

II Congreso Internacional de la Lengua
Valladolid, 16-19 Octubre de 2001

El uso de Internet en la Universidad Española

Arturo Azcorra Saloña
Univ. Carlos III de Madrid, azcorra@it.uc3m.es
Vicente Ortega Castro
Consejo de Universidades, Ministerio de Educación

1. Introducción

La naturaleza técnica de las redes telemáticas, como es el caso de Internet, permite la realización de un análisis automatizado y masivo de los diálogos que se producen a través de las mismas. Este análisis puede ser tan completo como se desee, si bien limitado por la capacidad de captura de tráfico y de análisis del mismo de los equipos que se utilicen para tal fin. La automatización de las medidas permite el manejo de volúmenes masivos de datos, pero tiene la restricción de que los datos obtenidos no requieran análisis interpretativo humano, viéndose limitados en algunos casos a la aplicación de criterios heurísticos informatizados.

En este trabajo se realiza una exposición e interpretación de distintas medidas y análisis de tráfico que han sido realizadas en la red académica RedIris. La red académica RedIris interconecta las universidades y centros de investigación españoles entre sí, así como con el resto de la Internet. Los datos fuente de medida se han obtenido a través de varios proyectos de investigación financiados por la Comisión Interministerial de Ciencia y

Tecnología. Estos proyectos son el proyecto CASTBA, el proyecto MEHARI y el proyecto MIRA, que han sido realizados por investigadores de la Universidad Carlos III de Madrid, la Universidad Politécnica de Madrid y la Universidad Politécnica de Cataluña.

Las medidas que sirven de base a este trabajo se realizaron a lo largo de los años 1997, 1998 y 1999, por lo es preciso señalar que la situación actual de tráfico sin duda habrá experimentado algunas variaciones sobre lo aquí presentado, dado el dinamismo en la evolución del uso de Internet.

2. Medidas

Las medidas de base se pueden dividir en dos tipos:

- Aquéllas cuyas conclusiones pueden derivarse del valor, de forma independiente (en principio) del enlace sobre el que se midan. Por este motivo se realizan sobre un enlace dado, siempre que sea de alto tráfico al efecto de tener datos estadísticamente más fiables. Ejemplos de estas medidas son las distribuciones horarias, o diarias, de tráfico y la composición del tráfico por servicios IP.
- Las que se consideran dependientes del enlace sobre el que se miden. Por este motivo se realizan sobre todos los enlaces. Un ejemplo de estas medidas lo constituye la distribución de tráfico de las distintas comunidades autónomas.

Las medidas presentadas según franja horaria corresponden al valor medio de la magnitud a medir, presentado a lo largo de las 24 horas del día. La media se calcula sobre los datos tomados durante los 96 intervalos de quince minutos de cada día dentro del período de medida (ej. datos acumulados de todos los intervalos de 13:00 a 13:15 horas dentro del periodo de medida). El objetivo de este tipo de medidas es determinar si se producen variaciones de la magnitud a medir a lo largo de las 24 horas del día.

3. Carga de tráfico en los enlaces

Estas medidas caracterizan la carga de los enlaces ATM de la red troncal de RedIris.

En primer lugar se ha medido la carga de los circuitos ATM en función de los distintos destinos de los mismos. Se ha distinguido entre el tráfico de entrada y salida de cada nodo o comunidad autónoma.

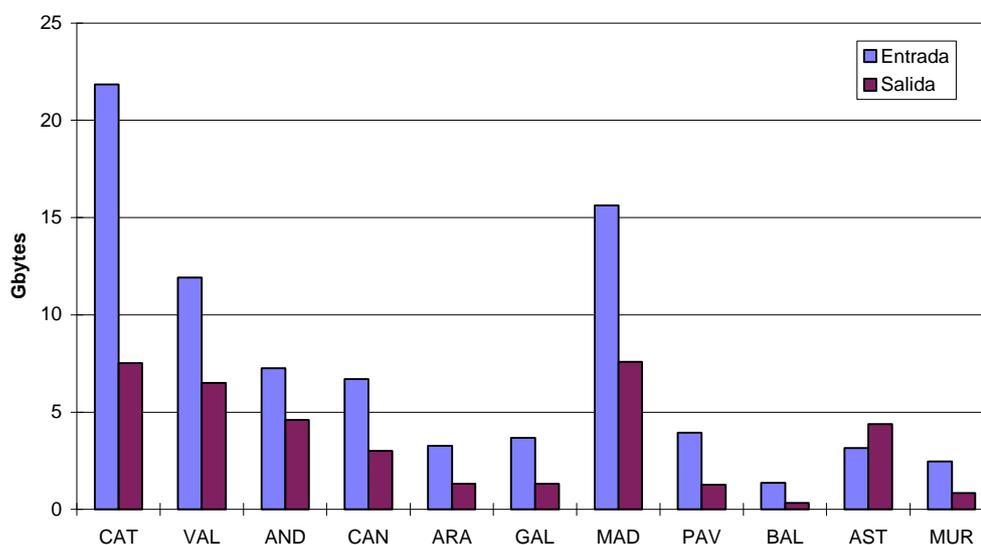
Adicionalmente se presentan medidas de la evolución horaria de la carga de los circuitos. Estas medidas identifican la ocupación media de determinados circuitos ATM de la red troncal de RedIris. Las medidas presentadas corresponden a la carga agregada de todos los circuitos nacionales de Rediris a las comunidades autónomas por ATM.

