

Estudio en rendimiento de redes en un palacio de congresos mediante simulación

Sonia Gómara, Carlos Juiz

Universidad de las Islas Baleares,
Carretera de Valldemossa km. 7.5, 07122 Palma, España

Resumen This work presents some results of the performance simulation of a model of a congress center. Different client profiles and the typical applications present in scientists congresses have been considered. Moreover, others applications of minor use have also been considered, as, for example, real-time download of the lecture presentation contents, and the mobility of users. The aim of the simulation is the study of the level of quality of service that can be offered by a defined network architecture. The simulation results clearly show the dependency of the performance on the congress timetable, due to the traffic peaks that occurred depending on the period that is being simulated. All the performance results can be easily transferred to any congress center in order to accurately dimension the corresponding networks and applications that have to be installed.

1. Introducción

Las Tecnologías de la Información y las Comunicaciones (TIC) están imbricadas en todas las actividades de la vida cotidiana. Particularmente, diversas aplicaciones basadas en Internet han dejado de ser actividades ligadas al sitio de trabajo y se han convertido en ubicuas. Así, por ejemplo, ya no se puede considerar asistir a un evento sin acceso a correo electrónico o a Internet desde los dispositivos móviles inalámbricos de los asistentes. Este fenómeno de utilización de las aplicaciones en cualquier sitio, de cualquier forma, irá sin duda creciendo hasta la total interoperabilidad de los dispositivos, las redes y las aplicaciones de forma transparente. Evidentemente, esto ha producido no pocos problemas a los administradores de sistemas desde el punto de vista no funcional, por ejemplo en seguridad, pero también en rendimiento.

Por otra parte, la necesidad de una mayor diversificación en el negocio turístico ha producido que aumente el número de eventos, que se celebran en grandes espacios, comúnmente denominados palacios de congresos o recintos feriales. Hasta hace pocos años, los asistentes a dichos eventos en esos espacios sólo demandaban, como mucho, cierta disponibilidad a ordenadores conectados a Internet en salas específicas. Sin embargo, el auge de la tecnología wifi ha producido que en un evento, por pequeño que sea, se demande conectividad inalámbrica segura, fiable y rápida. Desgraciadamente, todos los que hemos organizado o

asistido a eventos de cierta dimensión hemos experimentado que el rendimiento de la conectividad varía enormemente durante el desarrollo de los mismos.

En este artículo se estudia mediante simulación el rendimiento de un sistema en un recinto de congresos de nueva generación, caracterizado por la movilidad de los usuarios y la convivencia de aplicaciones tradicionales con nuevas aplicaciones de ambiente inteligente. El objetivo que se persigue es determinar cuáles son los requerimientos indispensables para que el sistema pueda soportar este tipo de tráfico, proporcionando la calidad de servicio adecuada. Este objetivo es especialmente oportuno en el momento en que se está construyendo un palacio de congresos en Palma de Mallorca, lugar caracterizado por un turismo “de sol y playa”, que pretende diversificar su turismo hacia clientes no estacionales y de mayor poder adquisitivo.

El artículo se desarrolla de la siguiente manera: en la sección 2 se describe el modelo de sistema creado, en las secciones 3 y 4 se especifican los parámetros de configuración, del modelo y la simulación, que se han aplicado y, a continuación en la sección 5, se analizan los principales resultados obtenidos en la simulación de este modelo.

2. Modelo del sistema

El modelo diseñado reproduce un entorno de un palacio de congresos donde una serie de clientes móviles se desplazan a través de diversas áreas, mientras ejecutan un conjunto de aplicaciones y acceden a los recursos del sistema. Por clientes móviles, entenderemos cualquier dispositivo fácilmente transportable (teléfonos móviles, PDAs u ordenadores portátiles), equipado con tecnología inalámbrica, que permita ejecutar aplicaciones basadas en TCP/IP y cuya función sea representar a las personas físicas asistentes al congreso.

Combina una red wireless formada por clientes móviles y puntos de acceso (APs), y una red cableada constituida por ordenadores de sobremesa, servidores y elementos de interconexión. En este entorno, los puntos de acceso sirven como elemento de unión entre ambas redes, permitiendo que los clientes móviles accedan a los recursos conectados a la red cableada.

