

ESTUDIO SOBRE LA DINÁMICA DEL INTERNET EN MÉXICO

Edmar Mota García and Rogelio Hasimoto Beltran

Comunicación Técnica No I-06-04/27-01-2006
(CC/CIMAT)



Estudio sobre la dinámica del Internet en México

Edmar Mota García, Rogelio Hasimoto Beltrán

Centro de Investigación en Matemáticas A.C. Callejón Jalisco S/N Mineral de Valenciana. Guanajuato,
Gto. C.P. 36000

edmar@cimat.mx, hasimoto@cimat.mx

Enero, 2006



Resumen

Se describe el trabajo de más de 200 horas de mediciones de las principales variables que caracterizan el comportamiento del Internet: a) tiempo de ida y vuelta (RTT), b) tiempo de una sola vía (OTT) y c) pérdidas de paquetes. Las variables son analizadas para diferentes tiempos de envío y condiciones de tráfico mediante el uso de herramientas basadas en los protocolos ICMP y UDP. Las herramientas basadas en ICMP, no requieren de un programa huésped ejecutándose en el servidor destino, permitiendo la obtención de datos de RTT de una manera muy sencilla. Por otro lado, las herramientas basadas en UDP permiten mayor control en la toma de los datos permitiendo obtener muestras confiables de OTT y pérdidas en ambas vías (ida y vuelta) de manera independiente. Se ajusta un modelo de distribución representativo a cada una de las variables analizadas y se comparan con los modelos propuestos en la literatura. Se discuten además los problemas que se encuentran al realizar este tipo de mediciones, cómo se pueden atacar y algunos resultados interesantes del análisis del comportamiento del OTT, RTT y pérdidas de paquetes.

1 Introducción.

La transmisión de datos en el Internet de manera robusta requiere de contar con modelos adecuados del tiempo de una sola dirección o vía (OTT-One way Transmisión Time), de ida y vuelta (RTT-Round Trip Time), así como de una cuantificación inicial de las pérdidas de los paquetes y su distribución. La calidad de una aplicación multimedia (video o audio) por ejemplo, esta determinada primordialmente por estas variables. Si un paquete se pierde, la calidad de la información se degrada a menos que sea recobrada con mecanismos de corrección de errores (FEC- Forward Error Correction) o retransmisión (ARQ-Automatic Repeat Request). Si el retardo de una vía es demasiado grande y se pierde la sincronía en la reproducción, conducirá a pérdidas de información por retraso. Por otro lado, un RTT grande degradará la interactividad de la aplicación. Estudios en la calidad de voz sobre IP, nos expresan que un retardo mayor a 300 ms. evita una comunicación eficaz entre ambos puntos en la red. Retardos menores son necesarios para una transmisión efectiva de video, debido a que la sincronía y procesamiento es mucho más intenso dada la naturaleza propia de la transmisión de video.

Muchos estudios se han realizado sobre la dinámica de los paquetes en el Internet, entre los más importantes se encuentran los de Paxson en 1997 ([5] y [6]). Paxson estudió alrededor de 37 servidores y ruteadores en Europa y Estados Unidos y obtuvo miles de mediciones bajo el protocolo TCP, obteniendo varias conclusiones interesantes que desde la publicación de su estudio han servido de referencia obligada al hablar sobre la dinámica de los paquetes en Internet. Entre sus conclusiones se encuentran:

- Las características del Internet son altamente variables y no existe un comportamiento que se pudiera considerar “típico”.
- Las pérdidas observadas no son un buen estimador de las pérdidas en el futuro.
- Es poco probable que un paquete utilice el mismo camino en su trayecto de ida y vuelta lo que permite indicar que dividir el RTT en partes iguales es un mal estimador del tiempo de una vía.

En este trabajo se estudia el comportamiento del Internet en México y su relación con el Internet a escala mundial. Dicho de otra manera, ¿los modelos aplicados a escala mundial son válidos a escala local o regional?. Para responder esta pregunta, se consideran únicamente tres variables importantes sobre las cuáles recaen la mayoría de los modelos propuestos que explican el comportamiento del Internet; estas variables son RTT, OTT y pérdidas de información (paquetes). Las mediciones de RTT se hicieron utilizando el protocolo ICMP (Internet Control Message Protocol), las correspondientes al OTT (tiempo de ida o vuelta) utilizando UDP (User Datagram Protocol), y para las pérdidas de paquetes se usaron ambos protocolos ICMP y UDP. La ventaja de usar ICMP para la medición del RTT es que la mayoría de los equipos responden a este tipo de mensajes (eco o de sello de tiempo) automáticamente (a través del sistema operativo) sin requerir una aplicación (herramienta) huésped en el equipo destino. A diferencia de ICMP, las aplicaciones basadas en UDP requieren de su ejecución en ambos equipos participantes en la medición. Situación complicada ya que se necesita acceder al equipo destino. A pesar de esto, las aplicaciones basadas en UDP nos proveen de datos suficientes para analizar el comportamiento de los retardos de una vía en ambas direcciones, así como también el comportamiento de las pérdidas de paquetes de información.

Los datos obtenidos corresponden a una serie de mediciones realizadas a través de varios meses de trabajo en varios sitios de la república Mexicana (Figura 1) y dos sitios adicionales en Canadá y Taiwán. Estos sitios en el extranjero son usados como puntos de comparación entre la dinámica del Internet interna y externa. En términos generales, los resultados obtenidos pueden en gran medida ayudar al desarrollo de tecnologías y arquitecturas que descansen en la infraestructura existente actualmente en el país, principalmente la infraestructura que pudiera considerarse pública (universidades y gobierno).



Figura 1 – Gráfica de Localización. En la gráfica se muestran los servidores propuestos para las pruebas ICMP. Aquí A) ITESM Campus Monterrey, B) UNAM, C) Universidad Autónoma de Sinaloa, D) INAOEP, E) Gobierno del Estado de Oaxaca, F) Instituto Tecnológico de Mérida, G) Universidad Autónoma de San Luis Potosí. GTO corresponde al CIMAT.

La organización del presente trabajo es como sigue. En la sección 2 se describe la metodología y problemas existentes en la toma de datos (RTT, OTT y pérdidas de paquetes). En la sección 3 se describe la metodología para el análisis de los datos. Los resultados y conclusiones se presentan en las secciones 4 y 5 respectivamente.

2 Adquisición de Datos.

2.1 Mediciones del RTT

El cálculo del retardo de los paquetes al atravesar varios enlaces dentro de una conexión punto a punto normalmente requiere de dos tiempos, el tiempo de envío y el tiempo de recepción del paquete una vez que recorrió la trayectoria de ida y vuelta. Esto es:

$$RTT = T_{rcp} - T_{env} \quad (1)$$

donde T_{env} es el tiempo en el que se envió el paquete y T_{rcp} es el tiempo en que se recibió la respuesta. Esta relación incluye el retardo en la cola del destino al procesar el paquete, lo pudiera sobreestimar su valor si deseamos únicamente el tiempo que pasó nuestro paquete en la ruta. Si además se desea tener mediciones de tiempos de ida y de regreso se necesita contar entonces con cuatro tiempos, el tiempo de envío desde el emisor (T_1), el tiempo de recepción al llegar al destino (T_2), el tiempo de envío desde el destino (T_3) y el tiempo de llegada al emisor (T_4). Con esto ahora podemos calcular el tiempo de ida y vuelta con:

$$\begin{aligned} RTT &= (T_4 - T_1) - (T_3 - T_2) \\ T_{ida} &= (T_2 - T_1) \\ T_{vuelta} &= (T_4 - T_3) \end{aligned} \quad (2)$$

Si se analiza (2) se puede ver que a la ecuación (1) se le restó el tiempo que tardó el paquete en la cola del destino, obteniendo así una medición más exacta del RTT al pasar por los enrutadores desde el emisor al receptor¹. El RTT es la suma entonces de los retardos en cada enlace. Cada retardo en cada enlace consta de cuatro componentes: retardo de procesamiento, retardo de encolado, retardo de transmisión y retardo de propagación. Teniendo un tamaño de paquete fijo y una ruta fija, se puede asumir que el retardo de los paquetes enviados sólo cambiara con el retardo de encolado, que en el Internet cambia con las fluctuaciones del tráfico. Entonces, para una ruta fija en Internet, el RTT $T(t)$ es una variable aleatoria en tiempo t , entonces se dice que $T(t)$ nos describe el proceso del retardo de ida y vuelta del paquete. Para medir el RTT se utilizó el método especificado en (2).

2.2 Mediciones de una vía (OTT).

Como se observa en el conjunto de expresiones de (2), es posible obtener las mediciones de los tiempos de ida y vuelta. Si además tenemos herramientas ejecutándose en ambos puntos, podemos contar con mediciones independientes de los trayectos de ida y vuelta. El problema en este caso es que involucramos los relojes de los dos equipos finales que realizan las mediciones y esto lleva a la presencia de diferencias en tiempo y velocidad.

Al problema de diferencias en tiempo se le conoce como *desfase* (offset). Existe el denominado desfase global, con respecto a un tiempo considerado exacto (un reloj atómico por ejemplo) y un desfase relativo, este involucrando dos relojes no exactos, entre sí. Si un reloj reporta un tiempo T_c y el tiempo

¹ Esto no es realmente cierto si se utiliza la respuesta automática de un servidor a ICMP, ya que la implementación usada, tanto en Unix como en Windows, coloca el mismo tiempo en los campos correspondientes a los tiempos de llegada y salida.

real es T_r , entonces el desfase relativo será $T_c - T_r$. A lo largo de la sección el desfase se refiere únicamente al desfase relativo.

Como se expresa en [6], es difícil tener sincronizados los relojes tanto de la máquina de medición como la máquina remota. Esto ocasiona que exista de manera inherente un desfase entre ambos relojes que debe ser tomado en cuenta al momento de obtener los tiempos de ida y vuelta. Lo que lleva a tratar de calcular un estimador del desfase entre los relojes para obtener mediciones de ida y vuelta acordes con la resolución de las mediciones de RTT.

Se considera el desfase entre ambos relojes en las rutas de ida y vuelta como:

$$\begin{aligned} \text{Offset}_{A-B} &= \min_i (T_1(i) - T_2(i)) \\ \text{Offset}_{B-A} &= \min_i (T_3(i) - T_4(i)) \end{aligned} \quad (3)$$

donde los tiempos T_1, T_2, T_3 y T_4 son los mismos de las expresiones en (2). Entonces un estimador del desfase total entre ambos relojes y considerándolo constante es:

$$\text{Offset}_{|A-B|} = \frac{\text{Offset}_{A-B} + \text{Offset}_{B-A}}{2} \quad (4)$$

En el caso de las mediciones con ICMP este desfase se calcula y se ajusta para un número pequeño de paquetes de manera dinámica (tomando siempre el mínimo desfase observado en ese intervalo de paquetes al momento de recibirlos) y se guardan los datos ya ajustados para su análisis.

Para el caso de mediciones de UDP se tomaron los datos crudos y se aplicó el método descrito anteriormente pero al conjunto de datos completos una vez recibidas todas las mediciones. Esto implica considerar que tanto $\text{Offset}_{A-B} = \min_i (T_1(i) - T_2(i))$ como $\text{Offset}_{B-A} = \min_i (T_3(i) - T_4(i))$ será aplicado a todo el intervalo y no solo a en un intervalo pequeño. Esto está más apegado al método de Paxson explicado en [6].

