

Mejora de la Calidad en un Sistema de Telefonía IP Mediante el uso de Técnicas de Multiplexión

José M^a Saldaña, Julián Fernández-Navajas, José Ruiz-Mas, Jenifer Murillo, José I. Aznar, Eduardo Viruete Navarro, Luis Casadesus
Grupo de Tecnologías de las Comunicaciones – Instituto de Investigación en Ingeniería de Aragón
Dpt. IEC. Centro Politécnico Superior Universidad de Zaragoza
Edif. Ada Byron, 50018, Zaragoza
{jsaldana, navajas, jruiz, jenifer.murillo, jiaznar, eviruede, luis.casadesus}@unizar.es

Resumen- En la actualidad muchas empresas utilizan la Voz sobre IP (*Voice over Internet Protocol*, VoIP) en sus sistemas de telefonía. Las empresas con oficinas en distintos países y áreas geográficas pueden construir un sistema de telefonía centralizado para compartir las líneas de sus sucursales y así incrementar la probabilidad de admisión y ahorrar costes en llamadas internacionales. Por tanto, es conveniente introducir un sistema que permita asegurar una Calidad de Servicio (*Quality of Service*, QoS) mínima para las llamadas. Uno de estos sistemas es el Control de Admisión de Llamadas (*Call Admission Control*, CAC). En este trabajo se estudian las mejoras en cuanto a probabilidad de admisión y calidad de la conversación (Factor R) que se pueden obtener cuando se usan técnicas de multiplexión RTP, puesto que en este escenario habrá múltiples llamadas con un origen y destino comunes. Se han realizado simulaciones para comparar el uso habitual de RTP con el del protocolo TC RTP (*Tunneling Multiplexed Compressed RTP*). Los resultados cuantifican la mejora que se obtiene usando multiplexión en términos de probabilidad de admisión y calidad en la conversación.

Palabras Clave- telefonía IP, Factor R, VoIP, QoS, centralita software, multiplexión RTP

I. INTRODUCCIÓN

En los últimos años muchas empresas están sustituyendo sus antiguos sistemas de telefonía basados en Red Telefónica Conmutada (RTC) por otros nuevos que usan IP, de manera que las llamadas y las videoconferencias se puedan establecer a través de redes de datos. Uno de los objetivos es la reducción de costes, que se puede lograr aprovechando las conexiones a Internet existentes, para transmitir vídeo, voz y datos.

Muchas de estas empresas tienen sus recursos descentralizados, de manera que cada oficina o sucursal es independiente del resto y se encarga de gestionar su conexión a Internet y líneas telefónicas. Si se realizase una gestión centralizada de los recursos globales de la empresa, se podrían ahorrar costes realizando parte de las llamadas a través de Internet, manteniendo, e incluso aumentando, la Calidad de Servicio (*Quality of Service*, QoS) mediante el control de la probabilidad de admisión de las llamadas y de sus parámetros de calidad.

Como solución centralizada, se puede usar una centralita (*Private Branch eXchange*, PBX) software para unir los sistemas de telefonía de cada oficina, creando un sistema que cuenta con varias ventajas, como compartir las líneas de diferentes sucursales. Esto mejoraría los costes, ya que las llamadas internacionales se podrían establecer en dos tramos:

uno utilizando Internet, hasta el país destino, y otro con llamada local desde la sucursal en ese país hasta el usuario final.

La Voz sobre IP (*Voice over Internet Protocol*, VoIP) es un servicio en tiempo real, que en muchos casos utiliza una red que fue diseñada para servicios *best-effort*. Sin embargo, los usuarios desean una calidad similar a la que acostumbran a tener utilizando los sistemas de telefonía RTC tradicionales. Este hecho ha llevado a los investigadores a buscar soluciones para añadir calidad a las redes IP. Una de las más utilizadas es el Control de Admisión de Llamadas (*Call Admission Control*, CAC), que acepta o rechaza las nuevas llamadas para evitar la degradación del servicio. Más específicamente, en este trabajo se usará el CAC basado en parámetros [1], que cuenta el número de llamadas que hay establecidas simultáneamente, y solamente acepta las nuevas llegadas si ese número está por debajo de un límite marcado por la calidad mínima aceptable para una llamada.

