

# Monitorización activa de altas prestaciones mediante la plataforma paneuropa ETOMIC

E. Magaña, U. Alonso, F. Astiz, D. Morató, M. Izal, F. Naranjo y J. Aracil  
Departamento de Automática y Computación  
Universidad Pública de Navarra  
C/ Campus Arrosadia, 31006 Pamplona  
Teléfono: 948 169853 Fax: 948 168924  
E-mail: eduardo.magana@unavarra.es

***Abstract** In this paper we present the first set of active measurements that we have made using the ETOMIC system. ETOMIC is a paneuropean traffic measurement infrastructure with GPS-synchronized monitoring nodes. Specific hardware is used in order to provide high-precision transmission and reception capabilities. Besides, the system is open and any experiment can be executed. Internet measurements with high infrastructure requirements are now possible like one-way delay, routes and topology changing, congestion detection and virtual path aggregation detection. We will explain the results and how easy is to implement these measurements using the tools provided by ETOMIC, specially the API for using the specific sending and receiving capabilities.*

## 1 Introducción

Uno de los grandes campos de investigación en la actualidad es la medición de red mediante técnicas de monitorización activa o pasiva, de forma que se pueda extraer suficiente información como para poder realizar un dimensionamiento correcto del crecimiento de las troncales, modelar el tráfico de los usuarios o detectar las necesidades de calidad de servicio de nuevas aplicaciones. Sin embargo, conforme aumenta la velocidad de las troncales y redes de acceso se hace necesario disponer de herramientas de monitorización más sofisticadas que las que habitualmente se han venido utilizando.

De esta manera, si nos queremos enfrentar al problema de realizar medidas en Internet nos encontramos con los siguientes requisitos. En primer lugar será necesario una alta precisión en las marcas temporales de las medidas a realizar, sobre todo conforme se trabaje con mayores anchos de banda. Además, si se quiere realizar mediciones en una Internet global, necesitaremos mecanismos que provean la sincronización precisa entre equipos dispersos geográficamente, de manera que se pueda realizar una interpretación correcta de los resultados. Por otro lado, para lo que a monitorización activa se refiere [1], se hará necesario disponer de hardware específico que permita generar ráfagas de paquetes con alta precisión temporal en el espaciado entre paquetes o que permita la recepción de paquetes sin pérdidas. Estos requisitos quedarán por lo general fuera de las capacidades de plataformas PC de bajo coste habitualmente utilizadas.

Precisamente el proyecto integrado EVERGROW<sup>1</sup> del VI Programa Marco de la Unión Europea tiene como objetivo fundamental el crear unas bases de conocimiento de la evolución de Internet hasta el

año 2025. Aplicando teorías de sistemas complejos se hace necesario disponer de datos reales de la Internet actual. Para obtener esos datos, dentro de este proyecto integrado se enmarca la plataforma de monitorización ETOMIC<sup>2</sup> (*European Traffic Observatory Measurement InfrastruCture*). El sistema ETOMIC consiste en un sistema central de gestión (CMS, Central Management System) que gestiona y supervisa nodos de monitorización distribuidos por toda Europa. En la actualidad se dispone de 12 nodos y se pretende llegar a tener 50 en localizaciones escogidas de toda Europa, principalmente en universidades, centros de investigación, operadoras y empresas de telecomunicaciones.

Las características principales de ETOMIC [2, 3] son las siguientes:

- Ofrece una plataforma de monitorización abierta a la comunidad investigadora, de manera que cualquier investigador pueda obtener una cuenta en el sistema y acceder a todos los recursos.
- Provee monitorización activa y pasiva, si bien en la fase inicial está enfocado a monitorización activa.
- Es totalmente reconfigurable, dejando al investigador la posibilidad de realizar cualquier tipo de medida al permitirle correr en los nodos el software que desee. Se dispone además de una selección básica de programas que permiten realizar las medidas más habituales.
- Todo ello dentro de un esquema hardware de alta precisión, tanto en temporización como en sincronización, para recepción y emisión de paquetes sobre la red.

<sup>1</sup>[www.evergrow.org](http://www.evergrow.org)

<sup>2</sup>[www.etomic.org](http://www.etomic.org)

Para realizar el presente estudio se ha utilizado la plataforma ETOMIC mostrando algunas de las medidas que es posible realizar y con ello la potencia real del sistema.

## 2 Plataforma ETOMIC

La plataforma ETOMIC provee a los investigadores de un interfaz web de acceso al sistema. Este interfaz es servido por el CMS que es el encargado de gestionar la realización de los experimentos que se deseen en los nodos distribuidos por Europa.