Los valores que se muestran en las gráficas de carga porcentual (no así en las de distribución porcentual de tipo de tráfico) corresponden al porcentaje de ancho de banda ocupado sobre el total disponible en la línea.

3.1. Origen y destino del tráfico

Los resultados se muestran en función de cada circuito de la red troncal. La medida corresponde al volumen de datos transferido durante un día laborable en media, y se expresan en Gigabytes (Un Gigabyte equivale a 2 elevado a 30 bytes, que aproximadamente son mil millones de bytes). Se calculan midiendo la información transferida en cada día laborable y sacando la media a lo largo del periodo de medida.

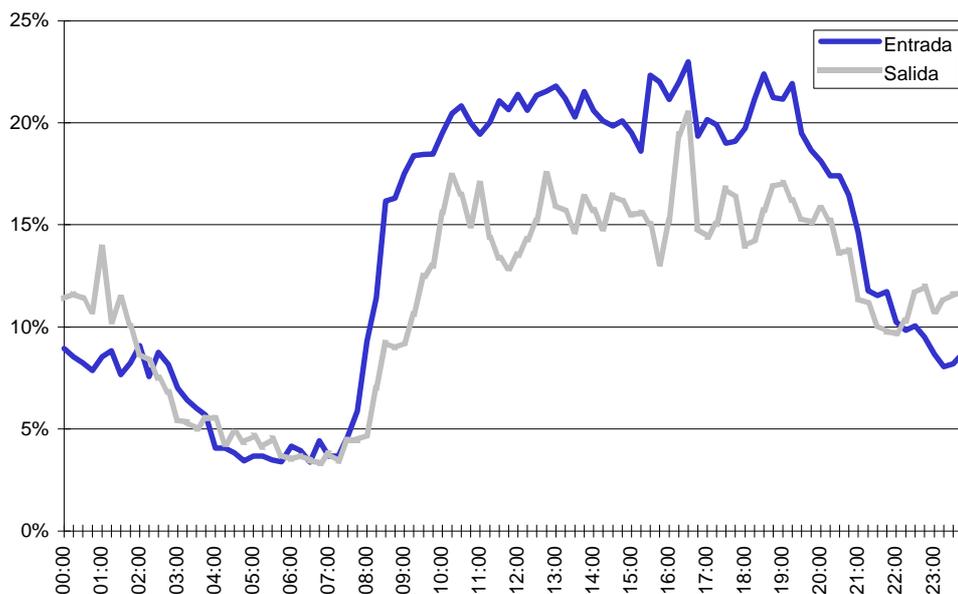
Como valor de referencia, una línea de 2.048 Kilobits/segundo es capaz de transferir 20,6 Gigabytes diariamente.



Nodo	Código	Entrada (Gbytes)	Salida (Gbytes)	Capacidad (Gbytes) Entrada	Capacidad (Gbytes) Salida	Capacidad (Mbit/s) Entrada	Capacidad (Mbit/s) Salida
Cataluña	CAT	21,83	7,52	80,47	40,23	8	4
Valencia	VAL	11,91	6,50	70,41	40,23	7	4
Andalucía	AND	7,25	4,60	60,35	30,17	6	3
Canarias	CAN	6,70	3,03	40,23	30,17	4	3
Aragón	ARA	3,27	1,32	40,23	20,12	4	2
Galicia	GAL	3,68	1,33	40,23	20,12	4	2
Madrid	MAD	15,62	7,57	1005,83	1005,83	100	100
País Vasco	PAV	3,94	1,26	40,23	20,12	4	2
Baleares	BAL	1,36	0,33	30,17	20,12	3	2
Asturias	AST	3,16	4,38	40,23	30,17	4	3
Murcia	MUR	2,45	0,83	30,17	20,12	3	2

3.2. RedIris Nacional en función de la franja horaria

La evolución a lo largo de la franja horaria se ha tomado para el total de los enlaces de RedIris a las comunidades autónomas por ATM (sin incluir el nodo de Madrid). El 100% equivale a 47 Mbit/s y 27 Mbit/s para las curvas de entrada y salida, respectivamente. La curva de entrada se corresponde con tráfico entrante a las comunidades autónomas.



