en el apartado 2, estas MAPs son solamente un ejemplo de los diferentes tipos de análisis que se pueden llevar a cabo por el sistema MEHARI. Nuestro objetivo principal describiendo los siguientes módulos es mostrar las capacidades del sistema y proveer una muestra de la gran cantidad de posibilidades que se pueden tener en cuenta.

3.1 Módulo de Clasificación del Tráfico

El proyecto MEHARI fue concebido como una continuación de los estudios mediante medidas de tráfico y caracterización de la utilización de red, llevados a cabo en el proyecto CASTBA sobre la Red Académica Española (RedIRIS). Los resultados del proyecto CASTBA revelaron una serie de dificultades en la caracterización del tráfico Internet utilizando los analizadores de tráfico convencionales. Se encontró que un simple análisis basado en protocolo y puertos TCP/UPD no proporcionaba suficiente exactitud [5]. El motivo es la existencia de aplicaciones que utilizan puertos no registrados, o puertos registrados con otras finalidades que las asignadas. Esto dio lugar a un importante porcentaje de tráfico (10-20%) cuya naturaleza no pudo ser determinada o lo que es peor, una indeterminada cantidad de tráfico clasificado en categorías incorrectas. Una de las principales motivaciones del proyecto MEHARI fue precisamente desarrollar una herramienta de monitorización y análisis de tráfico Internet que solucionara estas inseguridades. Lo que empezó como una modesta tentativa para resolver las limitaciones del análisis por puertos TCP/UDP se convirtió en el *Módulo de Clasificación del Tráfico*, el

cual ha aportado una alta efectividad a la caracterización de tráfico y servicios Internet.

La figura 2 muestra la estructura del Módulo de clasificación de Tráfico (MCT). El MCT analiza los datos provenientes del MPP, el cual está basado en el concepto de flujos IP y el reconocimiento de patrones. El MPP lleva a cabo una agregación estadística de todos los paquetes pertenecientes a un mismo flujo IP durante un período de tiempo configurable. Un flujo IP se define como la agregación de los paquetes que tiene el mismo cuádruplo dirección IP origen, puerto TCP/UPD origen, dirección IP destino, puerto TCP/UDP destino. Además, el MPP explora el contenido de los paquetes para tratar de determinar la presencia de patrones específicos, los cuales pueden ser programados por el usuario. Como resultado, el MPP genera periódicamente un fichero que contiene un resumen de los flujos IP detectados y el número de veces que un patrón ha sido detectado en él.

Los flujos IP son posteriormente analizados por el MCT, el cual realiza una correlación de los datos correspondientes al tráfico entrada y salida. El resultado es una colección de flujos IP bidireccionales (biflujos) con una lista de síntomas ponderada para cada sentido. El siguiente paso consiste en aplicar un conjunto de heurísticos, que pueden ser fácilmente configurados por el usuario. La aplicación de estos heurísticos produce que un flujo sea clasificado bajo una categoría de tráfico particular, la cual ha sido previamente definida por el usuario. Por ejemplo, en las pruebas de campo descritas en el apartado 4 de este documento, consideramos cuatro categorías de tráfico llamadas académico, comercial, ocio e indeterminado. Finalmente, el MCT genera un fichero resumen informando el volumen de tráfico dentro de cada una de las categorías consideradas. El apartado 4 muestra un ejemplo del informe de clasificación de tráfico que puede ser obtenido con el MCT.

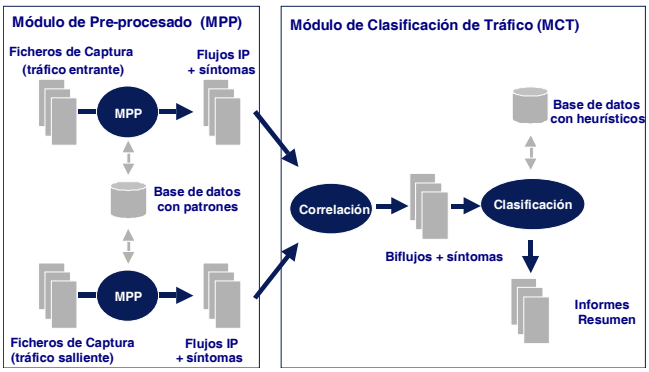

Figura 2. Módulo de clasificación de Tráfico (MCT).