Cada uno de los nodos está dotado del siguiente hardware que los convierte en herramientas de alta precisión:

- Plataforma PC: un PC con Debian GNU/Linux proveerá capacidades de almacenamiento de datos en disco, comunicación con el CMS, watchdog y el entorno para la ejecución de experimentos. Dispone de una tarjeta Ethernet para comunicación de gestión con el CMS, pero también utilizable para los experimentos.
- GPS Garmin 35 HVS: permite dotar al sistema de alta precisión en las marcas temporales de paquetes recibidos y en el espaciado de paquetes en emisión, así como sincronización de los relojes de todos los nodos. La precisión del sistema será uno de los aspectos estudiados en este trabajo.
- Tarjetas Endace DAG 3.6GE [4]: tarjetas Ethernet 10/100/1000 con conexión al GPS, permite colocar un *timestamp* por hardware a la recepción de los paquetes y definir la temporización de ráfagas de paquetes en emisión también por hardware, consiguiendo cotas de precisión muy por encima de los sistemas convencionales. Además se basan en una arquitectura de memoria compartida con lo que eliminan la sobrecarga de interrupciones de los interfaces de red habituales.
- Convertidor serie: permite adaptar las señales del GPS (RS-232) con la entrada de la tarjeta Endace DAG 3.6GE (RS-422) y el puerto serie del ordenador.

En la Fig.1 se muestran los componentes específicos del sistema: tarjeta Endace DAG, GPS y conversor. Para el proyecto se ha desarrollado un firmware para las tarjetas Endace DAG 3.6GE que ofrece funcionalidades de generación de tráfico (el firmware convencional únicamente tiene capacidades en recepción), proveyendo un esquema para la generación de ráfagas de hasta 256 paquetes donde es posible definir el tamaño y espaciado entre paquetes con resolución del orden de nanosegundos.

Para facilitar la labor del investigador se ha realizado un API que permite aprovechar las funcionalidades de las tarjetas Endace DAG a través de un interfaz muy sencillo para generación y recepción de

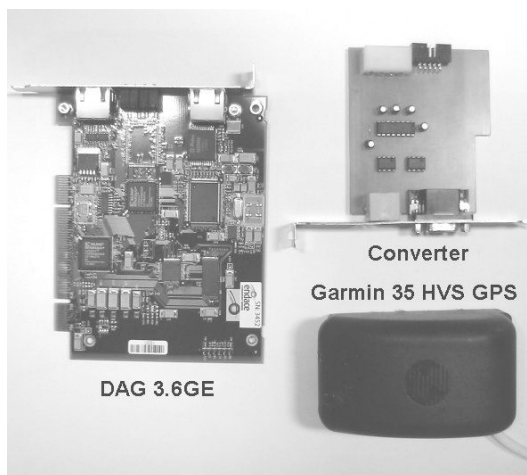


Figura 1: Endace DAG, GPS y conversor

paquetes, tanto para modo activo como pasivo. El API permite comprobar los programas desarrollados para las tarjetas Endace DAG sin disponer de ellas, realizando una emulación por medio de una tarjeta de red Ethernet convencional en el PC del investigador. Este API será de utilidad para aquellos investigadores que deseen lanzar sus propios programas en cada nodo, siempre dentro de las limitaciones de un usuario normal sin permisos de administración.

Actualmente se encuentran desplegados nodos en Suecia, Inglaterra, Alemania, Francia, Hungría, Italia, Israel y España. El CMS se encuentra localizado en la Universidad Pública de Navarra, Pamplona. Estos nodos requieren de una instalación especial debido a la necesidad de colocar el GPS con buena visibilidad aérea, en la mayoría de los casos conseguida colocándolo en el tejado (normalmente a decenas de metros de distancia del nodo). En la Fig. 2 se muestra una captura de la web de información del sistema con la disposición de los nodos operativos. Este interfaz web provee a los usuarios de todo lo necesario para realizar sus experimentos, desde subir y compilar los programas hasta reservar los nodos y seleccionar los programas a correr en cada nodo. Las fases típicas de definición y ejecución de los experimentos son las siguientes:

- Subir al CMS ficheros ejecutables, código fuente o ficheros de datos. Una vez en el CMS tendremos accesibles esos ficheros para programar los experimentos en los nodos.
- Crear un *bundle*. El bundle define las características de nuestro experimento sin concretar el momento en el que se va a realizar. Para ello deberemos seleccionar qué nodos necesitamos para el experimento y definir los ficheros que es necesario subir a cada uno de ellos, así como los puntos de ejecución que queremos insertar para lanzar nuestros programas con la temporización que consideremos oportuna.
- Definir un experimento. Consiste en asignar

la franja temporal en la que queremos lanzar determinado bundle definido con anterioridad.

- El CMS se encarga de esperar y planificar la ejecución del experimento de manera automática sin intervención del investigador. Antes de la hora programada para el experimento se encargará de subir los ficheros necesarios para cada nodo, y en el momento de cada ejecución lanzar los programas adecuados. Cuando termine el tiempo para el experimento se bajará los resultados de cada nodo al CMS.
- Descarga de resultados. El investigador, cuando haya finalizado el experimento y lo estime oportuno, podrá utilizar el interfaz web para bajarse los resultados de su experimento desde el CMS.

Toda la interacción del investigador y la plataforma ETOMIC se realiza a través del interfaz web del CMS, facilitando en gran medida la realización de experimentos complejos en múltiples nodos simultáneamente.