4. Caracterización de Servicios sobre IP

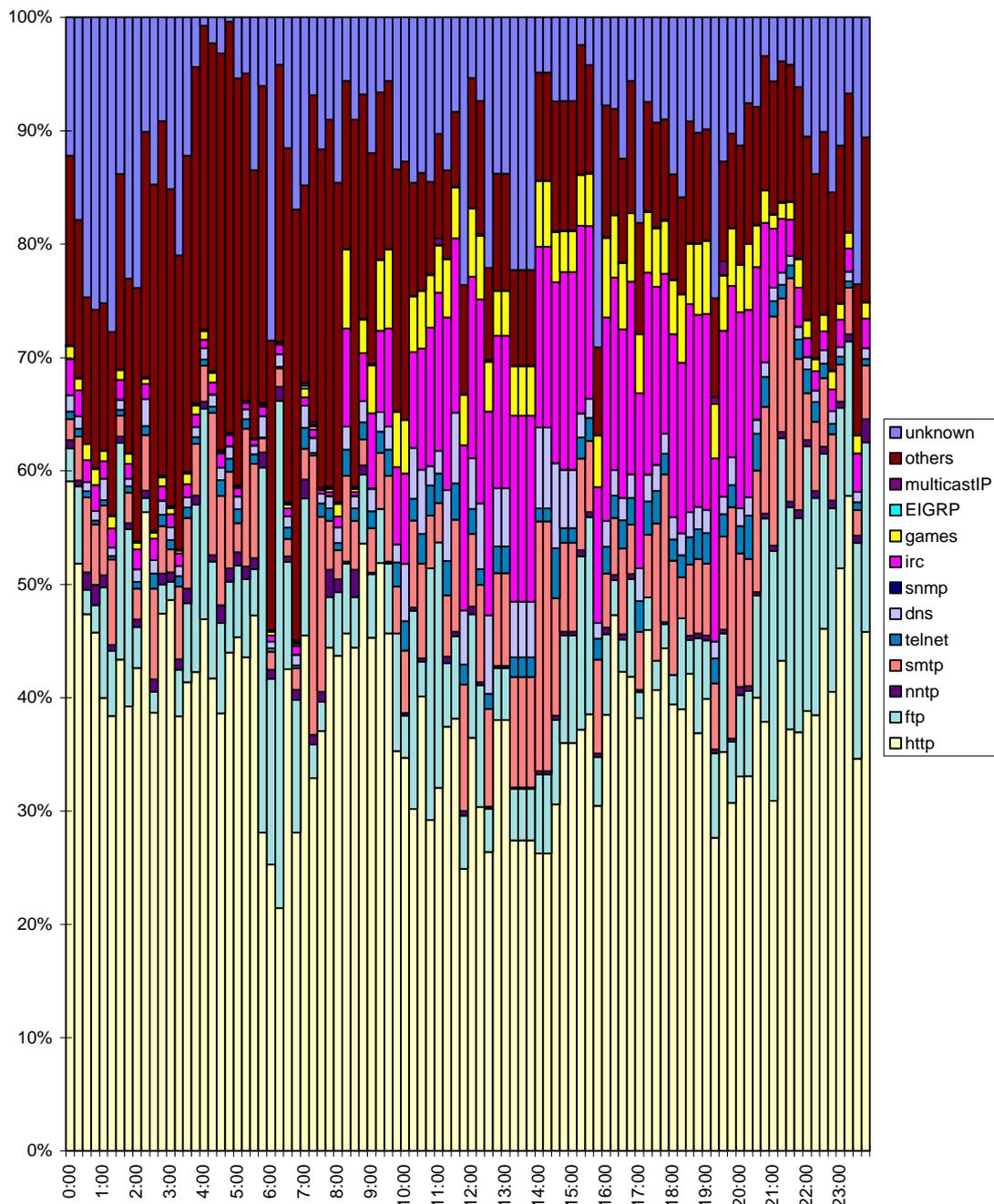
Estas medidas tienen por objeto caracterizar la naturaleza del tráfico IP cursado por el enlace ATM-MAD-CAT (Madrid - Cataluña) de la red troncal de RedIris, en base a los servicios o protocolos de aplicación utilizados.