3.2 Módulo Analizador de Orígenes/Destinos del Tráfico

El Módulo de análisis de Orígenes/Destinos de Tráfico (MODT) realiza una clasificación del tráfico basada en el Sistema Autónomo (SA) de origen/destino del flujo bidireccional identificado en el módulo MCT (ver apartado 3.1). El diagrama funcional del módulo de MODT es representado en la figura 3.

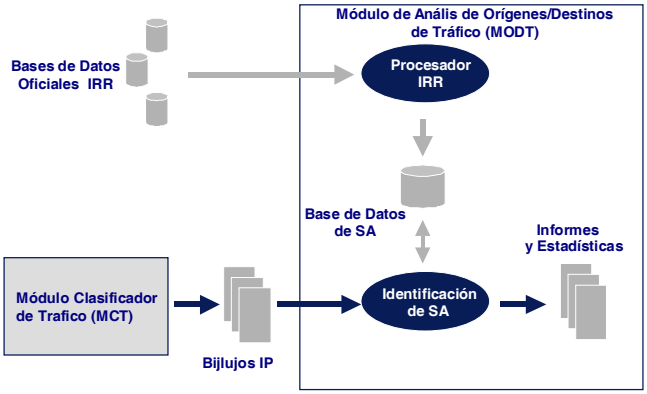

Figura 3. Módulo de Análisis de Orígenes/Destinos de Tráfico (MODT).

Para determinar el SA al cual pertenece una dirección IP origen/destino, el MODT utiliza la información almacenada en las bases de datos del Internet Routing Registry (IRR). Esta información es actualizada periódicamente por el *Bloque Procesador del IRR,* el cual genera una base de datos sobre los SA en local, que es utilizada por el *Bloque Identificador de SA*. Las bases de datos de los SA incluyen el nombre de la organización responsable para cada SA, así como la identificación de las subredes que pertenecen a cada SA.

El Bloque Identificador de SA analiza las direcciones IP buscándolas en la base de datos de SA local. Como resultado, este bloque periódicamente genera un fichero con informes estadísticos sobre el volumen de tráfico intercambiado entre cada subred perteneciente a la red en estudio (p.e. RedIRIS) y los diferentes SAs registrados en las bases de datos del IRR. El módulo de aplicación MODT puede ser utilizado, por ejemplo, para determinar la fracción de tráfico intercambiado con una red académica especifica, con una red comercial, o para obtener la lista de los SAs más visitados. Otras posibles aplicaciones del módulo MODT podrían ser las siguientes:

- *Evaluación del uso de enlaces externos:* El sistema produce una serie de informes mostrando el uso de recursos por cada subred perteneciente a la red

académica monitorizada. Esta información es esencial para un futuro rediseño de la red, para los esquemas de tarificación, etc.

- *Tarificación y facturación por el uso de los recursos de la red*: La distribución del tráfico según subredes y SA permite soportar modelos de tarificación y facturación basados en el destino donde la tarifa depende las direcciones IP origen y destino, de la mascara, del número de sistema autónomo o de la cadena (route).

3.3 Módulo de Análisis de la Cabeceras Internet

El Módulo de Análisis de las Cabeceras Internet (MCI) produce una clasificación del tráfico basada en el análisis de las cabeceras de los protocolos Internet: la cabecera del paquete IP, la cabecera del segmento TCP/UDP, y la cabecera de PDU de aplicación. La estructura del módulo MCI se muestra en la figura 4.

El principal objetivo del MCI es diferenciar tantos flujos de tráfico (secuencias de paquetes con el mismo origen y destino) como sea posible, y luego comprobar la coherencia entre los puertos TCP/UDP y las aplicaciones para cada flujo. Se diferencia entre el tráfico que se ajusta a estándares (el puerto TCP o UDP corresponde a una asignación estándar) y el tráfico que no. Se obtienen estadísticas de los servidores locales más visitados (dentro de la red académica) y remotos (fuera de la red académica).