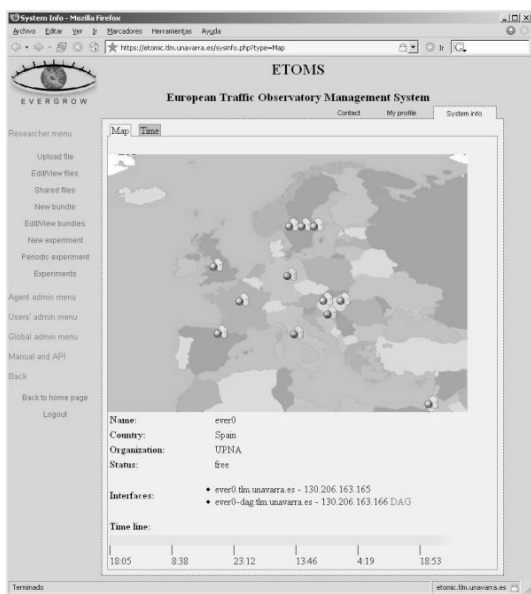


Figura 2: Interfaz web para el investigador

### 3 Precisión del sistema ETOMIC

En las redes actuales, un requisito previo para obtener mediciones correctas es disponer de un instrumento de medida con una resolución suficiente para manejar paquetes que se transmiten a altas velocidades. Más concretamente, la mayor parte de los enlaces troncales de las redes Rediris y Geant están formados por enlaces de 2.5 y 10 Gbps. Además de una buena resolución se requiere que los nodos estén sincronizados, o dicho de otra forma, los nodos tienen que compartir la misma hora con una gran precisión.

Como ejemplo de la resolución necesaria, un paquete Ethernet de tamaño máximo en un enlace

Gigabit tarda en transmitirse 12 microsegundos y en general se requiere una resolución de decenas de nanosegundos. Es importante disponer de gran resolución cuando, por ejemplo, se quieren detectar espaciados entre paquetes enviados de forma inmediatamente consecutiva (*back-to-back*). El envío de paquetes *back-to-back* se puede utilizar para la estimación del ancho de banda.

La sincronización entre los nodos se resuelve gracias al sistema GPS, el cual además de proveer los conocidos servicios de localización globales también ofrece servicio de sincronización. La sincronización es necesaria para medir correctamente el tiempo que tarda un paquete desde que se manda desde un nodo origen hasta que llega al nodo destino (retardo). El retardo de los paquetes además de indicar la velocidad de transmisión de los paquetes ofrece información sobre el nivel de congestión de las redes. Aunque la magnitud en las medidas de retardo es del orden de decenas de milisegundos, la sincronización debe ofrecer una mayor precisión o en caso contrario medidas como el espaciado entre paquetes *back-to-back* obtendrían valores oscilantes no tolerables.

El procedimiento básico para realizar mediciones es el de establecer una marca de tiempo cuando se transmite y otra cuando se recibe. Estas marcas de tiempo se deben establecer con el mínimo retardo y variabilidad posible. Para ello no se puede confiar en hardware convencional o en sistemas operativos de propósito general. El hardware convencional no ofrece servicios de marcas de tiempo en los paquetes y tienden a ofrecer unos tiempos de respuesta variables. Los sistemas operativos de propósito general no están diseñados para ofrecer marcas de tiempos con una alta precisión y en cualquier caso están supeditados a la no predecibilidad temporal del hardware convencional. Los errores de *timestamp* de este tipo de plataformas están alrededor de 0.1ms [5] con respecto al reloj real de la máquina, sin contar las desviaciones de éste y la falta de sincronización entre diferentes equipos.

La solución utilizada para resolver todos estos problemas ha sido utilizar hardware de propósito específico, concretamente se han utilizado tarjetas Gigabit-Ethernet DAG 3.6GE de Endance. Las principales características de estas tarjetas de red son:

- Utilizan GPS para su sincronización.
- Establecen marcas de tiempo en los paquetes en emisión y en recepción con una resolución aproximada de 60 ns (exactamente  $1/2^{24}$  segundos).
- Permite la programación del envío de paquetes con un espaciado concreto con la resolución mencionada en el punto anterior de forma que el sistema operativo no influye en la temporización de los paquetes.

La evaluación de la precisión se ha hecho realizando medidas del tiempo de retardo entre dos nodos a través de una red con retardo constante. Debido a la imposibilidad de conectar dos tarjetas DAG con un

cable cruzado se ha utilizado un concentrador y se ha comprobado que el retardo a través del concentrador es suficientemente estable. Más concretamente, en la evaluación se comparará una configuración en la que las tarjetas se sincronizan entre sí, es decir sin GPS y compartiendo la misma fuente de sincronización (una de las propias tarjetas en modo *master*) y otra configuración en la que cada tarjeta está sincronizada con su propio GPS. El hardware utilizado ha sido: tarjetas de red DAG 3.6GE de Endance, concentrador Fast-Ethernet OfficeConnect DualSpeed Hub 8 (3C16753) de 3Com, receptores GPS Garmin 35-HVS y tarjetas conversoras serie.

La primera prueba consiste en validar la prueba diseñada, esto se realiza utilizando la configuración sin GPS y observando por una parte la variación del tiempo de retardo y por otra la diferencia de tiempo entre paquetes en tiempo de emisión y recepción. Ambos tiempos deben de ser el mismo con una alta precisión, los resultados obtenidos en esta prueba indicarán la precisión de los resultados que se obtendrán en pruebas posteriores.