En el presente trabajo estimaremos la calidad de las conversaciones mediante el Factor R, propuesto por la norma ITU G.107 [2]. El Factor R va desde 0 hasta 100, y los valores que se consideran aceptables son $R > 70$. Se puede convertir fácilmente en MOS (*Mean Opinion Score*), que va desde 0 (mala calidad) hasta 5 (muy buena).

En el escenario propuesto ocurrirá con frecuencia que varias llamadas tendrán el mismo origen y destino (Fig. 1). Por tanto, el uso de técnicas de multiplexión puede suponer una mejora. Si introducimos en el mismo paquete IP las muestras de diferentes llamadas, podremos ahorrar ancho de banda añadiendo pequeños retardos. En este trabajo usaremos el RFC 4170: TC RTP (*Tunneling Multiplexed Compressed RTP*) [3].

Para poder estudiar los beneficios de los sistemas de telefonía distribuidos, se construyó un escenario diseñado en trabajos previos [4], similar al de una empresa con sucursales

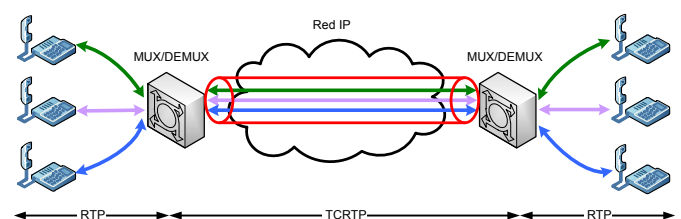


Fig. 1. Llamadas que comparten las mismas sucursales origen y destino

en diferentes áreas o países. En sus primeras etapas se implementó mediante emulación en un *testbed*, pero este hecho introducía una limitación de tamaño causada por el número de máquinas que se requerían.

Para solucionar este problema se ha usado simulación, que puede ayudarnos a estudiar el comportamiento del sistema cuando las líneas se comparten entre un número de sucursales y se usa multiplexión RTP. Para evaluar los resultados nos centraremos en la probabilidad de admisión y el Factor R. Como parámetros de simulación se usarán los resultados obtenidos previamente en el *testbed*, especialmente retardos y pérdidas de paquetes, utilizando por tanto un método de pruebas híbrido, que integra simulación y emulación.

El presente trabajo está organizado de la siguiente manera: la sección II trata sobre los trabajos relacionados. La arquitectura del sistema se presenta en la sección III. La siguiente sección detalla la plataforma utilizada. La sección V presenta las pruebas que se han realizado y los resultados. Las conclusiones cierran el trabajo.

II. TRABAJOS RELACIONADOS

En la actualidad, en entornos empresariales se está tendiendo a introducir nuevas soluciones como VoIP, en parte buscando disminuir costes. En [5] Intel publicó los resultados de un programa piloto en el que un grupo de empleados utilizó VoIP basada en SIP (*Session Initiation Protocol*) durante unos meses. La conclusión fue que esta tecnología es beneficiosa para las empresas, en términos de costes y también de productividad. En [6] se puede encontrar otro estudio que ilustra las mejoras obtenidas al usar VoIP en lugar de telefonía tradicional. Se obtenían ahorros en costes de equipos, aprovisionamiento, facturación, mantenimiento y servicio, y se recomendaba VoIP como la nueva solución de telefonía para las empresas.

Los sistemas CAC se pueden clasificar en dos categorías [7]: por un lado, los basados en medidas toman sus decisiones en función del estado de la red; por otro, los basados en parámetros requieren la realización de una serie de medidas durante la puesta en marcha del sistema, para obtener los parámetros que luego regirán su funcionamiento, como el número máximo de llamadas simultáneas.

Respecto a las políticas del *buffer* de salida de los *router* en el sistema, podemos decir que en los últimos años la regla usada tradicionalmente para dimensionarlos era el uso del producto del ancho de banda por el retardo. Pero esta regla ha sido cuestionada por el llamado "*Stanford model*". En [8] se presentó una comparativa y se sugirió el uso de un *buffer* limitado en tiempo. En este trabajo usaremos esta política, ya que es muy adecuada para mantener los retardos por debajo de una cota superior.

III. ARQUITECTURA DEL SISTEMA

Probaremos el sistema en modo *original*, en el que cada flujo RTP es independiente del resto, y en modo *multiplexión*, que multiplexa los flujos usando TCRTTP. El escenario se corresponde con el de una empresa con sucursales en varias áreas geográficas. Este escenario es el equivalente al de algunas soluciones comerciales [9].