En todas las medidas realizadas se ha distinguido entre el tráfico de entrada y de salida del nodo de Barcelona. Los servicios analizados se citan seguidamente, indicando entre paréntesis la leyenda que los identifica en las gráficas:

- Uso de Navegadores web (http)
- Transferencia de ficheros (ftp)
- Servicio de noticias o "news" (nntp)
- Correo electrónico (smtp)
- Terminal remoto de acceso a sistema informático (telnet)
- Resolución de nombres de dominio (dns)
- Gestión de red (snmp)
- Conferencias textuales o "chat" (irc)
- Juegos en red (games)
- Otros servicios de índole más técnica o tráfico desconocido (EIGRP, multicastIP, others y unknown)

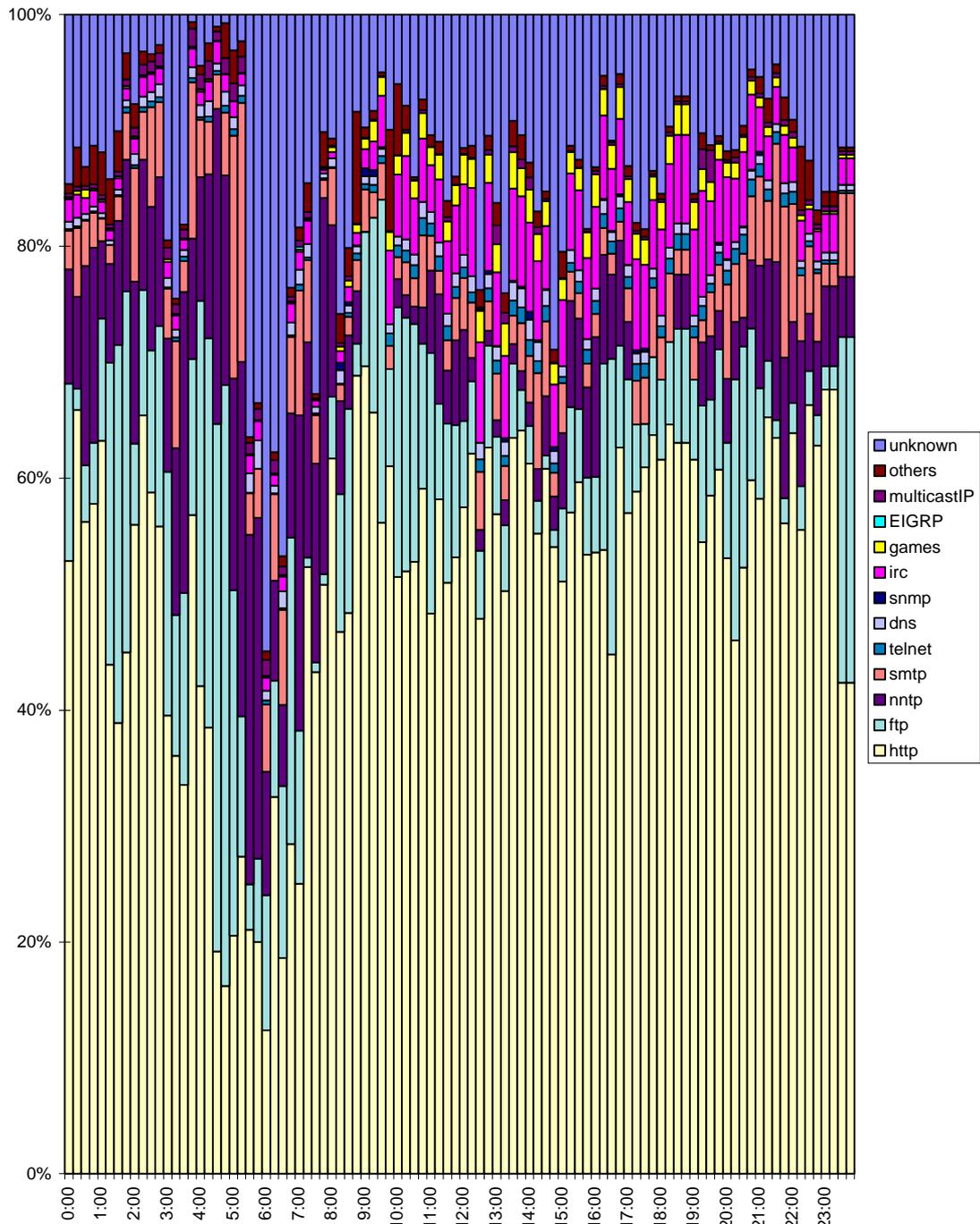
4.1. Distribución de octetos por servicio en función de la franja horaria (salida)

Esta medida identifica los porcentajes de octetos de los distintos servicios a lo largo de un día sobre el circuito de salida ATM-MAD-CAT (Madrid - Cataluña) de la red troncal.



4.2. Distribución de octetos por servicio en función de la franja horaria (entrada)

Esta medida identifica los porcentajes de octetos de los distintos servicios a lo largo de un día sobre el circuito ATM-MAD-CAT (Madrid - Cataluña) de entrada de la red troncal.