En las pruebas se obtuvo que el retardo promedio fué de 1276 ns, que la diferencia entre el retardo máximo y mínimo fué de 299 ns (menos de 30 bits en Fast-Ethernet.) y la desviación estándar 41 ns. Nos referiremos al valor del retardo promedio obtenido en esta prueba como *retardo calibrado* y será utilizado cuando comparemos los valores de retardo obtenidos cuando se realicen medidas utilizando GPS.

Por otra parte, la diferencia de tiempo entre paquetes en tiempo de emisión y recepción es como mucho dos veces la resolución del reloj de las tarjetas, es decir  $\pm 120$  ns, obteniendo el mismo tiempo tanto en emisión como en recepción el 81 % de las veces.

Una vez validada la prueba diseñada, se procede a evaluar la bondad de la sincronización mediante GPS. Cabe destacar que el dispositivo GPS utilizado es de bajo coste y no está diseñado específicamente para obtener un rendimiento óptimo en cuanto a sincronización de tiempos.

Se realizaron pruebas de 4 horas en las que se enviaban 2 paquetes por segundo, o lo que es lo mismo 28800 muestras. Una vez obtenidos los valores de retardo de los paquetes, se les resta el valor del *retardo calibrado* para obtener una medida del error. En la Fig. 3 se pueden apreciar los resultados obtenidos, en los que se obtiene un error promedio es de 750 ns con una desviación estándar de 140 ns.

En general es esperable que el desplazamiento entre los relojes sea del orden de decenas o centenas de nanosegundos. Algunas razones para ello son:

- Las prestaciones del dispositivo GPS.
- Diferentes longitudes de los cables de conexión con el GPS. No es extraño que pueda haber diferencias en cables de 100 metros que equivalgan a 500 ns en tiempo de propagación de la señal eléctrica.

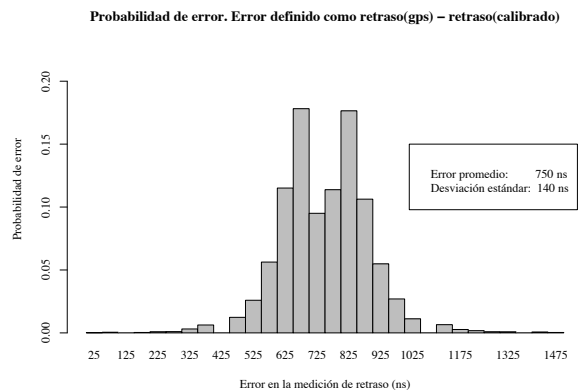


Figura 3: Histograma de la probabilidad de error utilizando GPS.

- La temperatura afecta a los relojes de cuarzo de los dispositivos electrónicos. En este caso concreto, los dispositivos GPS se encuentran en el exterior, afectándoles el clima y las tarjetas DAG pueden estar en entornos de ordenadores donde no suele haber buena ventilación. Es decir los relojes que proporcionan las marcas de tiempo de los diferentes nodos pueden tener unas derivas diferentes según la temperatura ambiente del reloj.
- El clima exterior (nieve, lluvia/humedad, nubes y temperatura) también afecta a la propagación de la señal de los satélites.

Debido a las cotas de precisión de los relojes, tanto la magnitud del desplazamiento como la de la oscilación del error no es problemático para medir el tiempo de retardo entre los nodos, ya que como se mencionó previamente tienen un retardo que está en el orden de milisegundos.

Un desplazamiento constante del error no afecta a las medidas de espaciados entre paquetes, sin embargo la oscilación del error sí. Esto último se debe a que el espaciado de los paquetes puede ser inferior a los microsegundos. A este respecto los resultados anteriores se deben interpretar como “cada 500 ms hay una oscilación con una desviación típica de 140 ns”. Es decir, en espaciados inferiores a estos 500 ms el error será proporcionalmente menor hasta llegar a la resolución del reloj (aprox 60 ns). En el caso de utilizar trenes de paquetes enviados *back-to-back* con una longitud mayor se deberá utilizar algún procedimiento estadístico para minimizar el error.

## 4 Medición del retardo *one-way* y pérdidas

### 4.1 Función de densidad

Algunas aplicaciones (Voz sobre IP, sesiones interactivas, chat, etc.) son sensibles al retardo en un sentido y a sus variaciones, así como a las

perdidas [6] [7]. ETOMIC permite realizar varios tipos de medidas del retardo, sin necesidad de técnicas de estimación [8]. Gracias a la sincronización de los agentes basta con leer en recepción la marca de tiempo introducida en cada paquete en transmisión. Estas marcas temporales las introduce por hardware la propia tarjeta Endace DAG 3.6GE.

Para realizar este tipo de medidas se han desarrollado dos programas. Uno de ellos (trafgen) manda paquetes con las características que se deseen (protocolo, puerto, ttl, etc.) al destino que se indique. Los tiempos entre 2 paquetes consecutivos se definen mediante un fichero, o indicando al programa que sigan una distribución determinada. El tamaño de los paquetes puede indicarse de la misma manera. El segundo programa (sink) filtra los paquetes que se reciben en la tarjeta DAG siguiendo las reglas que se deseen. De los paquetes que han pasado el filtro se obtiene la información necesaria.

Configuramos trafgen de forma que envíe con tiempo entre paquetes constante. El tamaño de los paquetes es constante e igual a 46 bytes. Utilizamos el protocolo UDP y los puertos de origen y destino son ambos 80.