La clasificación es el resultado de la verificación de los servicios contenidos en el tráfico cursado. Esta verificación se realiza comprobando la coherencia entre los puertos TCP/UDP y las cabeceras de las PDU de aplicación. Para poder conseguir esta verificación, se desarrollaron unos diccionarios de los principales servicios "well-known", conteniendo una serie de patrones extraídos de cada una de las especificaciones de los servicios. La verificación esta orientada al servicio, la cual proporciona un mejor funcionamiento y un menor consumo de recursos de la máquina. Algunas posibles aplicaciones del módulo de MCI pueden ser la tarificación, auditoría del uso de la red, verificación de los servicios cursados, grabación del tráfico desconocido/inusual, etc.

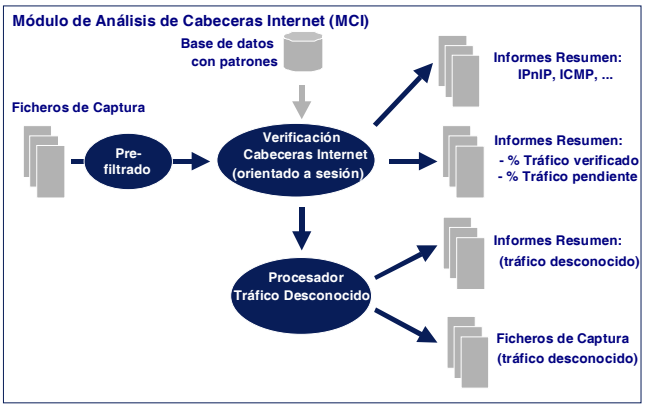

Figura 4. Módulo de Análisis de Cabeceras Internet (MCI).

Algunas posibles aplicaciones del módulo MCI podrían ser las siguientes:

- *Relación de Verificación*: La clasificación de tráfico se hace por: verificado (comprobación correcta), pendiente (los patrones no han sido hallados en los diccionarios), desconocido (los puertos no incluidos en los diccionarios), descartado (el protocolo de transporte es diferente a TCP/UPD).

- *Identificación de los principales servidores locales y remotos*: El tráfico verificado es clasificado según sean sus direcciones locales o remotas. Esta clasificación permite destacar los principales servidores remotos y locales, o los servicios que estos suministran.

- *Clasificación del tráfico desconocido*: Algunos de los siguientes heurísticos pueden ser útiles para clasificar los principales servidores y servicios que permanecen desconocidos para los administradores de red: Informes sobre las direcciones remotas y locales ordenadas por volumen de tráfico servido. Informes sobre posibles servidores detectados ordenados por número de posibles clientes conectados. Todos estos informes destacan direcciones, volúmenes, protocolos de transporte y puertos de aplicación que permiten a los auditores del sistema inspeccionar el tráfico y descubrir aplicaciones.

4. Pruebas de campo con el Sistema MEHARI

El sistema MEHARI ha sido probado satisfactoriamente en un entorno de red real: el troncal IP/ATM de la Red Académica Española (RedIRIS). Dos unidades del sistema MEHARI fueron desplegadas en dos diferentes puntos de monitorización del troncal de RedIRIS. Una de las unidades MEHARI fue instalada en el nodo central de RedIRIS en Madrid, según la configuración que se

muestra en la figura 5. La otra unidad MEHARI se ubicó en la red de acceso de Cataluña a RedIRIS, con un entorno de red muy similar al utilizado en el nodo central.

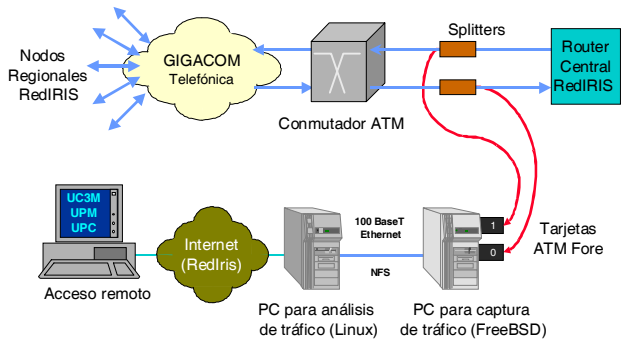

Figura 5. Configuración de la red de pruebas.