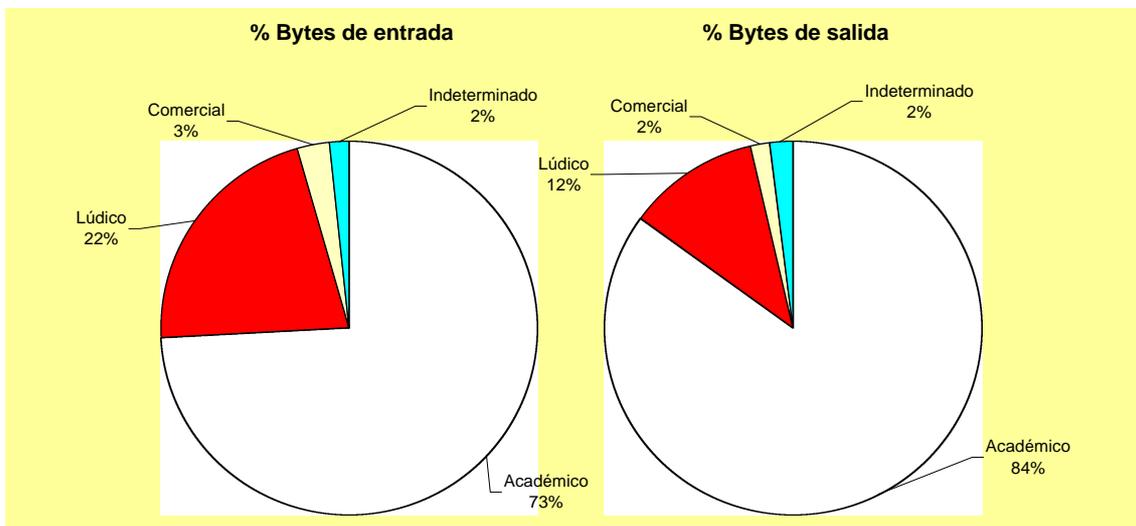
5. Resultados de Uso de la Internet

En este apartado se presentan datos de clasificación de tráfico por Comunidades Autónomas. El sistema de análisis utilizado, MEHARI/MIRA, tiene por objeto clasificar los flujos de tráfico que circulan por Internet bajo cuatro categorías: Académico (ACA), Lúdico (LUD), Comercial (COM), e Indeterminado (IND).

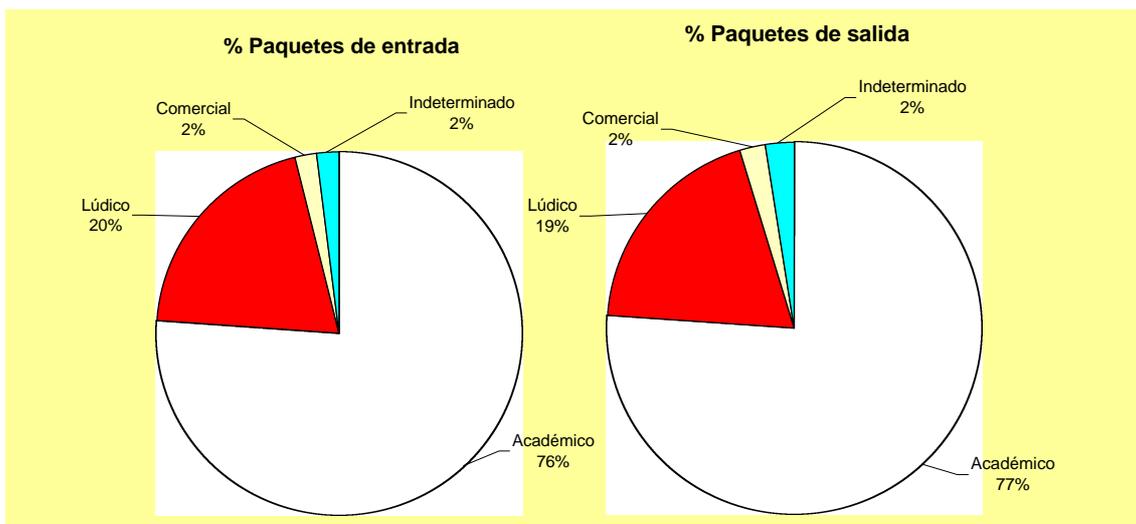
5.1. Resultados globales

Las gráficas de este apartado muestran, la distribución porcentual de octetos y paquetes correspondientes a cada categoría de tráfico (ACA, LUD, COM, IND), para el conjunto de enlaces regionales de RedIris. Se muestran gráficas separadas para los sentidos de entrada y salida.