El programa sink filtra los paquetes que no sean UDP con puertos de origen y destino 80. Entre toda la información que nos ofrece se encuentra el retardo en un sentido, calculado gracias al *timestamp* origen incluido en el paquete.

La Fig. 4 muestra la densidad de probabilidad de los retardos medidos desde un nodo en Pamplona (España) hasta otro situado en Birmingham (Gran Bretaña). Este experimento en concreto se ha realizado con un ratio constante de 150 paquetes por segundo. El tamaño de los paquetes es 46 bytes. En esta medida además de apreciar la distribución que sigue la medida del retardo se puede comprobar la precisión del sistema: nótese la escala del eje de abscisas, con cientos de puntos de medida por cada intervalo de 200ns.

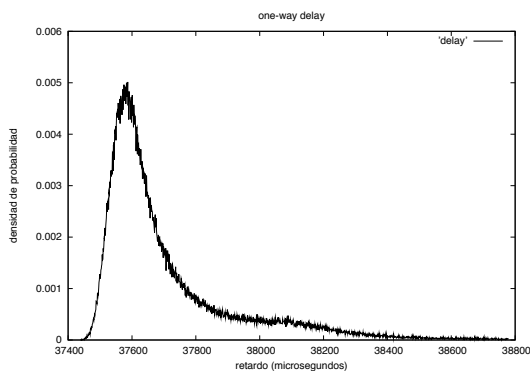


Figura 4: Retardo en un sentido Pamplona - Birmingham

Cabe destacar que el retardo no es el mismo en ambos sentidos. La Fig. 5 muestra la función densidad de probabilidad del retardo Pamplona-Birmingham en ambos sentidos, comprobándose la gran diferencia de

la media en ambas, debido posiblemente a diferente congestión en los enlaces en cada sentido.

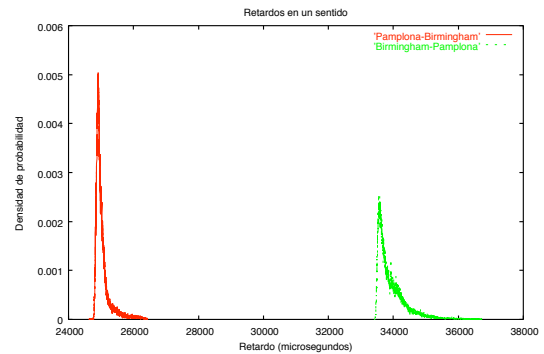


Figura 5: Retardos en ambos sentidos Pamplona - Birmingham

## 4.2 Detección de congestión

La medida de las pérdidas en un enlace puede utilizarse para detectar congestión o cuellos de botella. Para realizar esa tarea se envían dos ráfagas de paquetes utilizando las mismas aplicaciones explicadas anteriormente. Los paquetes se envían desde un nodo en Budapest ( Hungría) hasta un nodo en Jerusalem (Israel).

Transmitiendo ráfagas de paquetes a velocidades de hasta 8 Mbit/s no hay prácticamente pérdidas de paquetes y el retardo en un sentido tiene un valor constante para todos los paquetes. En la Fig. 6 se envían 2 ráfagas a una velocidad de 12 Mbit/s (representadas por la línea continua inferior en la figura). En este caso se ve claramente cómo un elemento está encolando los paquetes y haciendo que aumente su retardo. Cuando el buffer de ese dispositivo de red se llena comienza a descartar paquetes sistemáticamente (aspas sobre el valor 20 de retardo de la Fig. 6 indican paquetes perdidos) hasta que se vacía parcialmente y vuelve a cursarlos. Es decir, el mecanismo de planificación de la cola aplica cierta técnica de histéresis. En las dos ráfagas el comportamiento es idéntico y repetible en diferentes experimentos. Si además nos fijamos en la pendiente del retardo se observa cómo el cuello de botella está marcado por un tramo que está funcionando a 10 Mbit/s en el camino.

Este tipo de medidas, con ráfagas de paquetes, además de permitir la detección de congestión o cuellos de botella, sirven para obtener ciertas propiedades del camino [9]. En este caso se observa un comportamiento de un router que no corresponde con los algoritmos más habituales de descarte de paquetes [10]: FIFO, RED, CBQ, etc.

## 4.3 Efectos multicamino y del hardware de red

El protocolo IP utilizado por todos los dispositivos en Internet es un protocolo conceptualmente sencillo en el que cada paquete se reenvía de forma independiente