Se debe tener en cuenta que los resultados aquí presentados están dirigidos a suministrar algunos ejemplos de las posibilidades de los informes MEHARI. Las figuras incluidas en estos informes no tienen porque corresponder necesariamente a los valores reales de los parámetros de tráfico involucrados en la Red Académica Española. Sin embargo, estos valores pueden corresponder al proceso de monitorización durante un período de tiempo concreto, y por lo tanto no ser plenamente representativos.

La Figura 6 muestra algunos ejemplos de los resultados obtenidos por el Módulo de Clasificación de Tráfico (MCT) en el nodo central en Madrid. Para cada uno de los 17 enlaces, el histograma representa el porcentaje del tráfico de entrada correspondiente a las cuatro categorías consideradas en el proyecto MEHARI (académico, comercial, ocio e indeterminado). Los resultados corresponden a la media obtenida durante un período de captura de aproximadamente cuatro meses y medio (desde el 15 de septiembre de 1998 hasta el 2 de febrero de 1999).

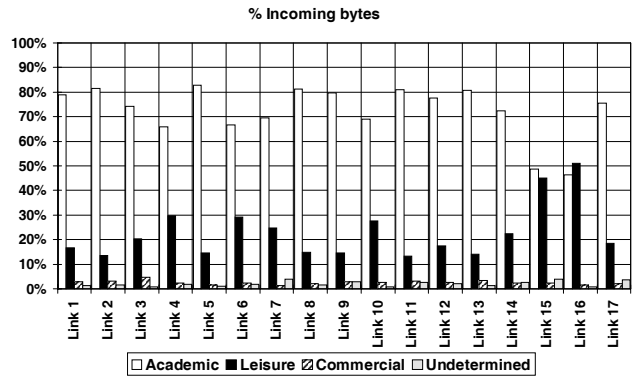

Figura 6. Clasificación del tráfico por uso (por nodos regionales de RedIRIS).

La figura 7 muestra el tráfico medio de entrada de los 17 nodos regionales de RedIRIS acumulados, para el mismo período de captura que la figura 6.

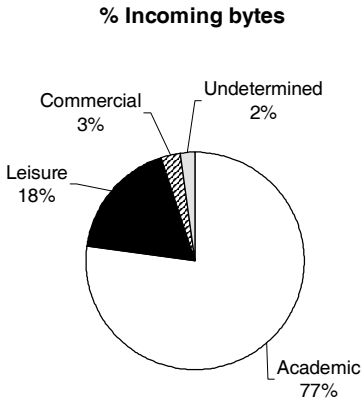

Figura 7. Clasificación del Tráfico por uso (Distribución global).

Un ejemplo de los resultados suministrados por el Módulo de Análisis de Orígenes/destinos de Tráfico (MODT), se puede ver en las figuras de la 8 a la 12. La figura 8 muestra los principales orígenes de tráfico de RedIRIS, obtenidos durante un período de captura de aproximadamente cuatro meses y medio (desde el 15 de septiembre de 1998 hasta el 2 de febrero de 1999). Las gráficas muestran el porcentaje de tráfico de entrada para cada uno de los 17 nodos regionales de RedIRIS y los siguientes orígenes: otros nodos de RedIRIS, la red comercial española, las redes académicas europeas (Ten-34/155), y el resto de Internet (a través del enlace de RedIRIS con USA).