5.1.1. Distribución de bytes



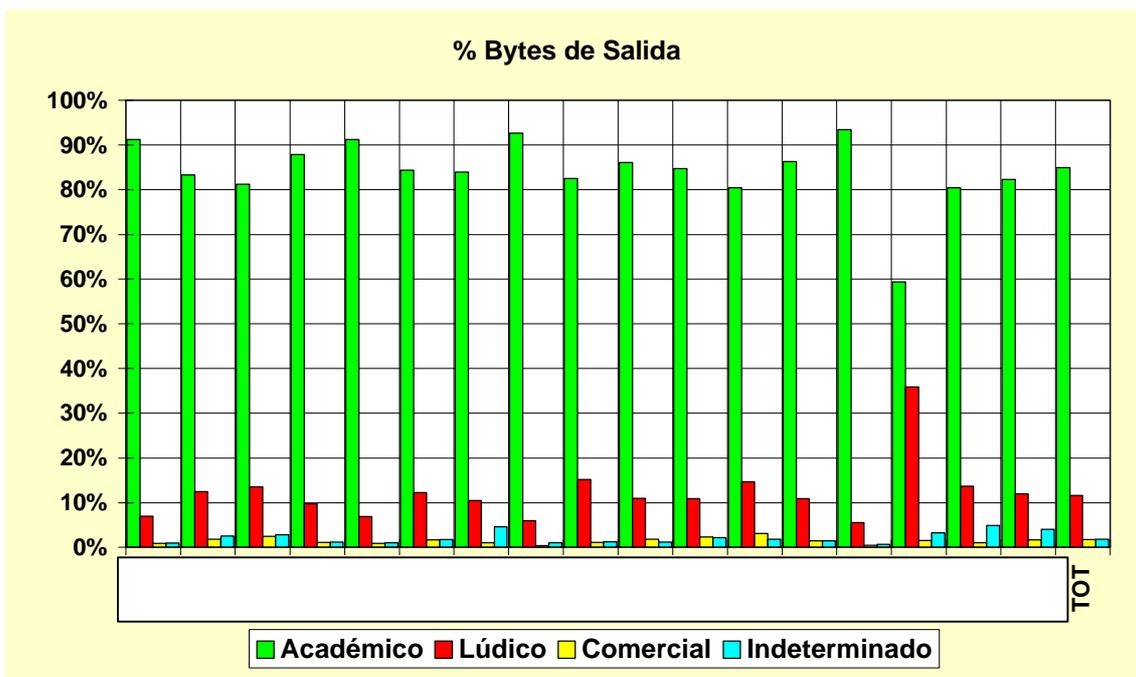
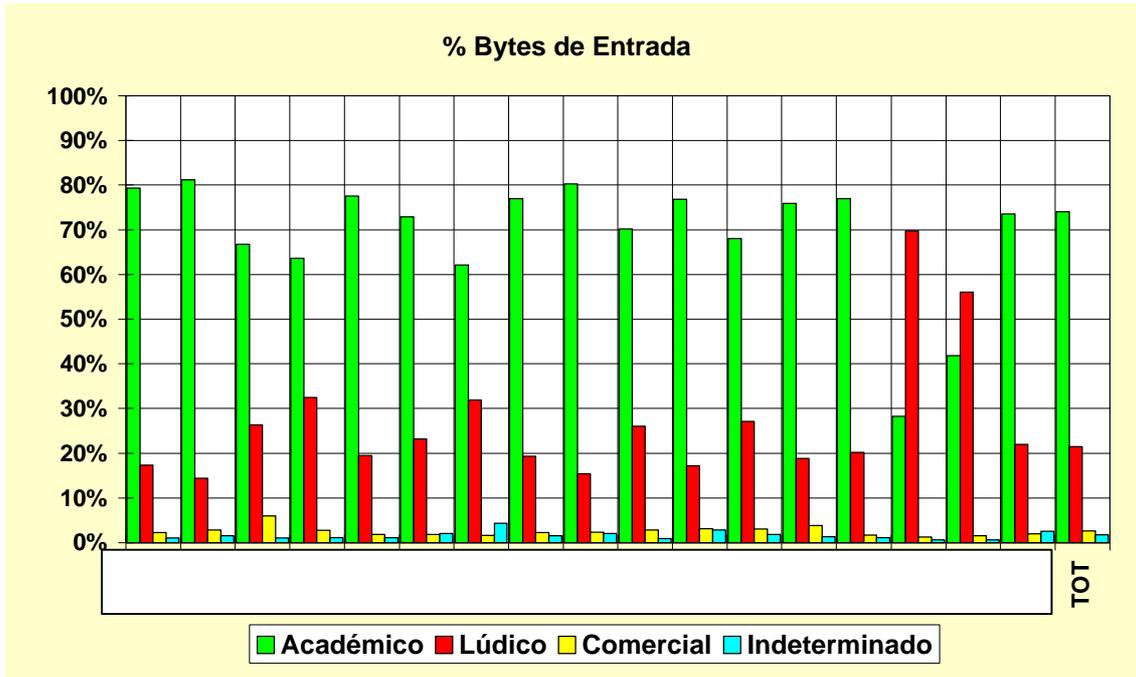
5.1.2. Distribución de paquetes