En el nivel de aplicación, los clientes combinan la ejecución de aplicaciones tradicionales, como son el correo electrónico o la navegación web, con nuevas aplicaciones de ambiente inteligente, como el envío personalizado de noticias o aplicaciones de chat entre clientes próximos. El modelo dispone de dos niveles: WAN externa y LAN interna.

La figura 1 muestra el nivel WLAN del modelo: la red del recinto se conecta al router de su proveedor de servicios mediante un enlace de fibra óptica de Gigabit Ethernet (1000BaseX). A su vez, el ISP se conecta a Internet a través de una conexión de fibra óptica SONET OC-12. Por otro lado, se han definido dos servidores específicos para atender las peticiones externas de web y correo electrónico, respectivamente. Estos servidores se conectan a través de enlaces de fibra de 1000BaseX a un router, que dispone de un enlace SONET OC-12 con la nube IP de Internet.

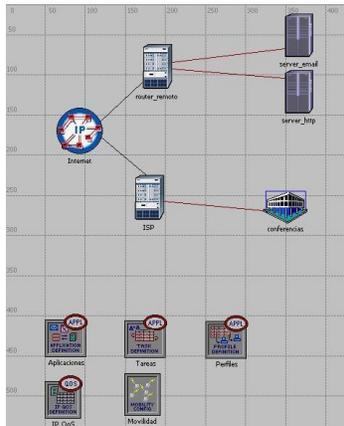


Figura 1. Nivel exterior del modelo (WAN)

La LAN del recinto, mostrada en la figura 2, se conecta al exterior a través de un *firewall*, que actúa como cortafuegos y router. Todos los elementos de esta red, exceptuando los clientes móviles, se encuentran conectados a un único switch central, a través de enlaces Fast Ethernet 100BaseT. Se han definido 5 áreas: hall, secretaría, sala PCs, sala conferencias 1 y sala conferencias 2. Un cliente móvil activo estará situado en todo momento en una de estas zonas, hecho que determinará qué aplicaciones puede ejecutar y a qué recursos del sistema puede acceder. Cada zona dispone de un punto de acceso a través del cual se conectan todos sus clientes móviles a la red cableada, para poder acceder a los recursos del sistema o a Internet. En la secretaría se han definido dos estaciones de trabajo y dos servidores de bases de datos para usuarios y noticias. En las salas de conferencias, se han dispuesto sendos servidores web y de ficheros.

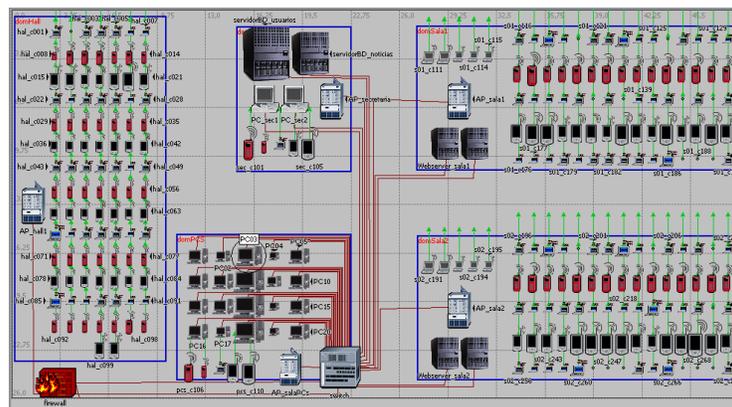


Figura 2. Nivel interno del modelo (LAN)

3. Parámetros del modelo a simular

La simulación se ha dividido en 8 etapas distintas, siguiendo un horario ficticio de una jornada de un congreso. Distinguimos periodos de descanso, donde no tienen lugar conferencias, y periodos de actividad; las acciones de los usuarios dependerán de la etapa que se esté simulando.

Para caracterizar el comportamiento de los diferentes grupos de clientes presentes en este tipo de eventos, se han definido 4 tipos de usuarios: genéricos, oyentes, ponentes y fijos, correspondiendo los tres primeros a clientes móviles y el último a los usuarios de los PCs de sobremesa. Cada tipo de clientes podrá ejecutar únicamente un determinado conjunto de aplicaciones. Los clientes móviles serán de un tipo u otro dependiendo del área donde se encuentren y del papel que tengan asignado.