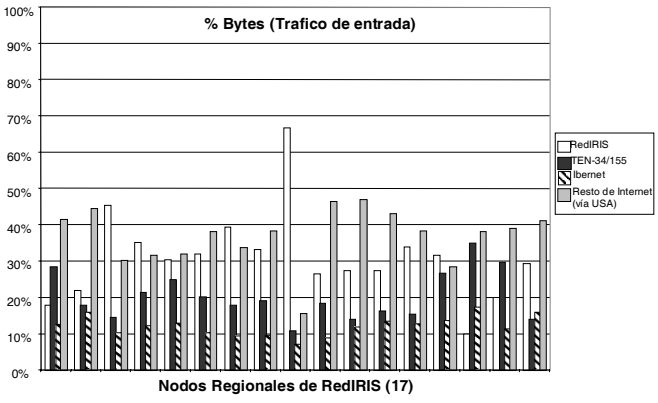

Figura 8. Principales orígenes del Tráfico de RedIRIS (por los nodos regionales de RedIRIS).

Nótese que, en ambos casos (figuras 6 y 8), el hecho de que la distribución sea dada en porcentaje de tráfico de

captura esconde las diferencias de cantidad de tráfico cursado por cada uno de los nodos regionales de RedIRIS.

La figura 9 muestra el tráfico medio de entrada de los 17 enlaces regionales de RedIRIS de forma acumulado, siendo el período de captura es el mismo que en la figura 8.

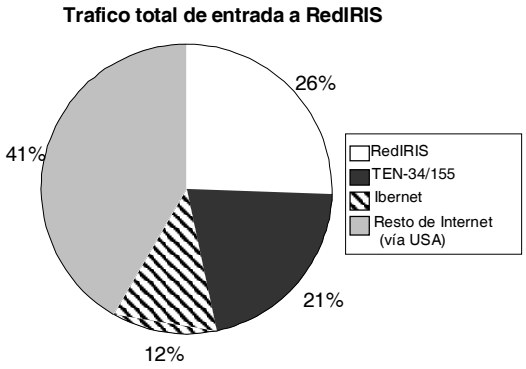

Figura 9. Principales orígenes del tráfico de entrada de RedIRIS (para la totalidad de la red de RedIRIS).

La figura 10 muestra un intento de medir el porcentaje de tráfico académico en el enlace entre RedIRIS y USA. Nótese que la figura 9 está basada en un enfoque pesimista, y sólo el tráfico proveniente de USA (tráfico de entrada a RedIRIS) cuyo origen sea una institución declarada como académica o centro de investigación dentro del Internet Routing Registry es considerado académico. El resto es considerado tráfico comercial ("commodity"). Este resultado corresponde a un mes de captura (entre septiembre y octubre de 1998).

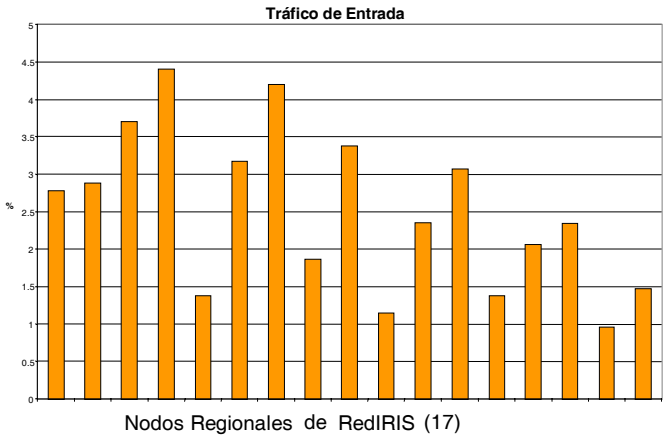

Figura 10. Porcentaje de tráfico académico encontrado en el enlace de USA a RedIRIS (de acuerdo con la descripción del IRR).

Como consideración a la clasificación del tráfico por destinos comerciales, el módulo MODT proporciona estadísticas separadas por cada uno de los nodos regionales de la Red Académica Española. La figura 11 muestra los principales orígenes comerciales del tráfico de entrada del nodo Catalán. El período de captura en este caso va desde enero hasta febrero de 1999. En esta figura solamente se pueden ver los 25 sitios más visitados: el resto (958 sitios para la gráfica del tráfico de entrada y 990 para la gráfica del tráfico de salida) están acumulados en la última columna. Nótese que los 25 sitios más recogen alrededor del 60% del tráfico capturado.