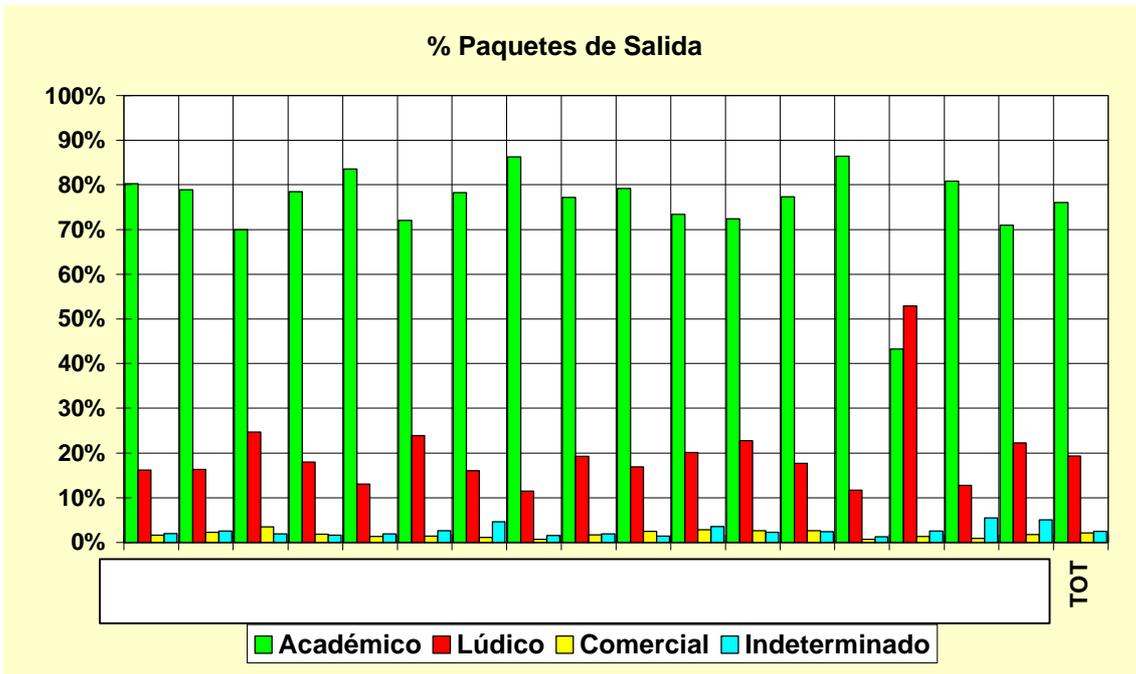
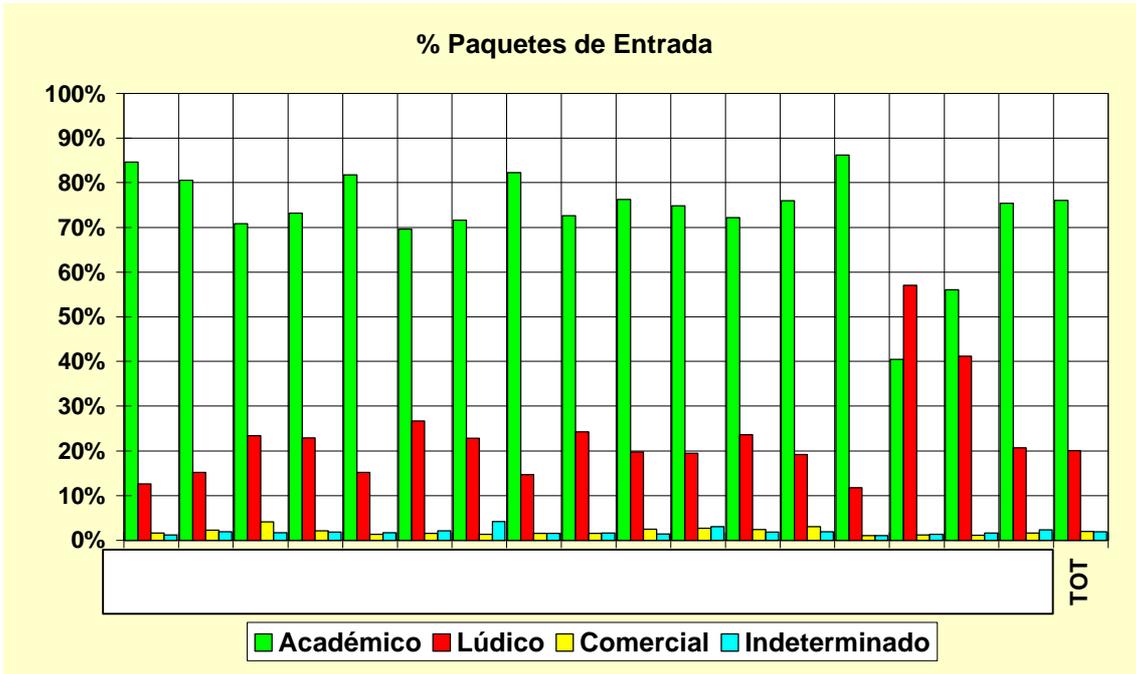
5.2. Resultados por CC.AA.