Por otro lado se intentó minimizar el efecto del desfase en ambos lados de la línea mediante el protocolo NTP (Network Time Protocol). No obstante que esto minimizó el efecto del desfase, se presentó otro efecto muy notable de “salto” debido a la manera en que NTP actualiza el reloj (ver Figura 2). Esto se puede eliminar al realizar adecuaciones al núcleo del sistema operativo y al adecuar NTP para cada caso. Estas adecuaciones no son viables en situaciones donde no se tienen los permisos para modificar al sistema huésped y por otro lado son lentas de implementar. Una solución sencilla para obtener mediciones sin desfase y exactas es contar en cada equipo con un reloj sincronizado mediante GPS y hacer que NTP use este reloj como base. Esto implica la instalación de hardware especial que tampoco era viable. El contar con relojes sincronizados es entonces un problema complicado que va más allá este reporte.

Otro problema inherente a las mediciones de OTT es la diferencia de resolución o de frecuencia del reloj entre dos equipos diferentes. Este problema se denomina *sesgo* y se origina por las diferencias entre las actualizaciones a nivel del núcleo del sistema operativo del tiempo devuelto a la aplicación. Esto normalmente genera incrementos lineales en las diferencias entre los relojes. Un ejemplo del sesgo se puede observar en la Figura 3.

En el caso de ICMP al aplicar la metodología dinámica para eliminar el desfase, se pudo notar que el sesgo parece relativamente pequeño (o casi cero) hacia los lugares probados. No obstante es difícil saber si efectivamente existía sesgo (dada la manera de ajuste del método dinámico), en que grado y si realmente se eliminó por completo de los datos obtenidos. Debido a esto, se consideró tomarlos sólo como referencia y utilizar los datos de UDP para el análisis de OTT.

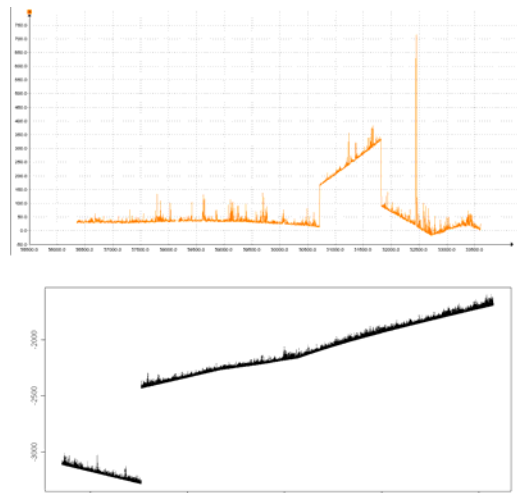


Figura 2. Ejemplos del efecto de salto en mediciones hechas con UDP debido al ajuste de NTP.

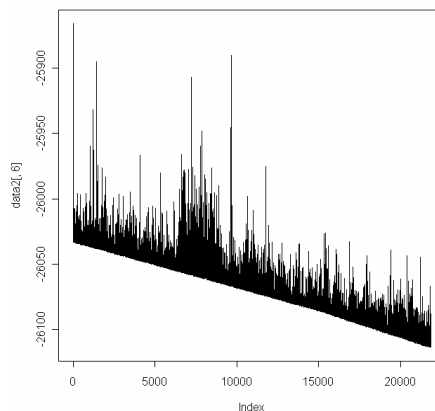


Figura 3. Ejemplo de la aparición del sesgo en mediciones de OTT con UDP.

Como se observa en la Figura 3, en el caso de las mediciones de UDP el efecto del sesgo es muy evidente incluso para intervalos de medición cortos (de 30 minutos) y se decidió eliminarlo mediante técnicas que se reportan en la literatura. En este caso se utilizó el algoritmo propuesto por Moon et al. [4]. Moon et al., utiliza una variante de un algoritmo de programación lineal de dos variables para encontrar un estimador del sesgo promedio observado, en el anexo se puede encontrar el algoritmo implementado en este trabajo. Al eliminar este sesgo (y agregando la metodología de Paxson para la substracción del desfase) podemos obtener buenas mediciones del OTT entre 2 equipos para efectos de su estudio.

2.3 Medición de las Pérdidas de Paquetes.

Para medir las pérdidas de paquetes a lo largo de la ruta de una conexión punto a punto se envían paquetes numerados de prueba y se anotan los números de secuencia que se reciben satisfactoriamente en el receptor. En el caso de una transmisión bajo el protocolo ICMP (utilizando un mensaje de eco o un mensaje de sello de tiempo) las pérdidas pueden ocurrir tanto en el trayecto de ida y como en el regreso, así que las mediciones obtenidas con este protocolo (a diferencia de aquellas realizadas con otros

protocolos que requieren de presencia en el destino como UDP o TCP) nos dirán la fiabilidad del ICMP, pero **sólo** cuantificarán las pérdidas del trayecto completo (ida y vuelta) atravesado por nuestro paquete. Es decir es imposible saber con exactitud si el paquete se perdió en el trayecto de ida o en el de regreso, por lo que no es posible estudiar si existe algún tipo de diferencia entre ellos.

En términos generales se espera que las pérdidas (en ICMP) sean menores que las reportadas en estudios de pérdidas utilizando TCP y UDP, esto es debido principalmente que el protocolo ICMP tiene una mayor prioridad en los encolados de los servidores (por su naturaleza de control). No obstante puede existir un rango de pérdidas mayores, siempre y cuando en una ruta específica cambie un enrutador y éste no acepte el reenvío de mensajes ICMP, teniendo por esta razón pérdidas.

Con respecto a las mediciones utilizando UDP también se obtuvieron una serie de datos de pérdidas ya diferenciadas entre trayectos de ida y vuelta y se obtuvo además una serie de porcentaje de pérdidas a través del tiempo. Estos resultados son más fiables que aquellos obtenidos con ICMP pero como se verá más adelante no se diferenciaron mucho de los resultados obtenidos con ICMP.

2.4 Descripción de los conjuntos de datos obtenidos.

2.4.1 Equipo utilizado.

Para realizar las mediciones se utilizó un equipo Pentium III a 900Mhz con 196 MB en RAM y disco de 40GB con sistema operativo Red Hat Linux 9. Para los equipos remotos de las pruebas UDP, el equipo de Culiacán es un Celeron M a 2.0 Ghz con 256MB en RAM y disco de 80GB con sistema operativo Fedora Core 4, y el equipo en INAOEP es un Pentium 4 a 2.0Ghz con 512MB en RAM y disco de 40GB con sistema operativo Red Hat Linux 9. No fue posible conocer las características de los equipos a los que se les realizaron pruebas con ICMP.

2.4.2 Mediciones con ICMP.

Se realizaron pruebas con la herramienta de ICMP con sellos de tiempo a distintos servidores en la República Mexicana. Se obtuvieron las rutas del equipo en CIMAT a cada uno de los servidores, mediante la utilería *traceroute* de linux. En la Tabla 1 se resumen estos datos. En algunos casos, como por ejemplo el servidor en Mérida, este no devolvía valores utilizables en el espacio dentro del paquete ICMP donde debería estar el sello de tiempo y por tanto se reportan únicamente sus mediciones de RTT ya que estas no dependen de lo que hace el servidor probado.

Para obtener los datos para el análisis reportado en la sección 3 en adelante, se realizaron más de 200 horas totales de pruebas a los servidores mostrados en la Figura 1, obteniendo mediciones del orden de 2 a 24 horas para cada servidor durante los meses de Marzo, Abril y Mayo de 2004. En las mismas mediciones se tomaron las pérdidas y el RTT, así como el tiempo de ida y vuelta en los casos que esto fue posible.

Se consideraron dos escenarios de comparación para ver el desempeño tanto de la red como de las colas de entrada de los servidores al momento de procesar mensajes ICMP. El primer escenario es el caso de envío en ráfaga con poco tiempo de recuperación del buffer de los servidores intermedios. Intuitivamente esto podría ocasionar desbordamientos en las colas de entrada ocasionando retrasos o pérdidas en los paquetes y una alta variabilidad en las mediciones de los retardos. Para este escenario se enviaron, desde el equipo en CIMAT, paquetes ICMP con sello de tiempo y carga de 540 bytes cada 80 y 160 ms, obteniendo una tasa de envío aproximada de 54 Kbps, y 27 Kbps, durante 24 horas continuas. Los datos son promediados para cada conjunto de 10 paquetes (lleguen o no) de acuerdo a su número de secuencia. Para cada conjunto se calculan sus pérdidas y su duración y se almacenan en un archivo de salida aparte. En la Tabla 2 se resumen el número de datos obtenidos para cada prueba.



Dirección del Host	Número de Pasos	Ubicación
www.itesm.mx (131.178.53.76)	9	Monterrey, N.L.
dragon.dgsca.unam.mx (132.248.10.7)	8	Ciudad de México, DF.
www.uasnet.mx (148.227.1.9)	7	Culiacán, Sin.
Inaoep.mx (192.100.172.1)	7	Puebla, Pue.
www.oaxaca.gob.mx (200.36.39.129)	12	Oaxaca, Oax.
www.uaslp.mx (148.224.17.118)	16	San Luis Potosí, SLP.
www.itmerida.mx (200.34.128.11)	11	Mérida, Yuc.
Nodo Uninet en Monterrey (148.223.15.37)	7	Monterrey N.L.
National Defense University (210.71.44.169)	14+	Longtan Shiang, Taiwan
Columbia Bible Collage (216.187.71.252)	14	Abbotsford, Canada

Tabla 1-Descripción de los Servidores. ICMP. Los números de pasos son el número de enrutadores atravesados por el paquete para cada Host. La ubicación es la ciudad donde se encuentra el servidor. En el caso de Taiwan, *traceroute* se detuvo antes de llegar al destino.

Tiempo	Tasa	Total de Paquetes	Total de Medidas
80 ms	54 Kbps	1080000	108000
160 ms	27 Kbps	540000	54000
1 seg.	4.32 Kbps	86400	8640

Tabla 2- Resumen de los tiempos

El segundo escenario es el envío de paquetes con tiempo amplios de recuperación. En este caso se esperaba no obtener pérdidas ni retrasos considerables (en el caso que no hubiera dependencia a largo plazo del retardo), ni variabilidad grande en los tiempos (RTT). Para este caso se enviaron paquetes ICMP con las mismas condiciones anteriores cada segundo. De igual manera cada conjunto de 10 paquetes es promediado y su resultado es guardado en un archivo.

2.4.3 Mediciones con UDP.

En el caso de UDP se probaron únicamente 2 servidores, uno en Culiacán, Sinaloa y otro en Tonanzintla, Puebla (cerca de la ciudad de Puebla). Estos fueron los únicos equipos donde se pudieron instalar las rutinas necesarias para realizar mediciones de este tipo. Los servidores se describen de igual manera que para la sección anterior en la Tabla 3.



Dirección del Host	Número de Pasos	Ubicación
www.tecedusin.gob.mx (200.52.176.30)	17	Culiacán, Sin.
Inaoep.mx (192.100.172.24)	6	Puebla, Pue.