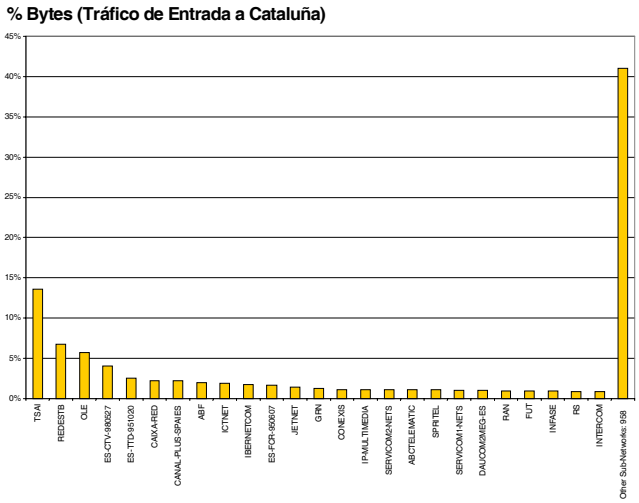

Figura 11. Los 25 sitios comerciales más visitados en Cataluña (de las capturas del tráfico de entrada).

Otro de los resultados interesantes es la clasificación del tráfico por SAs más visitados de la Red Académica Europea (TEN-155). Al igual que en los principales orígenes comerciales, el módulo MODT proporciona estadísticas separadas para cada uno de los nodos regionales de la Red Académica Española. Otra vez, y como ejemplo, la figura 12 incluye las estadísticas sobre los SAs más visitados del nodo Catalán, durante el mismo período de captura que el caso anterior (enero y febrero de 1999).

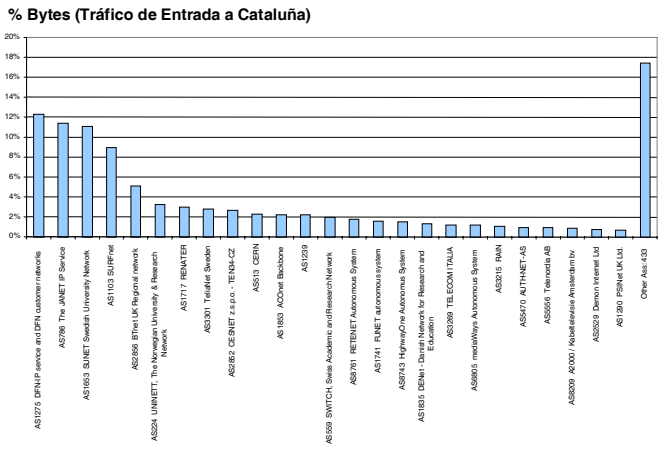

Figura 12. Los 25 SAs más visitados de TEN-155 en Cataluña (de las capturas del tráfico de entrada).

Las figuras de la 13 a la 16 muestran ejemplos de los resultados obtenidos por el Módulo de Análisis de Cabeceras Internet (MCI). La figura 13 muestra el porcentaje de tráfico verificado, tráfico pendiente (no se han encontrado patrones), tráfico desconocido (puertos no registrados), y tráfico descartado (otros protocolos de transporte distintos de TCP/UPD), para el tráfico capturado en la subred de Catalunya entre enero y febrero de 1999.

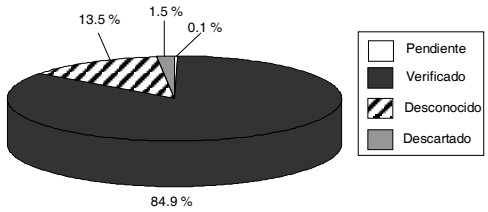

Figura 13. Resultados de la Verificación del Tráfico.

Las figuras 14 y 15 presentan una aplicación de la herramienta: encontrar los servidores más relevantes. Hosts y aplicaciones son determinados por los 25 servidores remotos y locales más visitados respectivamente partiendo del tráfico capturado durante enero y febrero de 1999. En estas gráficas el tráfico de entrada y salida es procesado de forma conjunta. Nótese que los servidores remotos recogen el 74% de la totalidad del tráfico capturado (ver figura 14). Los servidores locales recogen únicamente el 26% del tráfico capturado restante (ver figura 15). También se hace notar que en ambos casos, para gráficas de servidores remotos y locales, la última columna representa el tráfico recogido por el resto de servidores, los que no son los 25 más visitados. El número de estos servidores son 865,811 y 153,900 respectivamente.