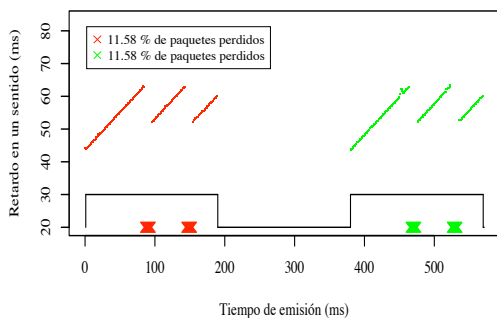


Figura 6: Ráfagas a 12 Mbps

entre los nodos intermedios hasta alcanzar el destino. IP no requiere que los paquetes lleguen de forma orientada, ni siquiera tiene cuidado de que lleguen, se dice que IP es un protocolo que no garantiza el servicio (*best effort*). TCP, encapsulado por encima de IP, es el protocolo de transporte más utilizado en Internet (más del 90 % [11]). Este protocolo es utilizado en una transmisión únicamente por el emisor y el receptor (no los nodos intermedios) y está diseñado para que el receptor avise al emisor de cuándo se producen pérdidas de paquetes. En este caso el emisor reenvía los paquetes perdidos y entiende que las pérdidas se producen por congestión de la red, por lo que baja el ritmo de emisión. Cuando se produce reordenación, aunque no se produzcan pérdidas TCP puede interpretar la reordenación como pérdidas: la conexión se vuelve más lenta porque utiliza la red de forma ineficiente (al reenviar paquetes innecesarios) y la tasa de transferencia se reduce ya que el emisor interpreta que hay congestión en la red.

Durante la realización de experimentos se comprobó que en el nodo de la universidad Hebrea de Jerusalén ([huji.ac.il](http://huji.ac.il)) solía producirse reordenación. Se puede apreciar este efecto en la Fig. 7, en la cual paquetes enviados de forma consecutiva obtienen un retardo muy diferente unos respecto a otros. Si nos fijamos en esta figura se puede comprobar cómo parece haber 4 líneas de tendencias diferentes en los retardos.

Se realizaron pruebas enviando tráfico desde todos los nodos hacia [huji.ac.il](http://huji.ac.il). Se comprobó que el efecto de reordenación mostrado en la Fig. 7, se producía con mayor o menor intensidad desde todos los nodos excepto el de la Universidad de Aston ([aston.ac.uk](http://aston.ac.uk)).

Utilizando la información de la topología de red se comprobó que el efecto no se producía por un cambio continuo en las rutas (*route flapping*, habitualmente debido a una mala configuración de los protocolos de encaminamiento) o porque se estuviera haciendo un reparto de carga de tráfico entre los diferentes caminos hacia [huji.ac.il](http://huji.ac.il).

A partir de la misma información topológica se comprobó que todos los nodos, cuando envían tráfico hacia [huji.ac.il](http://huji.ac.il), comparten el camino a partir del nodo 62.40.103.70 (nube C en la Fig. 8). Dado que

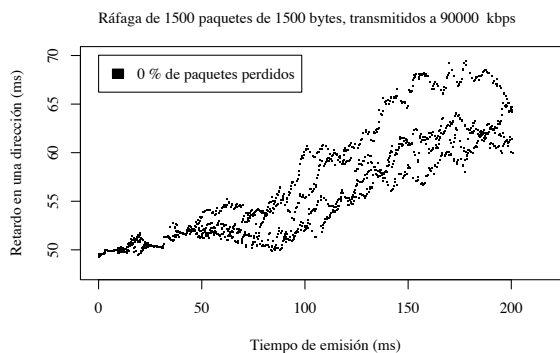


Figura 7: Incremento del retardo de los paquetes en 4 tendencias, produciendo reordenación.

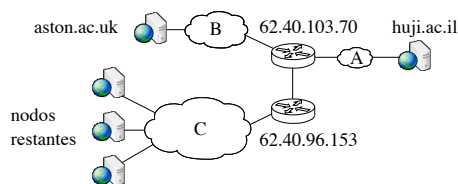


Figura 8: Topología de red de los nodos involucrados en el diagnóstico de la reordenación de paquetes hacia el nodo en Israel ([huji.ac.il](http://huji.ac.il)).

desde [aston.ac.uk](http://aston.ac.uk) no se produce reordenamiento, la causa del fenómeno debe estar en un punto intermedio compartido entre el resto de nodos y no por [aston.ac.uk](http://aston.ac.uk) (nube C). El causante puede ser cualquier salto intermedio compartido por el resto de los nodos y que esté situado entre 62.40.103.70 y éstos situados en la nube B. Afortunadamente, únicamente 62.40.96.153 (situado inmediatamente a continuación) cumplía esta condición.

Descartadas las causas de cambio de ruta, quedan las siguientes posibilidades:

- El reordenamiento de paquetes en 4 tendencias diferentes puede estar causado por el mal reparto de paquetes en un enlace lógico agregando 4 enlaces físicos (por ejemplo un OC-12 se puede componer con 4 enlaces OC-3).
- Una mala planificación de los paquetes en ese nodo.

El primer punto no es: observando el mapa de capacidades de Geant [12] el enlace es inferior a un OC-3 y este constaría de 3 (y no 4) enlaces físicos OC-1. Contactos con Dante (entidad que gestiona la red Geant) confirmaron que el enlace en cuestión se trataba de un OC-3 no agregado.

Por tanto, la posibilidad de la reordenación por la propia lógica del nodo intermedio parece correcta, ligándola en principio a una mala configuración. Éste no fué el caso: aunque la reordenación se produce en el procesamiento de los paquetes, se comprobó finalmente que la razón era que el modelo de router utilizado en 62.40.103.70 (Juniper M160) planifica los

paquetes repartidos en **cuatro** procesadores (SFMs – Switching and Forwarding Modules) que se dedican a conmutar y encaminar los paquetes entre las diferentes interfaces. En situaciones de congestión como la de nuestro caso de estudio, el retardo introducido por cada procesador es diferente. Nótese que el número de procesadores (4) corresponde con las líneas de tendencia del retardo de los paquetes.