Tabla 3. Descripción de los Servidores. UDP. Los números de pasos son el número de enrutadores atravesados por el paquete para cada Host.

El servidor de Culiacán se encuentra en el Departamento de Tecnología Educativa y cuenta con una conexión por cable. El servidor de Puebla se encuentra en las instalaciones del INAOEP y se encuentra conectado al CIMAT a través de Internet2. En particular, para el equipo en Puebla se logró sincronizar su reloj de manera exitosa con el reloj del equipo de CIMAT.

Para las mediciones de OTT y pérdidas utilizando el protocolo UDP se realizaron algo más de 50 horas totales de mediciones hacia los dos sitios desde enero a diciembre del 2005, obteniéndose mediciones del orden de 30 minutos a 6 horas. Adicionalmente se obtuvieron mediciones del comportamiento de las pérdidas con respecto al tiempo para intervalos de 24 horas.

Para el primer tipo de mediciones se realizó una utilidad que requería presencia en ambos lados de la ruta. La primera versión de la utilidad es idéntica a aquella utilizada con ICMP salvo que se realizó el cambio al protocolo UDP. Una segunda versión obtiene datos completamente diferenciados de la ruta en cada lado de la línea y es más sencillo analizarla. En ambos casos se realizaron mediciones con carga de 540 bytes y con carga de 1470 bytes a diferentes tasas de transferencia constante: 10, 20, 40, 50, 80, 160, 500 y 1000 ms.

En el caso de las mediciones de pérdidas para el comportamiento de estos a través del tiempo la rutina envía un número determinado de paquetes, en el caso que se reporta son 4500, y se detecta el número de ellos que llega de manera correcta.

3 Metodología de análisis de los Datos.

3.1 Introducción.

En esta sección se definen las técnicas utilizadas para el análisis de los datos recabados de RTT, OTT y pérdidas. En primer lugar se describe el análisis de presencia de dependencia a largo plazo (DALP), el cual es importante para poder suponer ciertas propiedades de la serie y poder realizar estudios sobre su distribución. Se presenta la función de correlación condicional, importante para estudios del comportamiento de los datos, así como también el cálculo de las funciones de autocorrelación, autocovarianza, y pruebas de hipótesis. Por último presentamos un resumen de la metodología utilizada para representar las series de pérdidas.

Antes de continuar es importante contar con la definición de estacionalidad de una serie de tiempo.

Definición 1: Una serie de tiempo se dice que es estacionaria en el sentido estricto, si las propiedades estadísticas se mantienen constantes a lo largo de toda la serie.

Definición 2: Una serie de tiempo se dice que es estacionaria en el sentido amplio (también llamada estacionalidad débil) si las funciones de media y covarianza se mantienen constantes a lo largo del tiempo.

Una vez expuestas estas definiciones podemos pasar a describir la metodología para el análisis de los datos.

3.2 Metodología para el proceso del Retardo.

3.2.1 El análisis de dependencia a largo plazo.

Para analizar la dependencia a largo plazo (DALP) dentro de una serie de tiempo se utilizó el método de varianza agregada o gráfica varianza-tiempo. La gráfica varianza-tiempo es obtenida al graficar $\log\{\text{var}[T^{(m)}]\}$ contra $\log(m)$, donde para cada $m = 1, 2, \dots$, que divide la serie de tiempo original $T = \{T_i, i \geq 1\}$ en bloques de tamaño m , y promediando para cada bloque, obtenemos el proceso agregado $T^{(m)} = \{T^{(m)}(k)\}, k = 1, 2, \dots$, y k es el índice de los bloques. Entonces si T no tiene dependencia a largo plazo se tendrá:

$$\text{Var}[T^{(m)}] \underset{m \rightarrow \infty}{\overset{\text{def}}{\approx}} m^{-1} \quad (5)$$

la pendiente obtenida de la gráfica varianza-tiempo debe ser igual a -1, por el contrario con dependencia a largo plazo, T puede ser caracterizada por

$$\text{Var}[T^{(m)}] \underset{m \rightarrow \infty}{\overset{\text{def}}{\approx}} m^{-\beta}, 0 < \beta < 1 \quad (6)$$

entonces la pendiente de la gráfica deberá alejarse de -1. Normalmente se utiliza el parámetro de Hurst H para medir la intensidad de la dependencia a largo plazo, y está relacionada al parámetro β por

$$H = 1 - \frac{\beta}{2}, \quad \frac{1}{2} < H < 1 \quad (7)$$

Para procesos con dependencia a corto plazo $H = 1/2$.

3.2.2 Análisis de la función de correlación condicional.

Una métrica más confiable para obtener datos sobre la dependencia en ráfaga del retardo (es decir la probabilidad de obtener bajo un intervalo l un retardo promedio mayor) es utilizar la métrica de complementariedad condicional de la función de correlación del retardo (CDF condicional) introducida en el trabajo de Bolot [1]. La CDF condicional se define como:

$$f(t) = P[d_i \geq t | d_{i-l} \geq t], l = 1, 2, 3, \dots \quad (8)$$

con l el intervalo, t es una medida del retardo y d_i el retardo observado por el paquete i , $i \in [1, n]$, n el número de paquetes totales. La ecuación anterior significa que si el paquete $i-1$ tiene un retardo $\geq t$, entonces con probabilidad $f(t)$ el paquete i también tendrá un retardo $\geq t$. Bolot [1] analiza la propiedad condicional de los retardos de ida y vuelta de paquetes consecutivos. Su conclusión es que tales retardos tienen una variación aleatoria en condiciones de carga ligera y cuando la carga del tráfico de fondo es alta, retardos consecutivos presentan picos. Un pico de retardo es una secuencia de retardos que comienzan con un retardo alto y decrecen casi linealmente después.

3.2.3 La función de autocorrelación (ACF).

Si se tiene $\{Z_i\}_{i=1}^{\infty}$ la secuencia estacionaria de variables aleatorias y además $\{z_i\}_{i=1}^n$ es obtenida a partir de $\{Z_i\}$, se define la función de autocorrelación para $\{z_i\}$ como:

$$\rho_z(h) = \frac{E[(Z_{i+h} - \mu_z)(Z_i - \mu_z)]}{E[(Z_i - \mu_z)^2]} \quad (9)$$

donde μ_z es la media y h es intervalo.

La función de la autocovarianza de la muestra, asumiendo estacionalidad es:

$$\hat{\gamma}(h) = n^{-1} \sum_{i=1}^{n-|h|} (z_{i+|h|} - \bar{z})(z_i - \bar{z}) \quad -n < h < n \quad (10)$$

con h el intervalo y \bar{z} la media de la muestra. Entonces la función de autocorrelación de una muestra será:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad -n < h < n \quad (11)$$

El intervalo también puede expresarse en términos de tiempo. Entonces, si se tiene a dt el intervalo en términos de tiempo entonces este será $dt = h\tau$, con τ el intervalo de muestreo.

Para un proceso estocástico independiente, la función de autocorrelación es cero, y para un proceso estocástico que es independiente en el intervalo h , la función de autocorrelación, $\rho_z(h)$, será cero en ese intervalo. En ciertos casos, principalmente al manejar muestras de una secuencia observada de un proceso estocástico independiente, se requiere contar con un valor de significancia.

Se tiene que, para una n grande, los valores de autocorrelación de una secuencia independiente e idénticamente distribuida (IID) con varianza finita son aproximadamente IID con una distribución normal $N(0, 1/n)$. Si $\{z_i\}_{i=1}^n$ es obtenida de esta distribución, el 95% de la autocorrelación de la muestra debe caer en el intervalo de confianza $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$. La escala de tiempo de correlación es entonces el menor intervalo en términos de tiempo d , más allá del cual la autocorrelación de la muestra es tan pequeña que puede considerarse insignificante.

Ahora, el estadístico de prueba se calcula a partir de la función de autocorrelación de la secuencia de muestras $\{z_i\}_{i=1}^n$ en el intervalo h como sigue:

$$S = \frac{\hat{\rho}(h)}{\sqrt{n}} \quad (12)$$

Si $\{Z_i\}_{i=1}^{\infty}$ es independiente hasta cierto grado en el intervalo h entonces $|S| \leq 1.96$, por el contrario se rechaza la independencia. Este es el método utilizado para encontrar los tiempos de mínimos para considerar a la serie estacionaria.

3.2.4 Ajuste de una distribución.

Después de analizar el comportamiento del retardo puede ser importante para poder describir un modelo de su comportamiento el ajustarle su distribución observada a alguna distribución conocida.

Para este trabajo se utilizó el proceso de máxima verosimilitud para el ajuste de los datos a una distribución conocida mediante el método de optimización Nelder-Mead, a través del ambiente R [8].

3.3 Metodología de análisis de pérdidas.

3.3.1 Representación de las pérdidas.

Existen dos maneras de representar los datos de pérdidas. La primera es un serie de tiempo binaria discreta $\{x_i\}_{i=1}^n$ tomada en los valores del conjunto $X = \{0, 1\}$, con 0 definiendo un evento exitoso (sin pérdidas). La traza también puede ser dividida en porciones de ceros consecutivos (llamadas corridas exitosas) y porciones de unos (llamadas corridas de pérdidas). Entonces una segunda manera de representar los datos es como la combinación de las dos secuencias de observaciones $\{g_i\}_{i=1}^e$ y $\{l_i\}_{i=1}^e$, con g_i es la i -ésima corrida exitosa (expresada en número de paquetes) y l_i la correspondiente longitud de la corrida de pérdidas. En el caso reportado aquí se utilizó la primera forma.