Como puede verse en la Fig. 2, cada oficina tiene un conjunto de usuarios, y se conecta a RTC mediante una pasarela o *gateway*. También dispone de un acceso a Internet,

y las llamadas de VoIP basadas en SIP son controladas por un sistema CAC. En el centro de datos hay una PBX *software* que también está conectada a la red IP. Se utiliza Internet para el tráfico telefónico, evitando de esta manera los costes de las líneas dedicadas.

En cada sucursal existe un agente local que incluye un *proxy* SIP que implementa las decisiones de admisión de llamadas. Se realizan una serie de medidas durante la puesta en marcha del sistema, y como resultado se asigna un número máximo de llamadas simultáneas permitidas a cada sucursal.

Como se ha comentado anteriormente, se usa la técnica de multiplexión TCRTTP. La Fig. 3 presenta el esquema de un paquete multiplexado. En primer lugar se usa ECRTTP para comprimir las cabeceras IP/UDP/RTP. Posteriormente, se utiliza PPPMux para multiplexar varios paquetes dentro de otro, que se envía usando un túnel L2TP. ECRTTP se usa extremo a extremo porque, al usar un túnel, no es necesario descomprimir y volver a comprimir las cabeceras en cada salto.

Cuando se usa el modo *multiplexión*, el ancho de banda ocupado por las llamadas será menor que el que ocuparían si se enviasen por separado. El sistema CAC debe tener esto en cuenta para tomar las decisiones de admisión, puesto que ya no puede contar las llamadas de la misma manera que en el modo *original*. Nuestra primera idea fue contar el ancho de banda de cada túnel según el número de llamadas que contiene, con lo que el sistema CAC limitaría el ancho de banda de cada sucursal en lugar del número de llamadas. De acuerdo con la Fig. 3, podemos obtener el ancho de banda de un túnel con k llamadas de la siguiente manera:

$$E[BW_k] = E[PS_k] / IPT = (CH + k (MH + E[RH] + S)) / IPT \quad (1)$$

Donde:

- **CH:** Cabecera Común (*Common Header*): Corresponde a las cabeceras IP/L2TP/PPP, que ocupan 25 bytes.
- **MH:** Cabecera de Multiplexión (*Multiplexing Header*). Es la cabecera PPPMux (2 bytes).

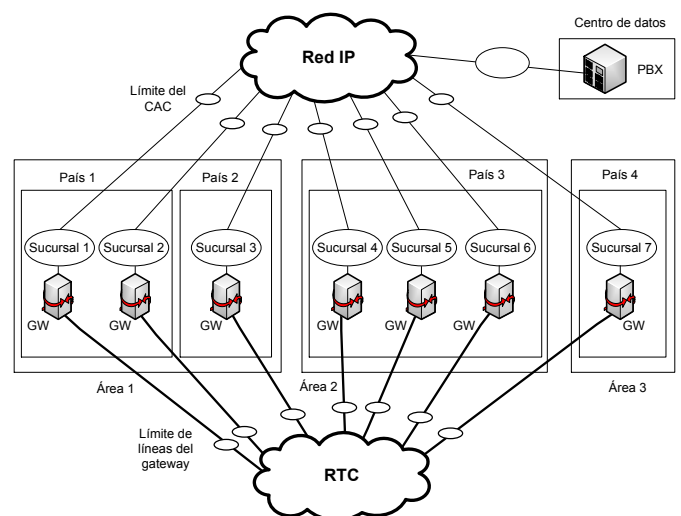


Fig. 2. Arquitectura del sistema



Fig. 3. Esquema de un paquete multiplexado

- **RH:** Cabecera Reducida (*Reduced Header*). Precede a las muestras de cada flujo RTP. Los protocolos de compresión producen cabeceras comprimidas de diferentes tamaños, por lo que $E[RH]$ se calculará según la probabilidad de obtener una cabecera de cada tamaño [10].
- **S:** Tamaño de las muestras (*Samples*). Tamaño en bytes de las muestras que lleva cada paquete RTP. En nuestro caso tendrá un valor de 20 bytes, al usar el *codec* G729a con 2 muestras por paquete.
- **IPT:** Tiempo entre paquetes (*Inter Packet Time*).