Las gráficas de esta apartado muestran , para cada CC.AA. y cada sentido de transmisión, los porcentajes de bytes y de paquetes correspondientes a cada categoría de tráfico (ACA, LUD, COM, IND). La identidad de cada CC.AA. se ha ocultado por motivos de confidencialidad.

5.2.1. Distribución de bytes



5.2.2. Distribución de paquetes



6. Conclusiones

En este apartado se exponen las principales conclusiones que pueden extraerse a raíz de las medidas realizadas referentes a la utilización de Internet que se realiza en las universidades españolas.

- Las comunidades autónomas con mayor volumen de tráfico de entrada son Cataluña, Madrid y Valencia, en este orden, que totalizan aproximadamente el 50%.
- El tráfico de entrada a los nodos regionales es aproximadamente el doble que el de salida (81 frente a 38 Gigabytes diarios). Esto implica que las universidades españolas son consumidores netos de información, recibiendo más del doble de información que la que les es consultada.
- Teniendo en cuenta la capacidad del enlace con Estados Unidos, que en el momento de realización de las medidas de tráfico era de 4Mbit/s (aprox. 40,23 Gbytes diarios), se concluye que el 50% de la información de entrada a la comunidad universitaria española tiene su origen en Estados Unidos.
- La aplicación más usada son los Navegadores Web, cuyo uso supone un 32% de los paquetes. En segundo lugar están las aplicaciones de multiconferencia textual (similar a los "party-line" telefónicos), tales como IRC y Chats (un 31% de los paquetes). Siguen la transferencia de ficheros (ftp) y el terminal informático virtual (telnet), con un 7% cada uno. Un volumen similar tienen las aplicaciones de juegos y el tráfico desconocido. El 9% restante corresponde a SMTP, DNS, NNTP, IPmulticast y otros. Los datos de partida de tráfico en paquetes no se han incluido en el artículo, pero si se midieron en los distintos proyectos que han servido de base.
- Desde el punto de vista del tráfico en bytes (no en paquetes) el servicio predominante es el Web, que supone un 52% del total del tráfico en bytes.
- El fin para el que los universitarios utilizan la internet difiere notablemente en las distintas comunidades autónomas. Por citar un solo hecho diferencial a modo de ejemplo, se puede observar en las gráficas que en ciertas comunidades predomina el uso lúdico (más del 50% del tráfico), frente a otras en las que predomina el académico.
- No se disponen de datos sobre el porcentaje de uso de los distintos idiomas en la información que fluye por Internet ya que no era uno de los objetivos de los proyectos de medidas de base. Sin embargo, es preciso resaltar que la herramienta de captura y análisis desarrollada en el proyecto MIRA podría configurarse sin dificultades para realizar dicho análisis, que sin duda tendría un elevado valor lingüístico.

7. Referencias

- [1] M.Álvarez et al. “CASTBA: Medidas de Tráfico sobre la Red Académica Española de Banda Ancha”, VIII Jornadas TELECOM I+D, Madrid-Barcelona, Octubre 1998.
- [2] M Álvarez-Campana et al “Caracterización de Servicios Telemáticos sobre la Red Académica de Banda Ancha”, Primer Seminario del Programa Nacional de Aplicaciones y Servicios Telemáticos (SPAST-I). Navarra, diciembre 1999.
- [3] J.Aspirdof et al. “OC3MON: Flexible, Affordable, High Performance Statistics Collection”, INET’97, Lamasya, june 1997.
- [4] P. Lizcano et al “MEHARI: System for Analysing the Use of the Internet Services”,Computer Networks 31 (1999), pag.2293-2307.
- [5] A. Azcorra, et al “Análisis del uso de los servicios Internet a partir del sistema MEHARI: prueba de campo sobre RedIRIS”, Primer Seminario del Programa Nacional de Aplicaciones y Servicios Telemáticos (SPAST-I). Navarra, diciembre 1999.
- [6] Carlos Veciana-Nogués, Jordi Domingo-Pascual y Josep Solé-Pareta. “Server Location & Verification Tool for Backbone Access Points”, ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management, Monterey (USA), Septiembre 2000.
- [7] Nevil Brownlee “Traffic Flow Measurements. Experiences with Netramet” IETF RFC 2123, March 1997