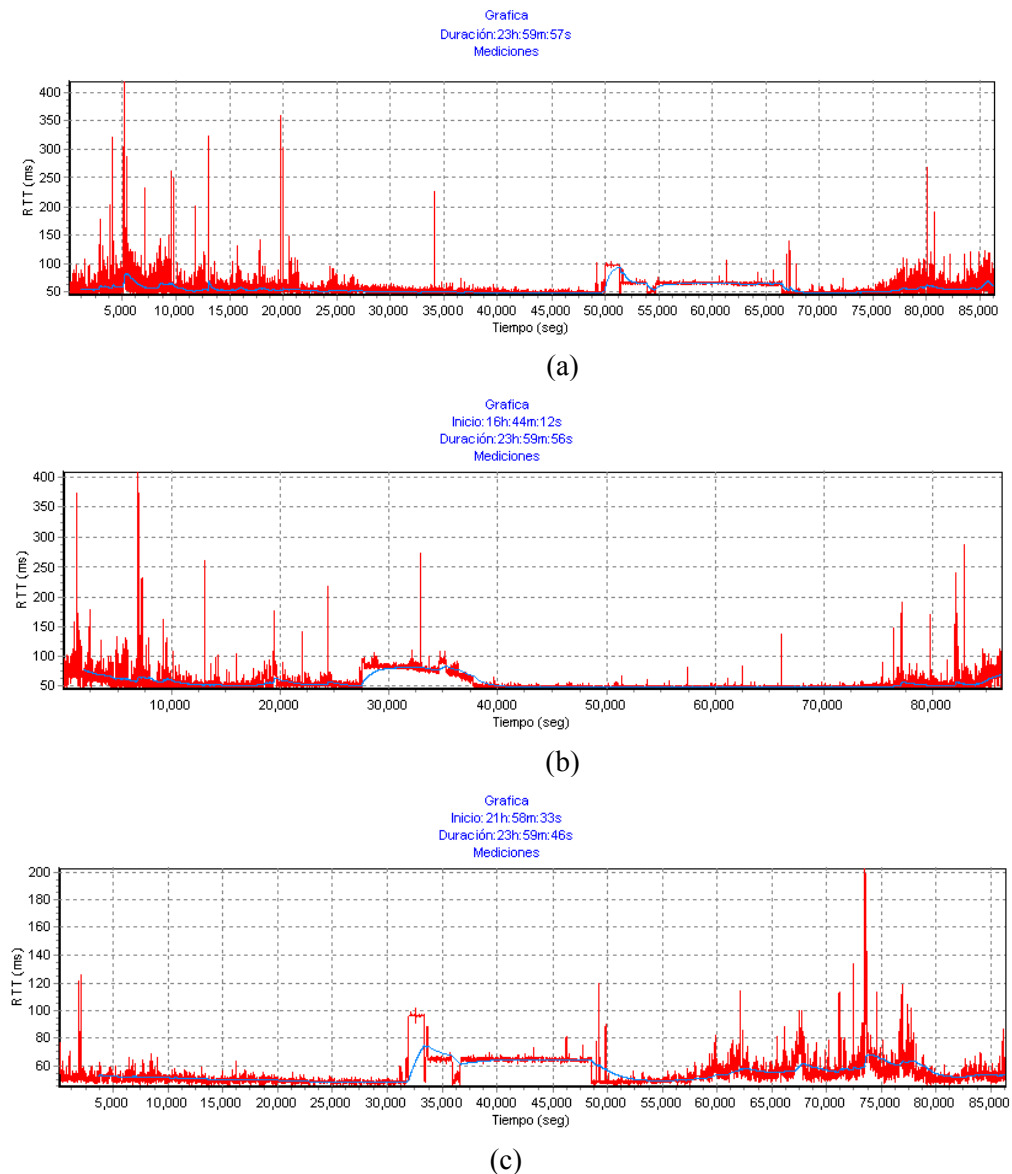


Figura 4. Series de tiempo para el sitio www.itesm.mx a) 80ms b) 160ms c) 1000ms. La línea sobre la serie es un moving average de 1hr.

4 Análisis de los Datos.

4.1 Análisis del RTT

4.1.1 Análisis simple de estacionalidad.

Para estudiar la posible estacionalidad de la serie de tiempo para el RTT se utilizó un filtro moving average para intervalos de 1 y 2 horas. En la Figura 4 se muestran las series de tiempo obtenidas para el sitio www.itesm.mx para cada uno de los valores de retardo, además de su moving average de intervalo

aproximado de 1hr. Datos similares fueron obtenidos para los otros sitios (ver sección 7). Como se puede observar en Figura 4, es sumamente difícil precisar estacionalidad en la serie, aún con el moving average. No obstante si nos enfocamos en el análisis de la serie tiempo para períodos de dos horas y de una hora y analizamos el comportamiento de la ACF, obtendremos los datos que se resumen en la Tabla 4.

Estos datos nos dan una pista de un valor promedio para considerar cualquier medición como estable. Estos valores no cuentan los segmentos en los que no se pudo precisar con seguridad si la serie es estacionaria (utilizando la ACF y considerando el intervalo completo 60 y 120 min). Los instantes en los que se presentaron porciones que pudieran considerarse estacionarias a un nivel menor a 10 segundos no necesariamente correspondieron a períodos diurnos o nocturnos. Por ejemplo hubo períodos de completa inestabilidad que darían un valor mínimo de 120 o 60 minutos a las 2:00 hrs. o a las 19:00 hrs. y valores mínimos de 1 segundo o menos a las 10:00 hrs. Puede pensarse que el comportamiento de la carga de trabajo de los enrutadores intermedios y del servidor no tuvieron un comportamiento predecible, ya que hubo servidores que aparentemente que tuvieron demasiada carga en períodos nocturnos y estuvieron relativamente estables en períodos diurnos o viceversa. No obstante en la mayoría de los casos los períodos de mayor trabajo (inestabilidad en el RTT) se presentaron en horario diurno como es de esperarse. Se realizó también análisis para intervalos de 15 minutos para la serie de 80 ms como comparación. En este caso se puede ver que el intervalo para considerar la serie estacionaria (es decir que podemos describirla a través de su estadísticas generales) disminuye a menos de 2 minutos, lo que es congruente con lo que se expresa en la literatura, de que si se toman mediciones a intervalos de 15 minutos sólo ocupamos valores en una amplitud de menos de un minuto para obtener estimación en ese intervalo.

Serie	Media y Mediana a partir de la cual se puede considerar a la serie estacionaria (1hr.)	Media y Mediana a partir de la cual se puede considerar a la serie estacionaria (2hr.)
80 ms	(18.858,8.425) min.	(31.047,11.2) min.
160 ms	(13.916, 2.25) min.	(22.206, 15.037) min.
1000 ms	(4.797,3.667) min.	(12.02,9.625) min.
80 ms (15 min.)	(1.939,0.217) min.	

Tabla 4- Valores centrales en tiempo para considerar a la serie estacionaria

4.1.2 Análisis de la dependencia temporal.

La dependencia a largo plazo (DALP) se caracteriza como se anotó en la sección 3.2.1 a través del parámetro de Hurst. Podemos considerar que un parámetro de Hurst menor 0.65 presenta efectivamente una dependencia a corto plazo, valores mayores nos dirán que la serie presenta una dependencia a largo plazo. En este caso de obtuvieron los valores para el parámetro de Hurst para los intervalos de 2 horas y de una hora y se promediaron los valores obtenidos, además se obtuvo un porcentaje de intervalos en los que el parámetro pudiera considerarse con bajo grado de dependencia a largo plazo, esto se resume en la Tabla 5. Esto es congruente con los hallazgos en [3], donde se observa que el parámetro de Hurst disminuye al disminuir el intervalo de muestreo (de muchas horas a pocas horas). Además los valores de Hurst bajos coincidieron con períodos estables en la serie (que pudieran considerarse estacionarios).

Serie	Media del parámetro de Hurst. (1hrs.)	Media del parámetro de Hurst. (2hrs.)	Porcentaje de intervalos considerados sin DALP (1hr,2hr)
80 ms (0.8796)	0.7569	0.7621	42,25
160 ms (0.8177)	0.7004	0.7591	67,42
1000 ms (0.901)	0.6530	0.76113	92,42

Tabla 5- Valores promedio del parámetro de Hurst de DALP. El número entre paréntesis es el valor de Hurst para toda la serie.

4.1.3 Variaciones en el RTT.

Para analizar las variaciones en el RTT se utilizaron varias técnicas descritas en [6], que utilizan la información intercuantil (se profundiza más adelante). En ellas se analiza la relación entre el máximo RTT y el mínimo RTT. En el este trabajo se utilizó este análisis para los tres tipos de pruebas (80, 160 y 1000 ms). Según Paxson [6], una manera de analizar la variación del RTT es a través del análisis de rango intercuantil (RIC). Para considerar datos normalizados se dividieron los rangos entre el mínimo para cada medición obtenida. Estos resultados se pueden ver en la Figura 5 (a), (b) y (c).

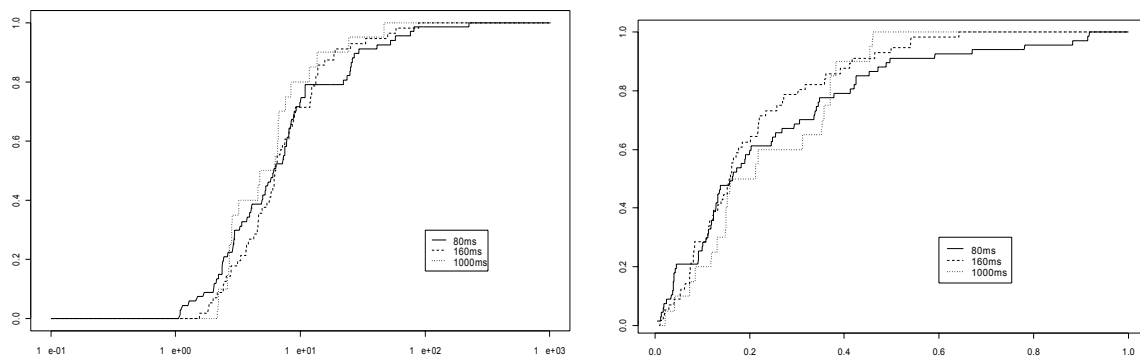
Paxson descubre que la relación mínimo RTT-máximo RTT en conexión TCP se comporta normalmente. Observando la Figura 5a (der) podemos ver que para ninguna de las velocidades de envío esto es válido. Por el contrario para la relación máximo-mínimo esto pareciera que si tiene una distribución gaussiana, específicamente para el retardo de 160ms. En el caso de la distribución del rango entre cuantiles normalizado entre el mínimo RTT podemos observar que tampoco se presenta una distribución normal para ningún caso, con lo cual el comportamiento del RTT para ICMP es diferente al observado para el RTT en el protocolo TCP.

La Figura 5c describe la distribución de las diferencias entre máximo y mínimo normalizadas por el rango intercuantil. Para el caso TCP esta distribución permite verificar que el utilizar media y varianza para obtener una estimación del máximo RTT no es una buena idea, a pesar de esto, el error se ve minimizado por el carácter dinámico de dichas estimaciones [6]. En el caso de ICMP se puede ver que específicamente puede parecer válido utilizar media y varianza para estimación de máximo RTT ya que el comportamiento de la distribución parece normal, principalmente para el caso de la medición a 80 milisegundos.

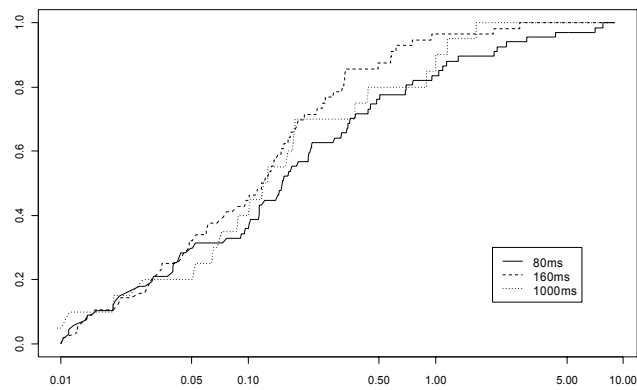
4.2 Análisis del tiempo de una vía (OTT).