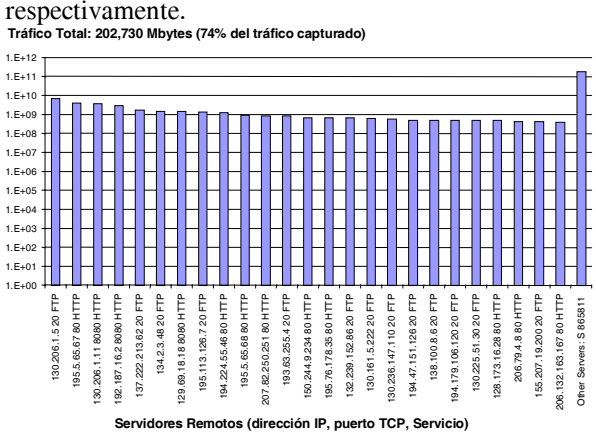

Figura 14. Los 25 servidores más visitados desde Cataluña.

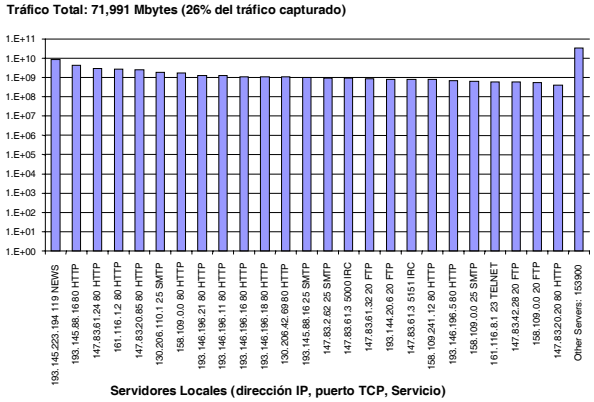

Figura 15. Los 25 servidores más visitados de Cataluña.

La figura 16 muestra un ejemplo de los posibles informes de tráfico desconocido que pueden ser proporcionados por el MCI. Estos resultados son una tentativa de encontrar posibles servidores a partir del número de clientes (número de orígenes que mandan tráfico un mismo destino). La gráfica de la figura 16 corresponde al tráfico capturado durante un solo día.

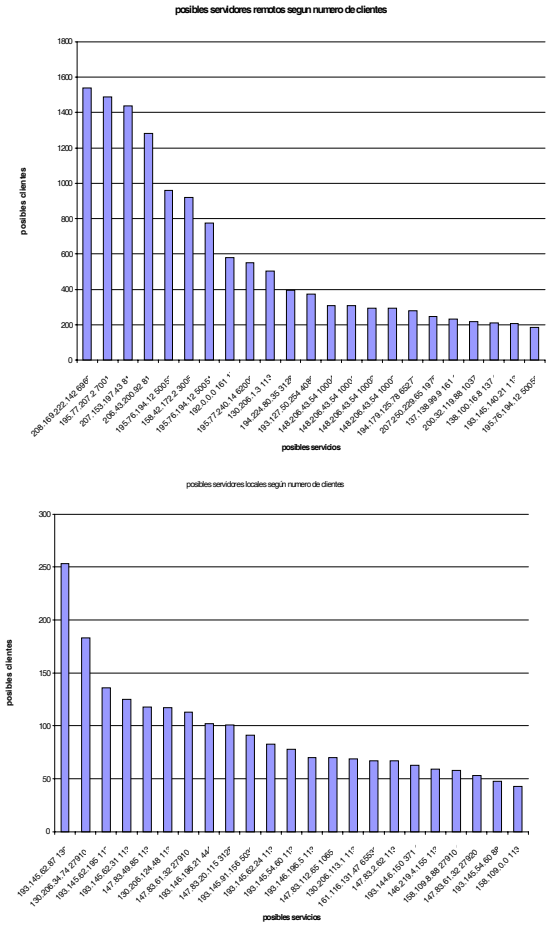

Figura 16. Los posibles servidores remotos y locales según número de clientes.