Podemos ver que (1) crece linealmente con k , por eso finalmente se ha decidido que el sistema CAC cuente el número de llamadas establecidas, independientemente del ancho de banda de los túneles.

En trabajos previos este sistema se probó en un *testbed* utilizando emulación en tiempo real, pero en este trabajo utilizaremos también simulación. Una ventaja de la simulación es que el número de sucursales del escenario puede ser mayor, como ocurría en [1]. Por otro lado, la emulación en tiempo real permite obtener otros valores, como el Retardo en un Sentido (*One Way Delay*, OWD) y las pérdidas de paquetes, necesarios para conocer el máximo número de llamadas que se pueden establecer simultáneamente con una calidad aceptable. Con estos parámetros se puede calcular el Factor R utilizando el E-Model [2].

IV. PLATAFORMA DE PRUEBAS

En primer lugar, hemos utilizado el *testbed* presentado en [11] para implementar el escenario. Se usa un sistema híbrido para las pruebas, combinando emulación y simulación. Lo explicamos con más detenimiento a continuación. Una prueba se divide en las siguientes etapas, como se ve en la Fig. 4:

- Emulación, en la que se usan tres máquinas. En primer lugar, el generador de tráfico envía paquetes RTP y tráfico de fondo. Después, estos tráficos atraviesan el *router*, que emula diferentes políticas de *buffer*. Finalmente, una máquina recibe y almacena el tráfico.
- Procesado *offline*, en el que se añaden algunos retardos al tráfico capturado en la fase anterior.
- Simulación del escenario global y resultados finales. Se usa Matlab para simular el escenario y se utilizan los resultados anteriores para obtener la probabilidad de admisión y el Factor R de las llamadas.

A. Red de emulación

Se ha usado el *testbed* para implementar una versión reducida del sistema, y así obtener los parámetros que determinan la QoS, y que serán incluidos después en las simulaciones. Se utiliza un generador para enviar tráfico de fondo y saturar de distintas maneras el acceso de cada sucursal. Para limitar el ancho de banda del acceso se utiliza la herramienta Linux *traffic control* (*tc*).

B. Procesado *offline*

Se han añadido al sistema los siguientes retardos: el de paquetización, que depende del *codec*; el tiempo de retención en el multiplexor, que se debe a que los paquetes deben esperar hasta que han llegado todos los paquetes a multiplexar. El retardo de encolado depende de la política del

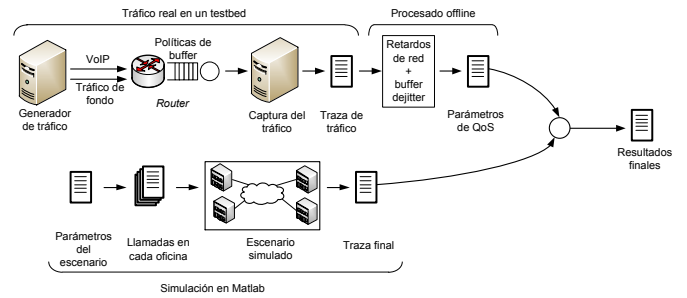


Fig. 4. Diagrama del sistema de medidas

buffer que se esté usando. También se ha incluido un *buffer de de jitter* en la máquina destino, que añade nuevos retardos.

C. Simulación

Se ha utilizado Matlab para generar realizaciones con diferentes parámetros: número de oficinas, usuarios, líneas de los *gateway*, países, zonas geográficas, retardos de establecimiento, etc. Las llamadas se generan según una distribución de Poisson. Su duración se ha modelado con una distribución Normal de media 180 seg. y varianza 30.

Posteriormente, cada realización se puede simular utilizando dos algoritmos. El primero, denominado *aislado* se usa como referencia. No implementa el sistema CAC, y cada sucursal es independiente del resto, por lo que el sistema no comparte los *gateway* y no se redirigen llamadas. El segundo algoritmo, denominado *compartido*, simula un sistema centralizado, que comparte todas las líneas de sus *gateway* entre las sucursales e implementa un sistema CAC, permitiendo al agente local redirigir llamadas a otras oficinas, buscando siempre la tarifa más barata.