4.2.1 Distribución del retardo

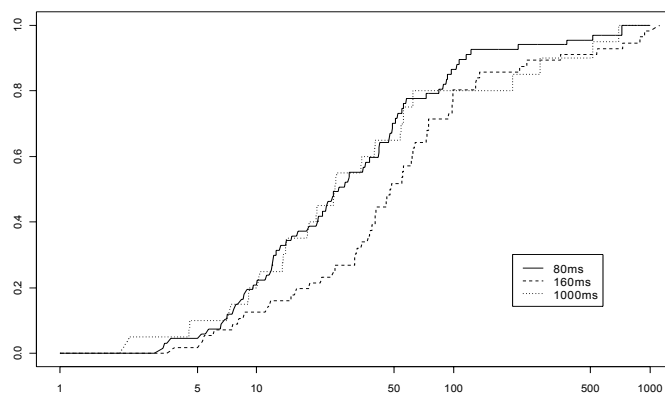
La distribución de probabilidad de los retardos de una sola vía es un aspecto importante para el desarrollo de técnicas que permitan reaccionar rápidamente a cambios en el comportamiento de la red; por ejemplo para evitar pérdidas de información. En la Figura 6a se muestra un ejemplo de la relación entre ida y regreso del sitio www.tecedusin.gob.mx para una muestra de 30 minutos con paquetes enviados cada 80ms. En la Figura 6b se muestra un ejemplo del comportamiento de las diferencias entre ida y vuelta contra el tiempo. En este caso se observa que casi siempre el trayecto de vuelta es más lento. En la Figura 6c, se presenta una comparación con el comportamiento de las diferencias normalizadas al máximo para mediciones hechas a www.itesm.mx con ICMP y al sitio www.tecedusin.gob.mx con UDP. Como se puede observar en el caso de ICMP el comportamiento muestra que en general para ICMP se puede asumir que el retardo de ida y de vuelta son iguales y se puede aproximar por una distribución normal, (más exactamente por una distribución Tukey-lambda). Por el contrario en el caso de UDP la distribución fue difícil encasillarla a una distribución en particular, pero consideramos que una distribución de colas pesadas es lo adecuado.



(a)



(b)



(c)

Figura 5. a) Distribución de la relación máximo-mínimo (izq) y mínimo- máximo (der). b) Distribución del RIC-normalizado. c) Distribución de las diferencias máx-mín normalizadas por RIC

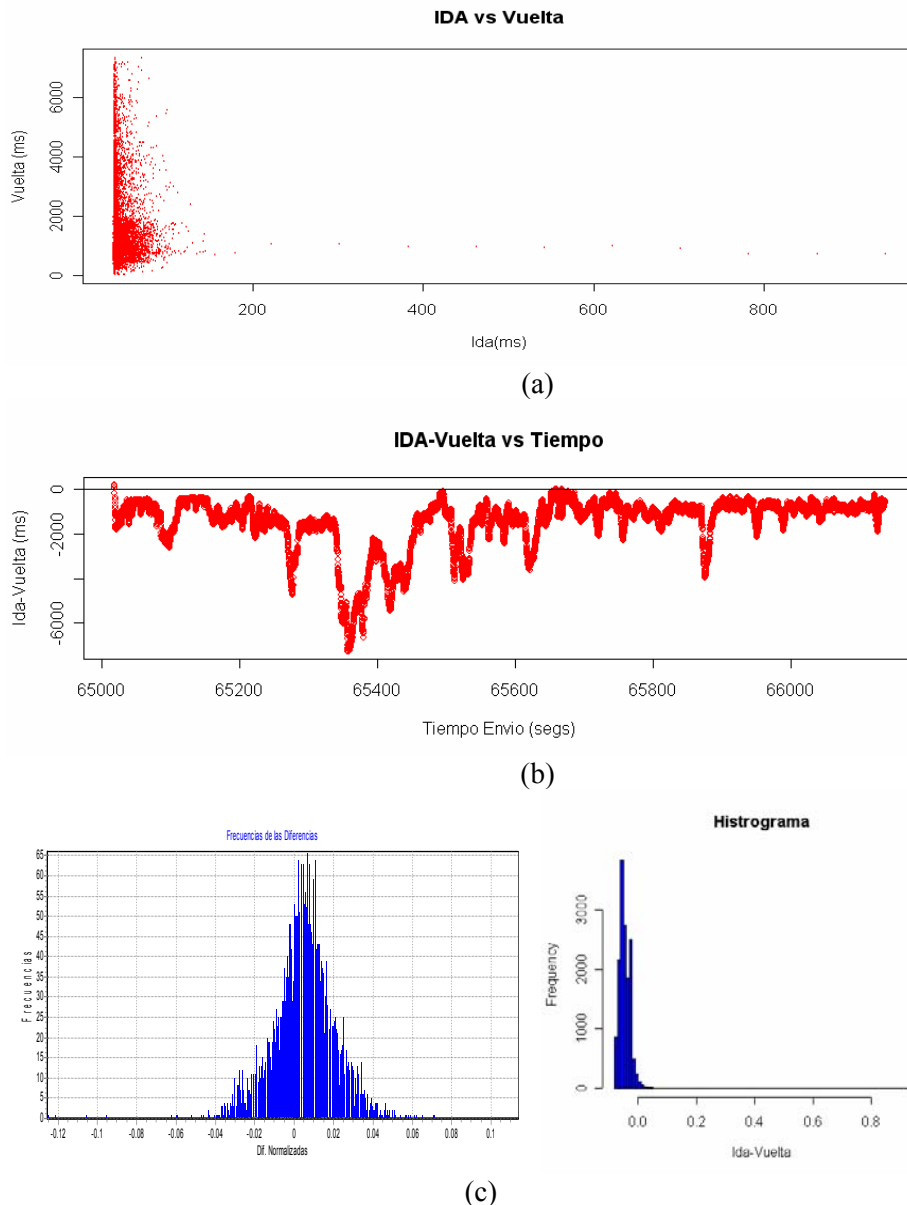


Figura 6. a) Tiempo de ida contra tiempo de vuelta de los paquetes en la ruta hacia www.tecedusin.gob.mx. b) Diferencias contra tiempo ida-vuelta hacia el sitio www.tecedusing.gob.mx. c) Distribución de frecuencias de las diferencias normalizadas, izquierda para ICMP y derecha para UDP.

En la Figura 7 se observa el ajuste que se hizo al retardo observado en la ruta Puebla-Guanajuato. Como se puede observar el retardo de una vía tiene un comportamiento con colas grandes comenzando en el retardo mínimo observado. Se le ajustaron varias distribuciones al histograma observado, utilizando principalmente distribuciones con colas pesadas, como Gamma y χ^2 . Se encontró que el mejor ajuste coincide con una distribución χ^2 corrida. Este ajuste no cambió mucho si se realiza el estudio a la serie cruda sin eliminarle el sesgo y el desfase, una comparación entre ambos se puede ver en la Figura 8.

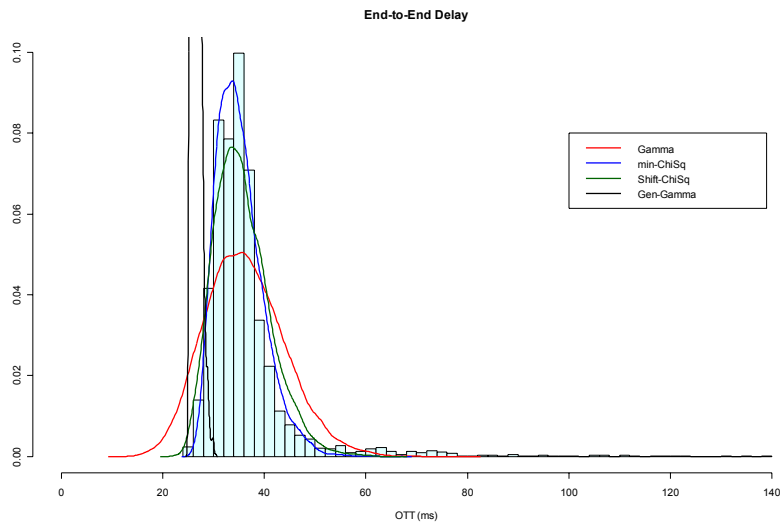
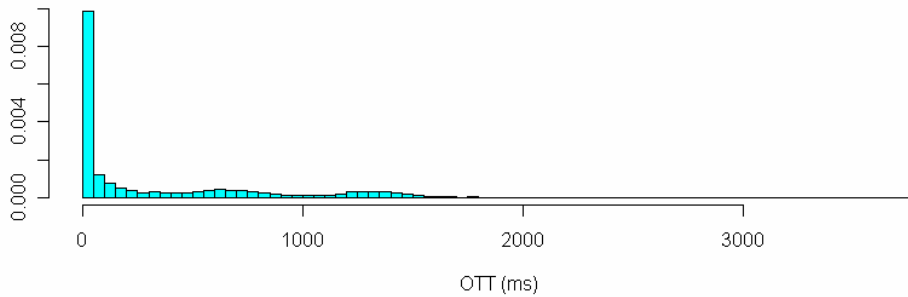


Figura 7. Ajuste de la distribución observada del retardo punto a punto de una vía en una ruta con equipos sincronizados (Puebla-Guanajuato). Las pruebas son a 80ms. Sitio www.inaoep.mx

Histograma OTT filtrado Cul-Gto 160ms



Histograma OTT-crudo Cul-Gto 160ms

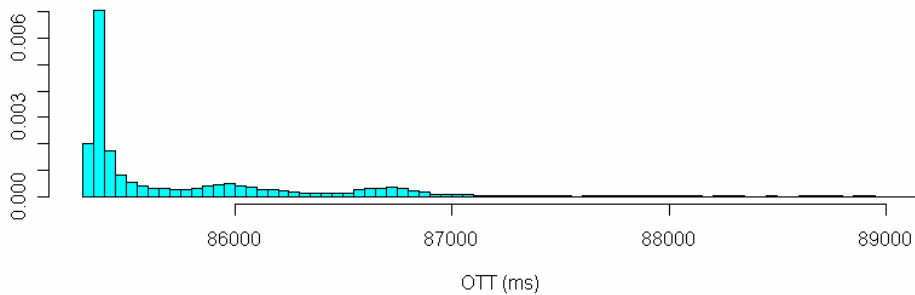


Figura 8. Histogramas de la ruta de regreso Cul-Gto filtrada para sesgo y desfase y cruda con tasa de prueba de 160ms. Sitio www.tecedusin.gob.mx

4.2.2 Causas del comportamiento observado.

En términos generales, se ha encontrado que el comportamiento entre los tiempos de ida y de vuelta es hasta cierto grado distinto (estadísticamente). En nuestras observaciones entre Guanajuato y Culiacán se encontró una diferencia significativa entre ambos retardos de 1:5, debida probablemente a que el servidor en el sitio www.tecedusin.gob.mx se encuentra conectado a la Internet mediante un módem de cable y este servicio tiene tasas y búferes diferenciados para los trayectos de ida y vuelta. Se comprobó que la capacidad del enlace de llegada al módem es de 1 Mb/s y la salida está valorada a 128 Kb/s. Además de esta posible causa, a través de la rutina *traceroute*, se pudo corroborar que los trayectos de entrada y de salida desde el sitio www.tecedusin.gob.mx son completamente diferentes. Esto puede agregar aún más retardo a las mediciones recibidas desde este sitio.

4.3 Análisis de las Pérdidas de Paquetes.

4.3.1 Análisis de frecuencias de las pérdidas.

Para el caso de las pérdidas de los paquetes utilizando ICMP se analizaron trazas de ocho sitios (ver Figura 9a). La distribución de las corridas de pérdidas tiene un comportamiento exponencial, lo que es congruente con la literatura.