Se han especificado 6 tipos de aplicaciones: navegación web, correo electrónico, noticias, chat, registro, envío de fichero y petición de fichero. Para las diferentes aplicaciones, se han definido dos versiones representando tráfico de alta y baja carga; los clientes ejecutarán una u otra versión atendiendo a la naturaleza de la fase de la simulación. La tabla 1 resume los principales parámetros de interés de las distintas aplicaciones.

Tabla 1. Principales parámetros de las aplicaciones

Aplicación	Parámetro	Valor
E-mail	Tamaño (KB)	Lognormal(7)
Registro	Tamaño (KB)	Constante(10)
Autenticación	Tamaño (KB)	Constante(1)
Envío de fichero	Tamaño (KB)	Pareto(500)
Petición de fichero	Tamaño (KB)	Pareto(500)
Noticias	Tamaño (KB)	Constante(1)
Chat	Datos (KB) x2	Pareto(40)
	Tiempo entre peticiones (seg)	Exponencial(30)
	Tiempo procesamiento (seg)	Exponencial(30)

De acuerdo con [5], se han definido 3 aplicaciones distintas para caracterizar los 3 diferentes tipos de tráfico web: estático, dinámico y seguro. Las peticiones web de los clientes de PCs o por los clientes móviles situados en las zonas comunes son atendidas por el servidor externo. En cambio, las generadas por los clientes móviles de las salas son respondidas con una probabilidad del 75 % por el servidor específico situado en la propia sala, o en un 25 % por el servidor web externo. La tabla 2 resumen los principales parámetros de estas aplicaciones web.

Tabla 2. Parámetros de las aplicaciones web

	Estática	Dinámica	Segura
Page Interarrival Time (seg)	Exponencial(540)	Exponencial(60)	Exponencial(300)
Exponencial(720)	Exponencial(120)	Exponencial(600)	
Componentes Página			
HTML (KB)	Lognormal(30)	Lognormal(30)	Lognormal(30)
CSS (KB)	Lognormal(15)	Lognormal(15)	Lognormal(15)
Script (KB)		Lognormal(70)	Lognormal(100)
Nº objetos	Pareto(40)	Pareto(40)	Pareto(40)
Tamaño/objeto (KB)	Lognormal(4)	Lognormal(4)	Lognormal(4)

4. Parámetros de la simulación

Para diseñar y simular nuestro modelo de sistema se ha utilizado la herramienta OPNET Modeler® [1], un simulador basado en eventos utilizado en el estudio y diseño de redes. Se ha definido una única simulación de 12 horas de duración, representando una jornada ficticia de un congreso. La figura 3 muestra la división del horario en periodos de conferencias (C) y de descanso (D).

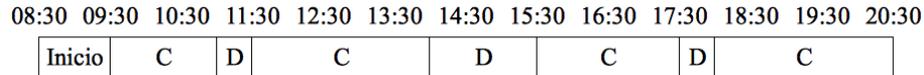


Figura 3. Horario ficticio en que se basa la simulación

5. Resultados de la simulación

Como ya se ha comentado en los puntos anteriores, el principal factor que determina el tráfico presente en el modelo radica en la naturaleza móvil de los usuarios, hecho que determina que la mayor parte del tráfico sea de origen inalámbrico y, que al mismo tiempo, condiciona los patrones de tráfico existentes, dependientes de las trayectorias de movimiento y de la actividad de los usuarios.

Para poder evaluar si este modelo satisface los requerimientos necesarios para proporcionar la calidad de servicio demandada por los clientes, es importante analizar la relación entre el volumen de tráfico que circula por el sistema y los tiempos de respuesta y retardo experimentados por los usuarios en la ejecución de las diferentes aplicaciones.