El modo *compartido* se usa para intercambiar la probabilidad de bloqueo en los *gateway* por probabilidad de bloqueo en el acceso a Internet, ya que normalmente es más barato contratar ancho de banda que aumentar el número de líneas. Por tanto, es de esperar que la probabilidad de admisión aumente a causa de las redirecciones, y que se consiga ahorrar costes en llamadas internacionales. Sin embargo, incrementar la probabilidad de admisión también tiene una contrapartida, y es que se introduce más tráfico en el acceso a Internet, por lo que la QoS puede verse afectada.

V. PRUEBAS Y RESULTADOS

A. Probabilidad de admisión al compartir las líneas

La Fig. 5 muestra cómo al usar el modo *compartido*, a mayor número de oficinas se consigue una mayor probabilidad de admisión. Se han incluido 25 usuarios por sucursal. Esto se corresponde con la idea de la fórmula de Erlang de que se consigue más probabilidad de admisión cuantos más recursos se comparten. Naturalmente, si la tasa de llamadas por hora y usuario λ se hace muy grande, la probabilidad de admisión se reduce.

Hemos realizado otras pruebas variando el límite del CAC. En la Fig. 6 se puede ver que el modo *compartido* da mejores resultados para la probabilidad de admisión que el modo *aislado*, por lo que se confirma que compartir las líneas es beneficioso. Por otro lado, vemos que según aumenta el límite del CAC, en el modo *compartido* aumenta la probabilidad de admisión, mientras que en el modo *aislado* no varía.

B. Factor R

Se ha configurado un escenario con cuatro zonas en el mismo país, y una sucursal en cada una. Existen 25 usuarios en cada oficina. Se ha usado un tráfico de fondo que satura el acceso en un 80%. Los parámetros variables son la tasa de generación de llamadas y el límite del CAC de las sucursales.

La Fig. 7 compara los dos modos en términos de Factor R en función del límite del CAC. En primer lugar, puede verse que la multiplexión supone una mejora para los mismos valores de λ y del límite del CAC. Este hecho se debe a que la cantidad de tráfico enviado en el modo *multiplexado* es menor que en el *original*. Por tanto, usando multiplexión, el límite del CAC puede ser mayor si se desea obtener el mismo valor del Factor R.

Por otra parte, si mantenemos λ fija, la figura muestra que a mayor límite del CAC, peor será la calidad en términos de R. Obviamente, incrementar el número de llamadas simultáneas sin aumentar el ancho de banda hará que la calidad disminuya.

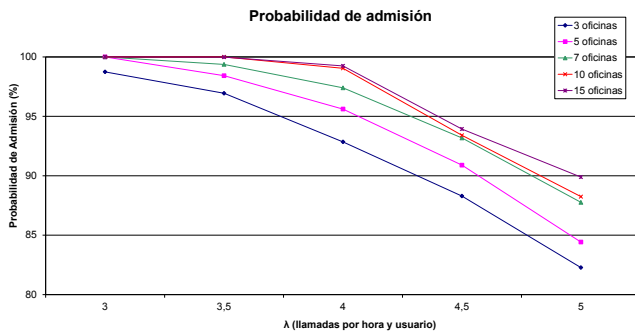


Fig. 5. Probabilidad de admisión en modo *compartido*

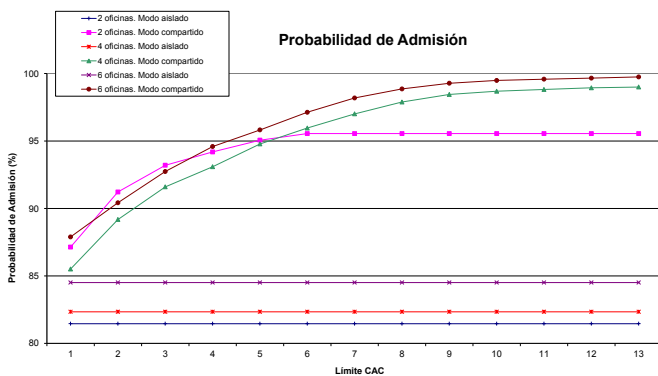


Fig. 6. Probabilidad de admisión en modos *aislado* y *compartido*

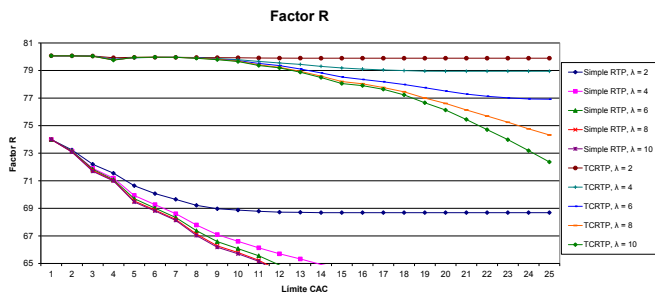


Fig. 7. Factor R en modos *original* y *multiplexado* en función del límite del CAC. No se representan los valores bajos del Factor R, porque se considera inaceptable por debajo de 70