Un punto a hacer notar es que los porcentajes de pérdidas de paquetes son menores que los reportados por los estudios con TCP y menores que los vistos en las pruebas UDP. Esto nos muestra que los paquetes ICMP (si no atraviesan un enrutador que los filtre) tienen una prioridad más alta que los paquetes comunes, prioridad no especificada en ningún momento por el protocolo. El máximo valor de porcentaje de pérdidas registrado fue de alrededor del 10% de pérdidas y el mínimo valor fue 0.04%, con un promedio de 3.7% de pérdidas, que es menor que los reportes anteriores para IP a escala global [5].

Para las frecuencias de pérdidas con ICMP se encontró que en las tres pruebas (80ms, 160ms y 1000ms) el porcentaje de sucesos con pérdidas consecutivas de 5 o más paquetes fue de menos de 0.5%, por lo que corridas más altas fueron muy raras. No obstante se presentaron series de hasta 40 paquetes perdidos consecutivos. En la Figura 9c se muestra las frecuencias para 80ms. En el caso de UDP también se observó este comportamiento donde menos del 2 % presentan pérdidas en bloque mayores a 6 paquetes (ver Figura 10).

Otro resultado interesante es que las pérdidas de paquetes se observaron justamente en las horas de más demanda, entre las 8:00 y las 21:00 hrs. Un ejemplo se puede observar en la Figura 9b.

4.3.2 Relación RTT y pérdidas (ICMP).

Por último se analizó el comportamiento de las pérdidas relacionado con el RTT. Se obtuvieron histogramas de la probabilidad de pérdida dado un valor de RTT (ver Figura 11). Para analizar adecuadamente esta relación, se compararon los paquetes perdidos con el correspondiente RTT máximo observado en cada grupo de 10 paquetes. Es conveniente recalcar que se utilizó el máximo y no el RTT promedio de ese intervalo de 10 paquetes dado que dada la teoría de colas es de esperar que las pérdidas hayan pasado entre dos mediciones de RTT altas (dentro de este intervalo). Se considero más justo para analizar y adecuado al contexto de utilización de estos resultados.

Otro análisis fue el de la distribución de los retardos máximos que sufrieron pérdidas. En la Figura 12 se muestra el comportamiento de estos retardos normalizados entre la mediana de los valores de RTT máximos observados. El eje horizontal de la gráfica está en escala logarítmica. Es interesante notar que un poco más del 70% de los pérdidas se obtuvieron en RTT máximos alejados 1.5 veces de su mediana. Esto es interesante ya que pudiera darnos pistas para la construcción de relaciones más adecuadas de inferencias de las pérdidas en relación al RTT (o sus valores estadísticos).

Se analizó también si pudiera existir una relación uno a uno entre el RTT y las pérdidas a través de una función de transferencia, cuyo resultado fue negativo.

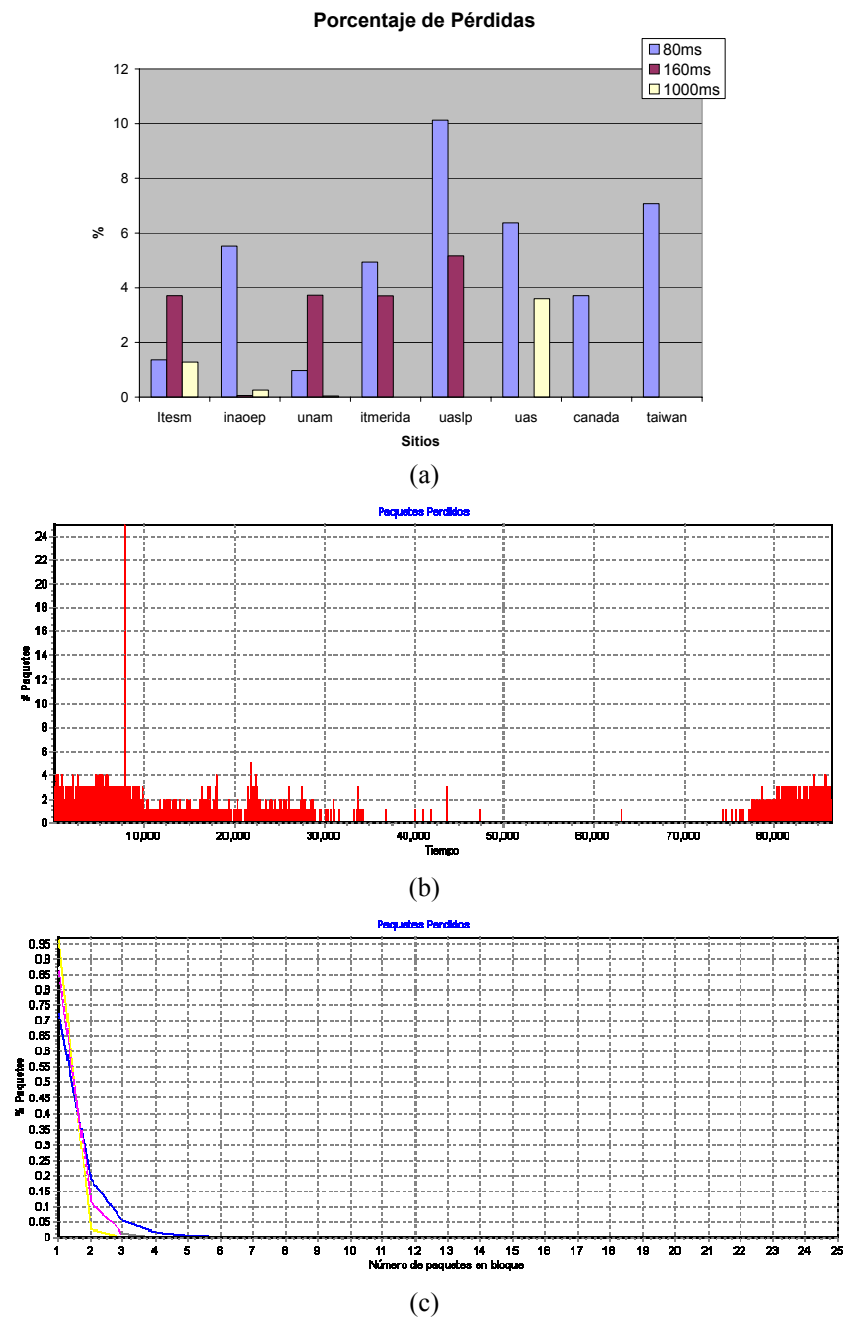


Figura 9. a) Porcentajes de pérdidas para distintos sitios. b) Serie de tiempo de las pérdidas, aquí el origen son las 12:30 horas tiempo del centro de México. b) Frecuencias de las corridas de pérdidas.

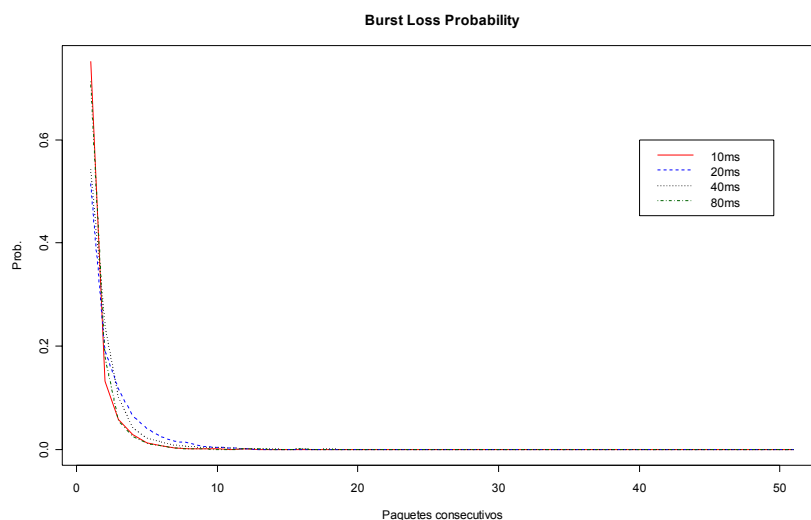
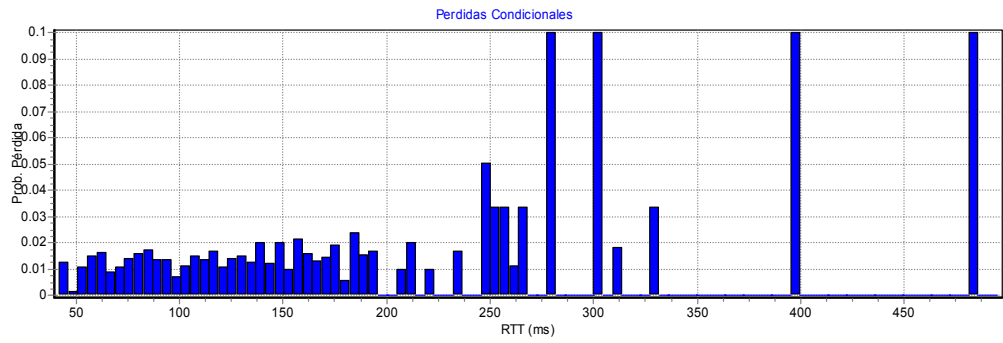


Figura 10. Probabilidad de pérdidas en bloque para UDP a distintas velocidades de transmisión. Sitio www.inaoep.mx

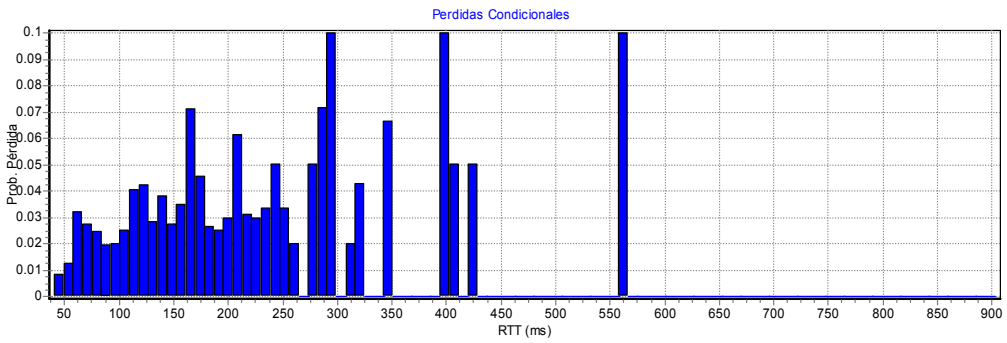
4.3.3 Análisis de las pérdidas en el tiempo (UDP).

En la Figura 13 se observa la razón de pérdidas (en porcentaje del total de paquetes enviados) para los trayectos de ida y vuelta hacia el sitio www.tecedusin.gob.mx durante 24 horas. Al analizar la distribución de las cantidades de paquetes perdidos se observaron comportamientos congruentes con la teoría, por ejemplo, que disminuía el número de paquetes perdidos con forme se disminuía la frecuencia de envío.