En este ejemplo hemos podido comprobar la utilidad de la información de la topología de red. Disponer de una información histórica de la misma puede resultar útil más aún cuando puede ser normal querer comprobar información anterior a cambios realizados en la red que no han tenido el rendimiento esperado. Para tener una información topológica suficientemente completa se requiere un buen número de nodos, cuantos más mejor. En este caso concreto, el número de nodos intermedios (routers) promedio entre cualquiera de los extremos de la red es de 15 y el número de nodos desde los que se realizaron pruebas 9.

## 5 Descubrimiento de topologías y evolución temporal de rutas

El sistema ETOMIC es capaz de ejecutar cualquier software y como tal cualquier aplicación conocida como la herramienta *traceroute*, por ejemplo con el fin de comprobar la topología de la red. En un estudio realizado durante el mes de Marzo de 2005 observando las rutas entre todos los nodos del sistema, las rutas permanecen mayormente estables. Si se considera como observación las ejecuciones periódicas de traceroutes entre un par de nodos durante un día, se han apreciado cambios en las rutas en menos del 7% de las observaciones. La mayoría de estos casos corresponden a cambios en la ruta que han durado unos minutos, para después volver al camino original. Estos cambios transitorios vienen acompañados de pérdidas de los paquetes que envía traceroute en un porcentaje apreciable de los casos.

En la Fig. 9 se muestran los RTT mínimos de los saltos desde un nodo situado en Estocolmo (Suecia) hasta otro situado en Jerusalem (Israel). Se ha escogido esta gráfica porque es una de las que tiene más casos de cambios de ruta. En la gráfica se han representado medidas separadas entre sí por al menos 24 horas, junto con los cambios que se han detectado. Cuando hay un cambio en una ruta se muestran los saltos nuevos que tiene la ruta. Por ejemplo, se observan cambios momentáneos en la ruta los días 1,3,9 y 10 de Marzo. En la gráfica se ve como aparecen los saltos nuevos de la ruta, representados por un símbolo '+' recuadrado y por una 'x'. El día 12 hubo un cambio de ruta que se mantuvo hasta el día 23. En ese periodo de tiempo también se observan cambios de ruta de corta duración, especialmente entre los días 12 y 17. Los últimos días del mes, se vuelve a cambiar la ruta por la misma que al principio del mes, que se

mantiene estable hasta el final de la medida.

En la Fig. 10 se muestra el grafo de interconexión entre todos los nodos de Evergrow, o dicho de otra forma, todos los enlaces y caminos que se pueden monitorizar en un instante determinado. Esta información de topología del sistema es muy útil para diseñar experimentos por parte de los investigadores. Funcionalidades de este tipo se van a ir incorporando al interfaz web del CMS sucesivamente.

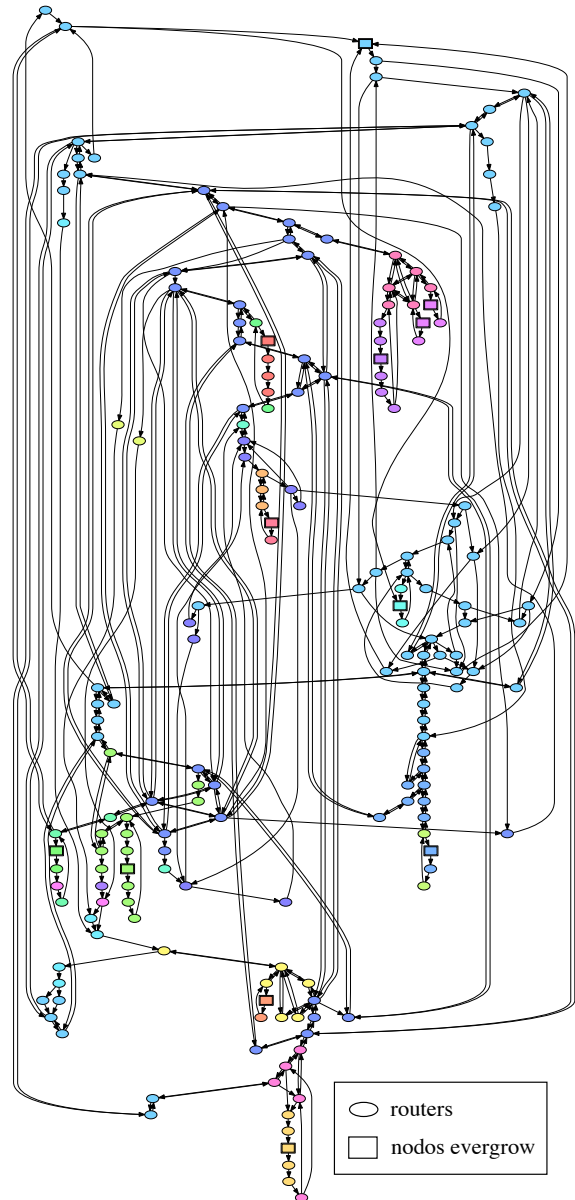


Figura 10: Grafo dirigido que representa los caminos entre todos los nodos de Evergrow. Los diferentes niveles de gris representan sistemas autónomos (AS).