Observamos que existe un compromiso entre la probabilidad de admisión y la calidad de las llamadas. Por tanto, se podría sacrificar parte del margen del Factor R que nos aporta la multiplexión, para lograr así mejores valores de probabilidad de admisión.

VI. CONCLUSIONES

En este trabajo se ha estudiado un sistema de telefonía basado en VoIP y que usa el protocolo SIP. El escenario se corresponde con el de una empresa con oficinas en diferentes zonas. Se ha incluido un sistema CAC que limita el número simultáneo de llamadas. Se ha usado un sistema híbrido de pruebas, obteniendo en un *testbed* en primer lugar los parámetros de QoS, y después usándolos como entradas para las simulaciones del escenario completo.

Por un lado, se quería comprobar si compartiendo las líneas de las sucursales es posible mejorar la probabilidad de admisión. Por otro lado, queríamos comparar el comportamiento del sistema en términos de calidad de las llamadas, si se usa RTP simple o túneles TCRTTP.

Se han realizado simulaciones del sistema en diferentes situaciones. Los resultados obtenidos muestran que la probabilidad de admisión aumenta si se comparten las líneas entre sucursales. También se observan mejoras en el Factor R cuando se usa multiplexión TCRTTP, que se pueden traducir también en mejoras de la probabilidad de admisión.

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por el Proyecto CPUFLIPI (MICINN TIN2010-17298), por el Proyecto MBACToIP, de la Agencia I+D del Gobierno de Aragón e Ibercaja Obra Social, y por el Proyecto NDCIPI-QQoE de la Cátedra Telefónica, de la Univ. de Zaragoza.

REFERENCIAS

- [1] S. Wang, Z. Mai, D. Xuan, and W. Zhao, "Design and implementation of QoS-provisioning system for voice over IP," *Parallel and Distributed Systems*, IEEE Transactions on, vol. 17, no. 3, pp. 276--288, 2006
- [2] ITU-T Recommendation G.107, "E-model, a computational model for use in transmission planning," 2003
- [3] B. Thompson, T. Koren, and D. Wing, "RFC 4170: Tunneling Multiplexed Compressed RTP (TCRTTP)," 2005
- [4] J. Saldana, J. Aznar, E. Viruete, J. Fernández-Navajas, and J. Ruiz-Mas, "QoS Measurement-Based CAC for an IP Telephony System," *QShine 2009, The Sixth International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*. Las Palmas de Gran Canaria (Spain), 2009
- [5] S. Sacker, M. Santaiti, and C. Spence, "The Business Case for Enterprise VoIP," Intel Corporation, 2006
- [6] B. Athawal, "Replacing Centric Voice Services with Hosted VoIP Services: An Application of Real Options Approach"
- [7] R. Solange, P. Carvalho, and V. Freitas, "Admission Control in Multiservice IP Networks: Architectural Issues and Trends," *IEEE Communications*, vol. 45, no. 4, pp. 114--121, 2007
- [8] A. Dhamdhere, and C. Dovrolis, "Open issues in router buffer sizing," *Comput. Commun. Rev.*, vol. 36, no. 1 pp. 89--92, 2006
- [9] VoIP Call Admission Control, http://www.cisco.com/en/US/docs/ios/solutions_doc/voip_solutions/CAC.pdf
- [10] G. Dimitriadis, S. Karapantazis, F.-N. Pavlidou, "Comparison of Header Compression Schemes over Satellite Links", In *Proc. International Workshop on IP Networking over Next-generation Satellite Systems (INNSS'07)*, Budapest, Hungary, Jul 2007.
- [11] J. Saldana, E. Viruete, J. Fernández-Navajas, J. Ruiz-Mas, and J. Aznar, "Hybrid Testbed for Network Scenarios," *SIMUTools 2010, the Third International Conference on Simulation Tools and Techniques*. Torremolinos (Spain), 2010