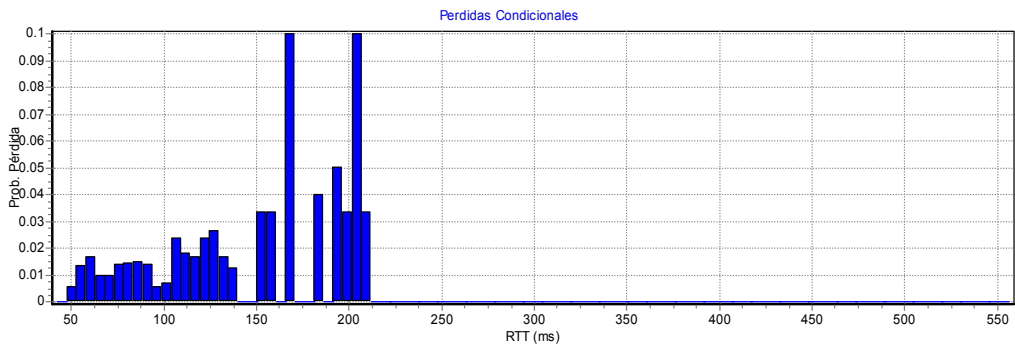
La relación entre porcentajes de pérdidas para cada trayecto es de casi 12 veces en promedio para esa tasa de transferencia, pero disminuye hasta casi 1.5 en tasas más altas. Esto es explicable por la diferencia entre los anchos de banda de entrada y salida propios de un módem por cable como se había hecho notar anteriormente.



(a)



(b)



(c)

Figura 11. Histograma de Frecuencias de probabilidad de pérdidas para cada nivel de RTT para el sitio www.itesm.mx a) 80ms , b) 160ms, c) 1000ms

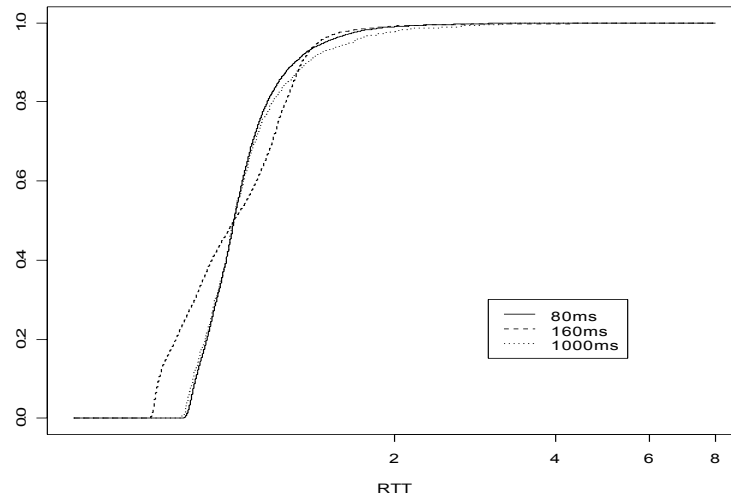
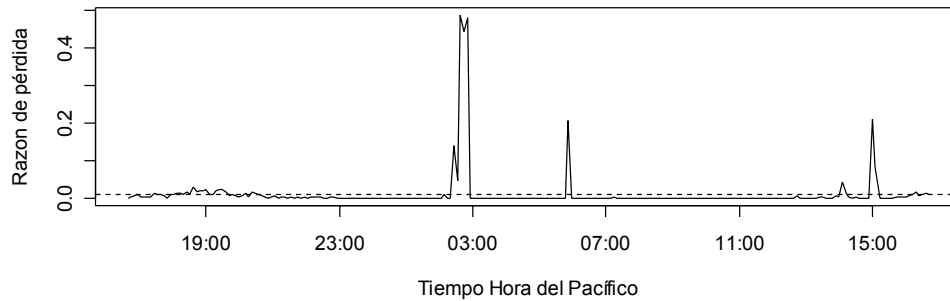
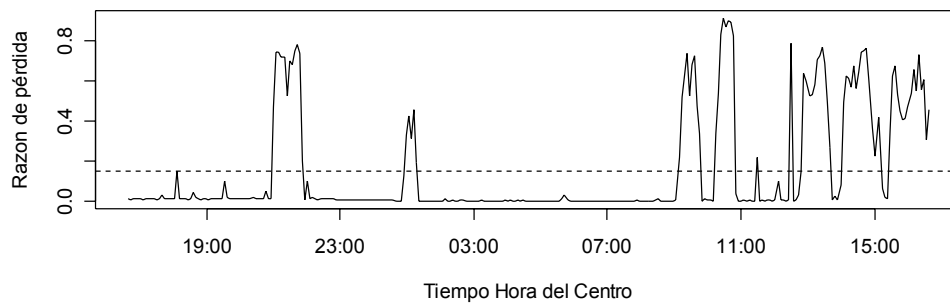


Figura 12. Distribución de probabilidades dado un RTT máximo. Sitio www.itesm.mx.

Perdidas Gto-Culiacan 40ms Prom= 1.20343251422764 % Med= 0.0888889 %



Perdidas Culiacan-Gto 40ms Prom= 15.0876630636364 % Med= 0.822222 %



(a)

Figura 13. Pérdidas vistas en la ruta Culiacán-Guanajuato. Sitio www.tecedusin.gob.mx

5 Conclusiones y Discusión.

El presente trabajo se presenta un estudio preliminar de la dinámica del Internet en México. Se analizaron las variables que en términos generales describen mejor su comportamiento como son el RTT, OTT y pérdidas de información. Entre los resultados más relevantes se encuentra que el comportamiento del RTT es congruente con lo expresado en los distintos trabajos revisados al respecto para Estados Unidos y Europa. Los datos preliminares expresan que en intervalos menores a 10 minutos podemos considerar como estable cualquier medición de RTT, menores de 30 minutos para el caso de OTT a la cual se le hicieron los estudios de ajuste de distribución. No obstante se requiere un estudio más detallado sobre los puntos donde efectivamente se da esta estabilidad y un análisis respecto a la hora del día en que se presenta. En este trabajo se noto que los períodos donde aumentaba la incertidumbre de estabilidad no necesariamente correspondían a períodos de carga alta (horas diurnas), si no que pareciera que en el caso de ICMP depende más de las tareas que el ruteador ejecuta fuera de su trabajo habitual de manejo de paquetes. Por ejemplo, en algunos casos se observaron períodos altos de trabajo en horas no diurnas lo que pudiera parecer que en ese período el ruteador o ruteadores hacia el sitio (e incluso el servidor mismo de respuesta) estaban ocupados realizando operaciones de alta carga de cpu.

El comportamiento observado del RTT parece darnos una pista de que los **supuestos** considerados en la literatura para TCP no siempre se cumplen para otros protocolos. Se encontraron varias diferencias entre los datos obtenidos en la literatura y los obtenidos con las mediciones realizadas (ICMP y UDP). Esto nos dice que cada protocolo de transporte tiene un comportamiento diferente y por tanto dicho comportamiento debe ser analizado antes de iniciar cualquier desarrollo de herramientas que utilicen de forma eficiente dicho protocolo. Además se encontraron pistas de que la dinámica del retardo (tanto de OTT como RTT) a través del tiempo pudieran ser explicadas a través de modelos Heteroscedásticos ya que las medias cambian con respecto al tiempo y no así la distribución de las varianzas, que siempre presentaron comportamientos similares a través del tiempo.

Algo interesante observado en el estudio de los datos de ida y vuelta con UDP, es que las diferencias entre estos dos tiempos tienen un comportamiento no normal, muy diferente a lo observado inicialmente en ICMP, comprobando así lo expresado en la mayoría de los estudios sobre las dinámicas del OTT. La presencia de colas pesadas, por otro lado, y la indicación de que en ningún caso puede existir un retardo menor a cierto valor mínimo dependiente del medio de transporte, permite afirmar que el retardo (tanto de una vía como de dos vías) puede ser modelado a través de una distribución Gamma o similares, específicamente la distribución χ^2 con corrimiento. Las distribuciones presentadas en la literatura (Weibull, Pareto y una mezcla de distribuciones normal y exponencial) no parecieron ajustar muy bien a los datos obtenidos del OTT en todos los casos observados.

Para las mediciones de pérdidas con ICMP, se encontró que el número de paquetes perdidos disminuye con la frecuencia de envío, además que las pérdidas se presentan en las horas de mayor trabajo. No obstante se obtuvieron muy pocas pérdidas en general, lo cual permite decirnos que el ICMP es un protocolo con prioridad alta. Otro resultado interesante es la distribución de las pérdidas de acuerdo a un RTT máximo observado en un intervalo de 10 paquetes (nuestros intervalos de prueba para los tres retardos). Hasta el momento, existe muy poca evidencia que pudiera indicarnos que el simple aumento o disminución gradual del retardo (tanto el OTT como el RTT) predice algún tipo de aumento en las pérdidas en el mismo sentido. Solo se pudiera afirmar que aumentos realmente considerables (del orden de 5 veces o más de la mediana del retardo, si esto llegara a presentarse) efectivamente conducen a eventos de pérdidas. Trabajos futuros pudieran realizarse al integrar las series de pérdidas con las series del retardo a través de métodos de análisis más sofisticados, un ejemplo interesante pudiera ser analizarlas por co-integración con innovaciones.



6 Referencias.

- [1]. Bolot, J.C. (1993) "Characterizing end-to-end packet delay and loss in the Internet". *Journal of High Speed Networks*. 2(3):305-323.
- [2]. Jiang, W. y Schulzrinne, H. (1999) "Modeling of Packet Loss and Delay and Their Effect on Real-Time Multimedia Service Quality"
- [3]. Li, Q y Mills, D.L. (1998) "On the Long-Range Dependence of Packet Round-Trip Delay in Internet". *Proc. IEEE International Conference on Communications*. Atlanta, GA. pp. 1185-1191.
- [4]. Moon, B. Sue, Skelly, P y Towsley, D. (1999) "Estimation and removal of clock skew from network delay measurements". *Memorias del IEEE INFOCOM '99*, New York, NY, March 1999
- [5]. Paxson, V. (1997) "End to End Internet Packet Dynamics". *ACM SIGCOMM Computer Communication Review*. Vol. 7. Núm. 4.
- [6]. Paxson, V. (1997) "Measurements and Analysis of End-to-End Internet Dynamics," Ph.D. dissertation, University of California, Berkeley.
- [7]. Postel, J. (1981). "Internet Control Message Protocol". RFC 792. Network Information Center, SRI International, Menlo Park, California.
- [8]. R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [9]. Yajnik, M., Moon, S., Kurose, J. y Towsley, D., (1999) "Measurement and modeling of the temporal dependence in packet loss". In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*. NY.

7 Anexos.

A continuación se presentan algunas gráficas de las series de tiempo obtenidas para distintos sitios.