## 6 Conclusiones

La plataforma ETOMIC provee de una infraestructura de monitorización de altas prestaciones, útil para realizar monitorización activa y pasiva entre nodos distribuidos por toda Europa. Las características de alta precisión del sistema abren la posibilidad

Traceroute from ericsson to huji

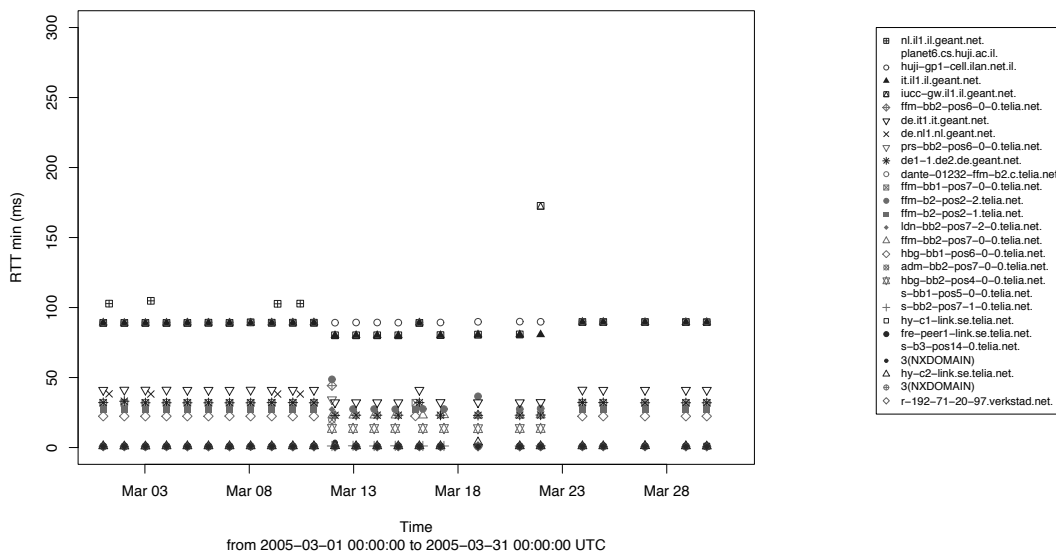


Figura 9: Cambio de rutas a lo largo del mes de Marzo de 2005

de realizar nuevas medidas activas y proponer herramientas que extraigan mayor información de la red. En este trabajo se ha estudiado el estado de enlaces de la Internet europea en base a algunas mediciones activas, observándose fenómenos de interés como la detección de mecanismos de retardo variable introducido por routers debido a su arquitectura de multiconmutación paralela. También se han presentado medidas de retardo en un sólo sentido, y algunas conclusiones que se pueden obtener directamente de ese parámetro como la detección de situaciones de congestión o estimación del ancho de banda del cuello de botella del camino entre dos nodos.

## Referencias

- [1] A. Pasztor and D. Veitch. On the Scope of End-to-End Probing Methods. *IEEE Communications Letters*, 6(11):509–511, 2002.
- [2] E.Magaña, D.Morató, M.Izal, J.Aracil, F.Naranjo, F.Astiz, U.Alonso, and et al. The european traffic observatory measurement infrastructure (ETOMIC). In *IEEE International Workshop on IP Operations & Management (IPOM 2004)*, Beijing, China, October 2004.
- [3] D.Morató, E.Magaña, M.Izal, J.Aracil, F.Naranjo, F.Astiz, U.Alonso, and et al. The european traffic observatory measurement infrastructure (ETOMIC): A testbed for universal active and passive measurements. In *Testbeds and Research Infrastructures for the Development of NeTworks and COMmunities (Tridentcom 2005). Best testbed award*, Trento, Italy, February 2005.
- [4] Endace Measurement Systems. <http://www.endace.com>.
- [5] Darryl Veitch, Satish Babu, and Attila Pasztor. Robust synchronization of software clocks across the internet. In *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 219–232, New York, NY, USA, 2004. ACM Press.
- [6] G. Almes, S. Kalidini, and M. Zekauskas. RFC2679: A one-way delay metric for IPPM, September 1997.
- [7] Y. Kitatsuji, M.Tsuru, S. Katsuno, and Y. Oie. Usefulness of precise time-stamping for exposing network characteristics on high-speed links. volume 5598, pages 163–170, October 2004.
- [8] O. Gurewitz and M. Sidi. Estimating one-way delays from cyclic-path delay measurements. In *Proceedings of Infocom, 2002*, 2002.
- [9] C. Dovrolis, P. Ramanathan, and D. Moore. Packet dispersion techniques and capacity estimation. *IEEE/ACM Transactions on Networking*, 12(6), 2004.
- [10] B. Braden, D. Clark, and et al. RFC2309: Recommendations on queue management and congestion avoidance in the internet, April 1998.
- [11] Richard Nelson, Daniel Lawson, and Perry Lorier. Analysis of long duration traces. *SIGCOMM Comput. Commun. Rev.*, 35(1):45–52, 2005.
- [12] Geant. Network topology. URL, December 2004. [http://www.geant2.net/upload/pdf/GEANT\\_Topology\\_12-2004.pdf](http://www.geant2.net/upload/pdf/GEANT_Topology_12-2004.pdf).