5.1. Comportamiento usuarios

Para evaluar el rendimiento del sistema, en primer lugar es necesario comprender los flujos de tráfico existentes; el movimiento de los nodos y los patrones

de actividad de los clientes pueden conducir a un comportamiento no esperado del modelo. En la gráfica 4(a) se muestra el tráfico, en bytes/segundo, generado por algunas de las aplicaciones definidas.

La aplicación que genera un mayor tráfico de datos es la aplicación de petición de fichero, ya que comporta la transmisión de mayores flujos de información. Al inicio de las distintas etapas, principalmente de aquellas en que tienen lugar las conferencias, se produce un incremento en el tráfico de datos asociado a esta aplicación, ya que es cuando los clientes móviles situados en las salas de conferencias desean acceder a la información. En cambio, la aplicación de envío de fichero no experimenta estas oscilaciones tan pronunciadas ya que las subidas se van realizando a lo largo de la jornada, si bien es cierto que hay un ligero incremento al inicio de las diversas etapas. De cualquier manera, el tráfico producido por esta segunda aplicación es considerablemente menor al de la aplicación de petición, ya que el número de clientes que la ejecutan es 15 veces menor.

Si continuamos con el análisis del tráfico, observamos como el tráfico enviado por la aplicación de noticias es muy reducido; al utilizar tecnología multicast posibilitamos que para cada noticia validada, únicamente tenga que enviarse una copia a cada uno de los 5 grupos multicast, en lugar de a cada uno de los nodos móviles activos. Así, en lugar de tener que enviar 150 KB de información (considerando un tamaño constante de noticia de 1 KB y una media de 150 clientes activos), únicamente es necesario enviar 5 KB.

Por otro lado, la aplicación de registro origina una mínima fracción del tráfico global, ya que implica un escaso intercambio de información y se ejecuta un número limitado de veces. Este tráfico se concentra durante las primeras etapas, cuando se produce el registro de la mayoría de asistentes; únicamente, después del descanso de la comida, se observa un ligero incremento, debido al registro de los asistentes que acuden exclusivamente a las conferencias vespertinas.

Hemos visto como el horario del congreso condiciona absolutamente los patrones de tráfico registrados; dependiendo de la actividad que se esté llevando a cabo, las fuentes de tráfico se concentrarán en una determinada área y se deberán, en diferente medida, a la ejecución de una u otra aplicación. Ante esta situación, se pueden adoptar dos aproximaciones: sobredimensionar el sistema, para que pueda hacer frente a estos niveles extraordinarios de tráfico; o aplicar técnicas que permitan priorizar el tráfico de determinadas aplicaciones para que, en los instantes de congestión, se prioricen unos flujos de datos, en detrimento de los retardos experimentados por otras aplicaciones.

5.2. Tiempos de respuesta

En este proyecto, el estudio del rendimiento se centra en la perspectiva de los usuarios, esto es, la medida del rendimiento vendrá dada por los tiempos de respuesta que éstos experimenten en la ejecución de las distintas aplicaciones.

Para evaluar la adecuación de los tiempos obtenidos, se ha utilizado como base comparativa la norma ITU G.1010 [6] que define un modelo de QoS para las aplicaciones multimedia desde el punto de vista del usuario. Así, el usuario

de este sistema priorizará la ejecución de las aplicaciones relacionadas con la transferencia de ficheros y el registro en el sistema.

Para el modelo desarrollado, de acuerdo a los parámetros de simulación definidos, se satisfacen todos los requerimientos relacionados con el tiempo de respuesta. Para ilustrarlo, la gráfica 4(b) muestra el cumplimiento de la SLA definida sobre las diversas aplicaciones de navegación web, considerando un tiempo de respuesta por página menor a 4 segundos, con un cumplimiento del 90 % sobre intervalos de 30 minutos. En la parte superior vemos el valor medio por intervalo, mientras que la barra vertical muestra los máximos y mínimos alcanzados en cada periodo. En la parte inferior, la línea punteada muestra el valor umbral definido, y podemos ver como las barras siempre se encuentran por encima de ese umbral, reflejando el cumplimiento del mismo.