Figura 16. Estadísticas del tráfico desconocido.

7. Conclusiones

La flexibilidad del protocolo IP combinada con las redes de banda ancha crean diferentes dificultades para el dimensionado, administración y operación de este tipo de infraestructuras. La continua disponibilidad de información detallada y actualizada acerca del tráfico de red es critico para la puesta a punto de las redes, control de carga, imposición de una aceptable política de uso y motivos de seguridad.

La plataforma MEHARI permite tasas de captura de datos sobre líneas STM-1 con unas relaciones de precio/rendimiento diez veces superiores que los de analizadores de protocolos comerciales, también teniendo un mayor rango de VCIs observados. Como la captura se produce al nivel de AAL5, es independiente de los protocolos superiores, y puede ser adaptada para capturar solamente una parte de los paquetes, o ciertos tipos de paquete. Además, su diseño modular permite que las sondas de captura de tráfico sean distribuidas por toda la red, consolidando el pre-procesado de flujos de información en un nodo central.

Los módulos de procesado de tráfico MEHARI actualmente proporcionan información acerca de estadísticas de tráfico convencional, matrices de tráfico por sistemas autónomos según grupos de usuarios, verificación de la asignación de puertos a aplicaciones, y heurísticos sobre la detección de flujos de tráfico inaceptable. Éstos pueden ser personalizados y/o extendidos a conveniencia de los cambiantes requerimientos de información sobre el tráfico de los administradores de red.

El sistema ha estado en marcha en las pruebas de campo 24 horas al día, 30 días al mes durante varios meses. Los resultados de las pruebas de campo han sido contrastados para verificar la validez, la objetividad de las medidas y la validez estadística de los resultados heurísticos sobre los distintos tipos de tráfico. Estos resultados dan una gran confianza en la potencia y la exactitud del sistema MEHARI.

Se cuenta con que el trabajo actual mejore las capacidades de captura y la plataforma de pre-procesado, de cara a obtener mejores relaciones de precio/calidad en la captura. Más módulos de procesado de tráfico están siendo desarrollados, principalmente los que se refieren a seguridad y tarificación.

Dos aplicaciones prácticas del sistema MEHARI son la monitorización de tráfico para soportar mecanismos de facturación y tarificación para Redes Académicas y Corporativas, y para auditar las Políticas de Uso Aceptable (AUP) de las Redes Nacionales de I+D.

Agradecimientos:

El proyecto MEHARI fue financiado por la CICYT (Comisión Interministerial de Ciencia y Tecnología) bajo los contratos TEL97-1897-E y TEL97-TEL97-1893-E. Los autores agradecen la colaboración de RedIRIS, Telefónica I+D y C^4 (Centre de Computació i Comunicacions de Catalunya), quienes han estado siguiendo el trabajo realizado en este proyecto, y por sus valiosos comentarios y sugerencias. Finalmente, agradecer al equipo de desarrollo de OC3-MON por toda la ayuda prestada en la adaptación y modificación del firmware de FORE para los módulos de captura de MEHARI.

Referencias

[1] Kevin Thompson et al., *"Wide-Area Internet Traffic Patterns and Characteristics"*, IEEE Network, November/December 1997.

[2] M. Alvarez-Campana et al., "CASTBA: Medidas de Tráfico sobre la Red Acedémica Española de Banda Ancha", TELECOM I+D, Madrid, October 1998.

[3] J. Aspirdof et al., "OC3MON: Flexible, Affordable, High-Performance Statistics Collection", INET'97, Malasya, june 1997.

[4] David Newman et al., "Intrusion Detection *Systems, Suspicious Finds"*, Data Communications, August 1998.

[5] M. Alvarez-Campana, A. Azcorra, J. Berrocal, D. Larrabeiti, J.I. Moreno, J.R. Pérez, "CASTBA: Internet Traffic Measurements over the Spanish ATM Network", HP-OVUA Workshop, Rennes (France), April 1998.