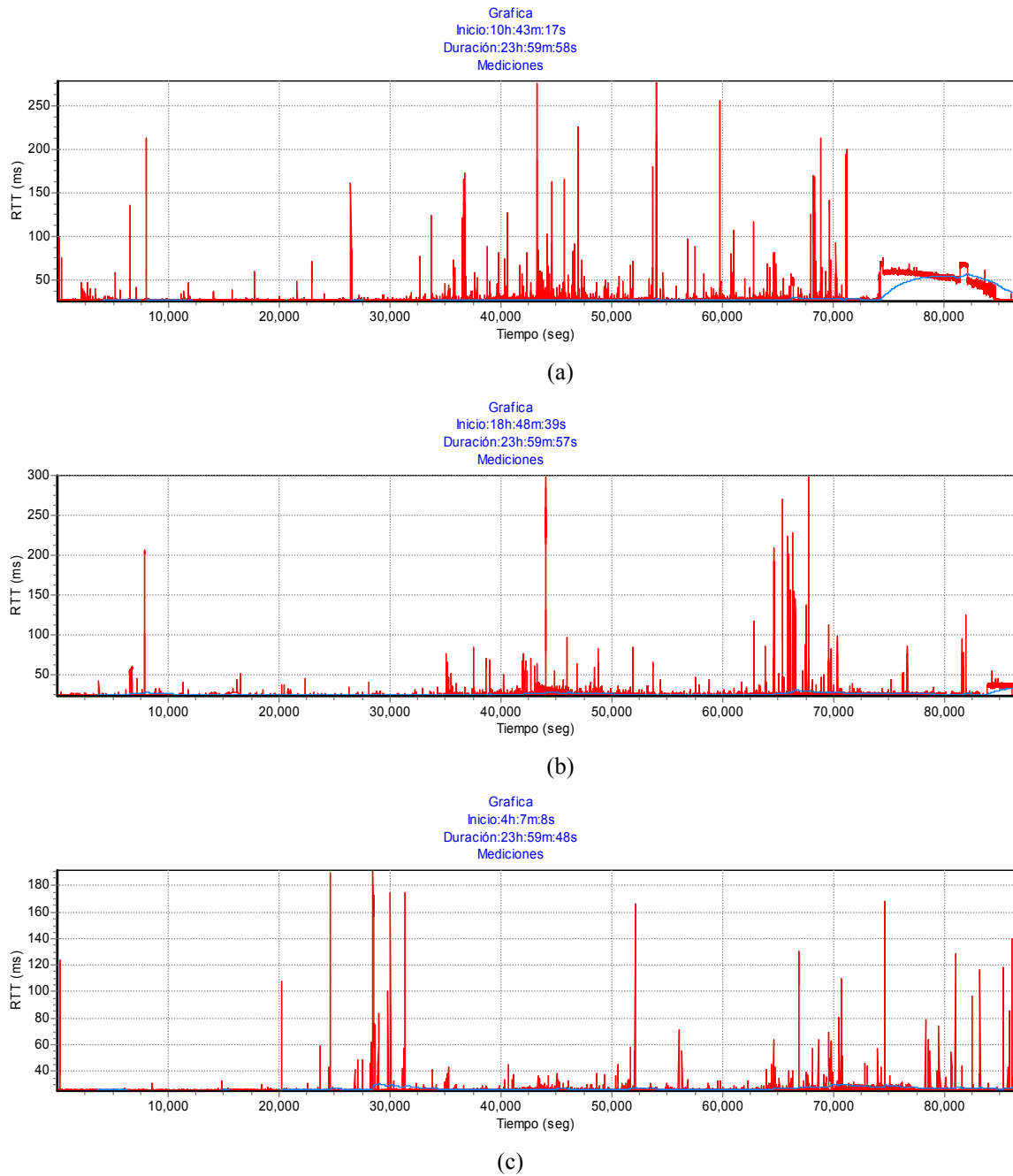
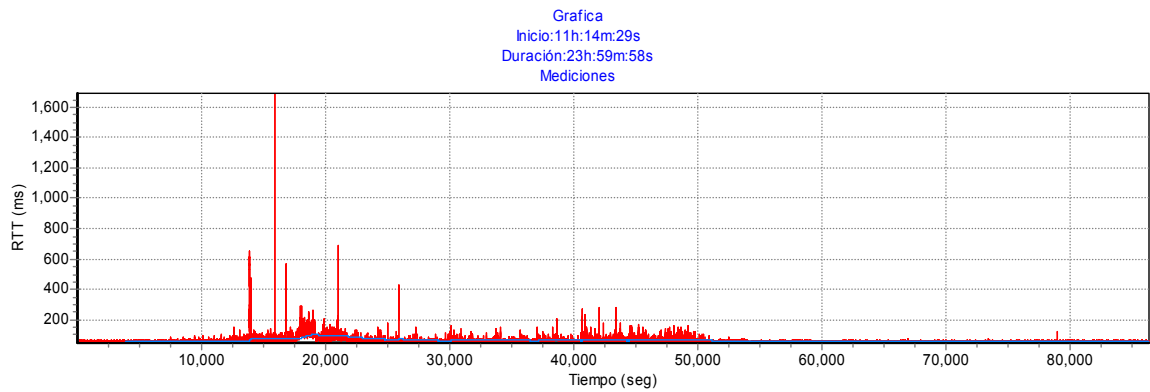
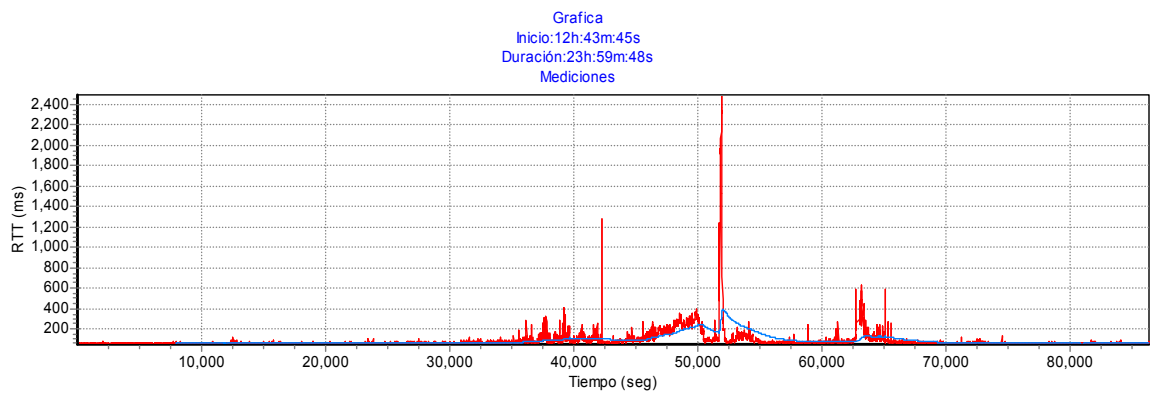


Figura 14. Series de tiempo para el sitio www.unam.mx a) 80ms b) 160ms c) 1000ms. La línea sobre la serie es un moving average de 1hr.

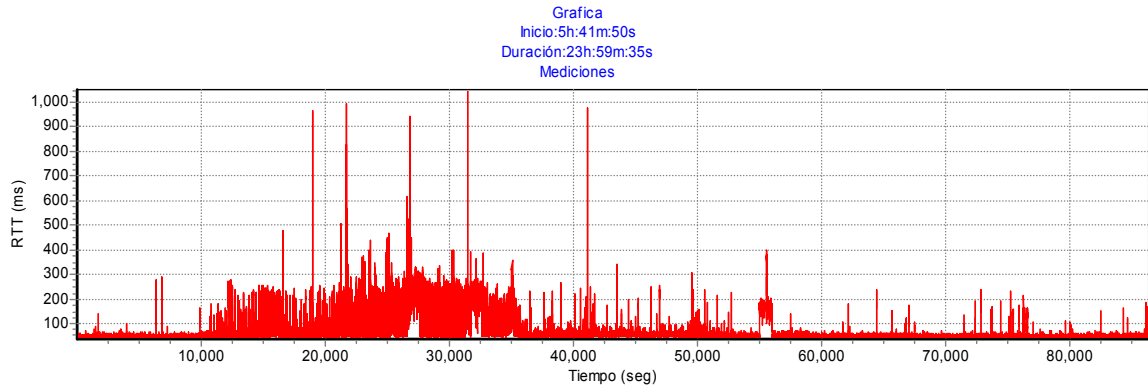


(a)

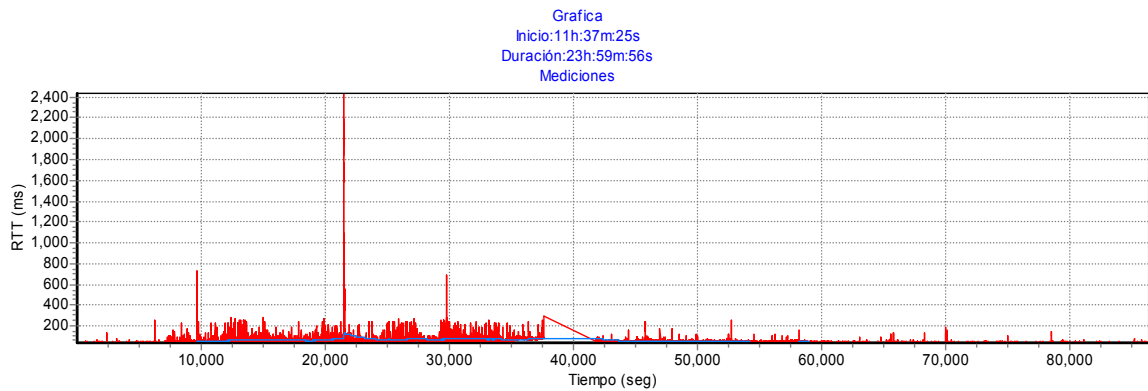


(b)

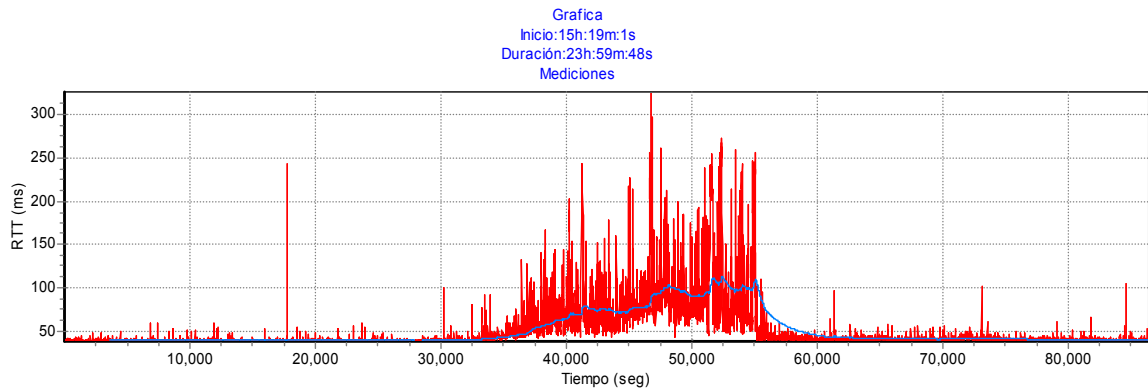
Figura 15. Series de tiempo para el sitio www.inaoep.mx a) 160ms b) 1000ms. La línea sobre la serie es un moving average de 1hr.



(a)



(b)



(c)

Figura 16. Series de tiempo para el sitio www.oaxaca.mx a) 80ms b) 160ms c) 1000ms. La línea sobre la serie es un moving average de 1hr.