5.3. Retardo

Al evaluar los tiempos de respuesta, hay que considerar que éstos son la suma de los tiempos de servicio, ajenos a este estudio, y de los diferentes tiempos de retardo experimentados. En este modelo, se distinguen dos fuentes principales de retardo: TCP, ya que la mayoría de las aplicaciones utilizan este protocolo de transporte; y WLAN, debido a que la mayor parte del tráfico es de origen inalámbrico. A continuación, se estudia la raíz de estos dos tipos de retardo.

En la gráfica 4(c) se muestran los valores de retardo de paquete recibido para todas las capas TCP definidas. Son valores relativamente constantes, pero los menores tiempos se producen en las etapas de descanso. Estos tiempos de retardo son mucho mayores que los producidos a nivel de WLAN y, dado que la mayoría de aplicaciones utilizan TCP como protocolo de transporte, las optimizaciones del sistema deberían ir enfocadas a reducir estos tiempos.

Como ya se ha comentado, en nuestros días, los asistentes a este tipo de eventos demandan una conexión inalámbrica segura, fiable y rápida que les permita acceder a Internet, así como a los diferentes recursos dispuestos en la red del sistema al que se conectan, ya que en la mayoría de casos los usuarios son personas que requieren estar permanentemente conectados al mundo. En la gráfica 4(d) podemos ver los tiempos de retardo producidos en el nivel inalámbrico. Podemos ver como los instantes de mayor nivel de tráfico WLAN coinciden con el inicio de los periodos de descanso, ya que es menor la varianza en los tiempos de inicio de las aplicaciones, con lo que se producen un gran número de ejecuciones en un breve intervalo. Estos *delays* se producen porque al inicio de estos intervalos, una gran cantidad de nodos acceden simultáneamente al medio inalámbrico, lo que se traduce en un aumento del tamaño de las colas de procesamiento y, con ello, de los tiempos de retardo. En esta gráfica se ve el retardo global (línea superior) y el retardo de acceso al medio inalámbrico (línea inferior); el retardo global se debe principalmente a las esperas para acceder al medio, que se encuentra saturado, lo que conduce a que los usuarios experimenten elevados tiempos de contención antes de poder iniciar la transmisión de datos.

5.4. Cuello de botella

En el análisis de rendimiento resulta indispensable determinar cuál es el cuello de botella del sistema, ya que es este elemento el que deteriora, en primera instancia, el rendimiento del mismo. Como era de prever, en este modelo el cuello de botella lo constituyen los puntos de acceso, ya que son el único elemento de interconexión entre los numerosos clientes inalámbricos y los recursos del sistema. No obstante, al contrario de lo que podría suponerse, la saturación no se produce en el nivel WLAN, si no que se encuentra en su capa IP. En este nivel, los puntos de acceso deben procesar todo el flujo de datos existentes entre cada uno de los nodos móviles que tenga asociados y los diferentes recursos del modelo, tanto internos como externos. Esto significa que cualquier petición realizada por algún cliente móvil, debe ser procesada por esta capa en alguno de los APs del sistema. La gráfica 4(e) muestra el tráfico descartado en la capa IP por los 3 puntos de acceso saturados: hall, sala1 y sala2, ya que en la secretaría y en la sala de PCs, el número de nodos móviles activos es demasiado reducido. El tráfico descartado por cada AP dependerá del periodo simulado: alcanzarán la saturación durante los intervalos en que se incrementa el número de nodos que tienen asociados.

6. Conclusiones

A partir del análisis del modelo actual, se puede determinar que éste satisface las necesidades de calidad de servicio para los asistentes a este tipo de eventos. En cualquier caso, debemos tener presente que el patrón de movimiento de los usuarios es predecible atendiendo al horario establecido de actividades. Así, el sistema debe planificarse para que pueda responder de manera adecuada durante estos instantes de saturación.

El cuello de botella lo constituyen los puntos de acceso, ya que son el único elemento de conexión entre la red cableada y la red inalámbrica; por un lado, se produce una saturación en la capa IP, ya que no pueden procesar todo el tráfico recibido y, por otro, se experimentan elevados tiempos de contención en el acceso al medio wireless. De esta manera, una buena optimización del sistema consistiría en incrementar el número de puntos de acceso, con lo que conseguiríamos reducir el tráfico descartado y los tiempos de espera, lo que a su vez, repercutiría en una reducción de los tiempos de respuesta proporcionados por las diferentes aplicaciones.

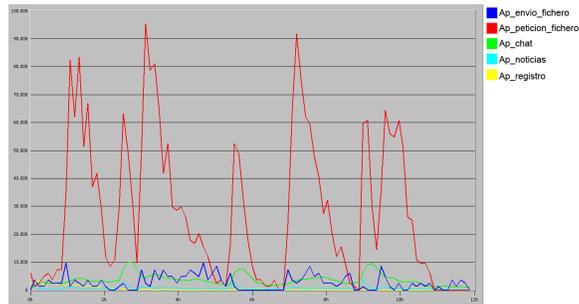
Por otro lado, TCP no fue diseñado para operar en entornos wireless por lo que se debería adoptar alguna de las variantes diseñadas para optimizar su funcionamiento en estos entornos inalámbricos, como las definidas en [7] [8] [9].

7. Agradecimientos

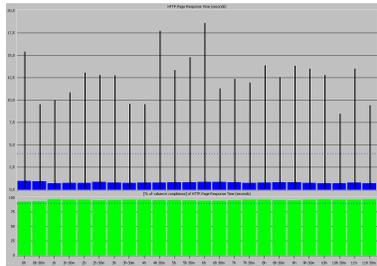
Este trabajo ha sido financiado parcialmente por el Ministerio de Ciencia e Innovación del Gobierno de España a través del proyecto TIN 2007-60440.

Referencias

1. OPNET Technologies <http://www.opnet.com/>
2. The average web page. <http://www.optimizationweek.com/reviews/average-web-page/>
3. Svoboda, P.; Kamer, W.; Rupp, M.: Modeling E-Mail Traffic for 3G Mobile Networks (2007)
4. The Average Web Page Size Triples Since 2003. <http://www.websiteoptimization.com/speed/tweak/average-web-page/>
5. Casalicchio, E.: Cluster-based Web Systems: Paradigms and Dispatching Algorithms (2002)
6. ITU-T Recommendation G.1010: End-User multimedia QoS categories (2001)
7. Elaarag, H.: Improving TCP Performance over Mobile Networks Stetson University
8. Sarolahti, P.: Performance Evaluation of TCP Enhancements over Slow Wireless Links Department of Computer Science, University of Helsinki
9. Balakrishnan, H.; Padmanabhan, V. N.; Seshan, S.; Katz, R. H.: A Comparison of Mechanisms for Improving TCP Performance over Wireless Links Computer Science Division, Department of EECS, University of California at Berkeley



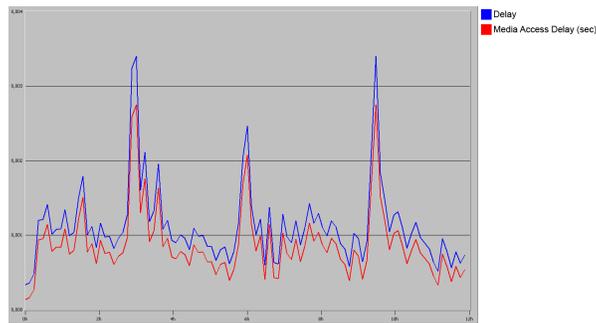
(a) Tráfico enviado (bytes/seg)



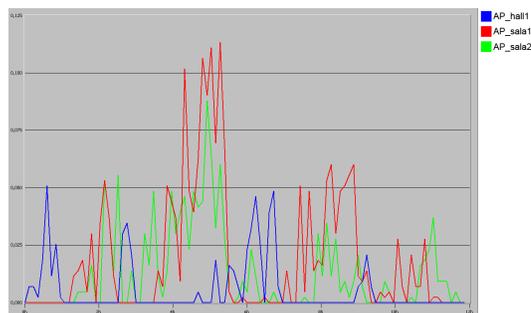
(b) HTTP - Tiempo de Respuesta (seg)



(c) Retardo TCP (seg)



(d) Retardo WLAN (seg)



(e) Tráfico Descartado IP (paquetes/seg)

Figura 4. Principales estadísticas obtenidas a partir de la simulación del modelo