

UNIVERSIDAD CENTROCCIDENTAL

“LISANDRO ALVARADO”

**DESARROLLO DE UNA HERRAMIENTA AUTOMATIZADA QUE
PERMITA MODELAR EN TIEMPO REAL EL CONSUMO DE ANCHO DE
BANDA DEL ENLACE WAN DE UNA RED**

OLGA CRISTINA PALMA URDANETA

Barquisimeto, 2006

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”
DECANATO DE CIENCIAS Y TECNOLOGÍA
POSTGRADO EN CIENCIAS DE LA COMPUTACIÓN

**DESARROLLO DE UNA HERRAMIENTA AUTOMATIZADA QUE
PERMITA MODELAR EN TIEMPO REAL EL CONSUMO DE ANCHO DE
BANDA DEL ENLACE WAN DE UNA RED**

Trabajo presentado para optar al grado de
Magíster Scientiarum en Cs. de la Computación

AUTOR: OLGA CRISTINA PALMA URDANETA
TUTOR: ARSENIO ANTONIO PEREZ PEREZ

Barquisimeto, 2006

**DESARROLLO DE UNA HERRAMIENTA AUTOMATIZADA QUE
PERMITA MODELAR EN TIEMPO REAL EL CONSUMO DE ANCHO DE
BANDA DEL ENLACE WAN DE UNA RED**

Por: OLGA CRISTINA PALMA URDANETA

Trabajo de grado aprobado

(Jurado 1)

Tutor

(Jurado 2)

(Jurado 3)

Barquisimeto, ___ de _____ de 200_

INDICE GENERAL

	Página
INDICE DE GRAFICOS.....	v
INDICE DE TABLAS.....	vi
INDICE DE FIGURAS.....	vii
RESUMEN.....	ix
INTRODUCCIÓN.....	1
CAPITULO	
I EL PROBLEMA.....	3
Planteamiento del Problema.....	3
Objetivos de la Investigación.....	6
Justificación e Importancia.....	6
Alcances y Limitaciones.....	7
II MARCO TEORICO.....	9
Antecedentes.....	9
Bases Teóricas.....	16
III MARCO METODOLOGICO.....	59
Tipo de Investigación.....	59
Población y Muestra.....	60
Fases del Estudio.....	61
IV ANÁLISIS DE LOS RESULTADOS.....	63
Elaboración de la herramienta.....	63
Prueba de la herramienta.....	69
Análisis y Comparación de Resultados.....	80
V CONCLUSIONES Y RECOMENDACIONES.....	81
VI SOFTWARE DE IMPLEMENTACION.....	84
REFERENCIAS BIBLIOGRAFICAS.....	91
ANEXOS	94
A. Tabla T-Student	94

INDICE DE GRÁFICOS

pág.

GRÁFICO

1	Correspondientes a los datos analizados el día Lunes.....	77
2	Correspondientes a los datos analizados el día Martes	78
3	Correspondientes a los datos analizados el día Miércoles.....	78
4	Correspondientes a los datos analizados el día Jueves.....	79
5	Correspondientes a los datos analizados el día Viernes.....	79

INDICE DE TABLAS

		pág.
TABLA		
1	Operacionalización de las Variables.....	57
2	Varianza de los Errores del Modelo, Varianza y Desviación estándar de los protocolos (DNS, FTP, HTTP ,POP3, SMTP).....	75
3	Valores de los Intervalos de Confianza para cada protocolo (DNS, FTP, HTTP ,POP3, SMTP).....	76

INDICE DE FIGURAS

		pág.
FIGURA		
1	Modelo de Capas de TCP/IP.....	18
2	Servicio basado en TCP/IP.....	22
3	Tipos y ancho de banda de la línea WAN.....	24
4	Tecnología WAN.....	25
5	Comparación de los tipos de gráficos en la WAN.....	25
6	La capa de enlace de datos WAN.....	27
7	Protocolo de Enlace de datos de las WAN.....	28
8	Conmutación de Circuitos.....	29
9	Conmutación de Paquetes.....	30
10	Opciones de enlace WAN.....	32
11	WAN con X.25.....	33
12	Frame Relay.....	34
13	ATM.....	35
14	Tecnología DSL.....	36
15	WAN con ISDN.....	38
16	WAN con una línea Alquilada.....	38
17	Ventana principal de Ethereal luego de una captura.....	48
18	Iris Analyzer.....	48
19	Etherpeek.....	49

20	NTOP.....	49
21	MRTG.....	53
22	RRD Tools.....	54
23	Routers2.cgi	55
24	STG.....	55
25	PRTG Traffic Grapher	56
26	Configuración de los Protocolos a Estudiar (DNS, FTP, HTTP, POP3, SMTP)	72
27	Captura de las 40 Observaciones.....	72
28	Modelo Generado en base a las 40 Observaciones.....	73
29	Intervalos de Confianza de cada Protocolo.....	75
30	Calculo del Ancho de Banda (BW).....	76
31	Escoger la fecha a Graficar.....	77
32	Menú Principal de la herramienta Modelo de Regresión Lineal Múltiple.....	84
33	Pantalla de Configuración de Protocolos.....	85
34	Pantalla de Captura de Datos para el Modelo de Regresión Lineal Múltiple.....	86
35	Pantalla de Selección de Directorio.....	87
36	Pantalla del Modelo de Regresión Lineal Múltiple.....	88
37	Pantalla que muestra el Archivo de Texto.....	89
38	Pantalla para Calcular el Ancho de Banda.....	89
39	Pantalla para Escoger la Fecha para graficar.....	90
40	Pantalla de Gráfica, clasificado por hora y fecha.....	90

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”

DECANATO DE CIENCIAS Y TECNOLOGÍA

POSTGRADO EN CIENCIAS DE LA COMPUTACIÓN

**DESARROLLO DE UNA HERRAMIENTA AUTOMATIZADA QUE
PERMITA MODELAR EN TIEMPO REAL EL CONSUMO DE ANCHO DE
BANDA DEL ENLACE WAN DE UNA RED**

Autor (a): Olga Cristina Palma Urdaneta

Tutor (a): Arsenio Antonio Pérez Pérez

RESUMEN

El presente trabajo tiene como objetivo desarrollar una herramienta automatizada que permita modelar en tiempo real el consumo de ancho de banda del enlace WAN de una red. Para lograr dicho objetivo se realizaron las siguientes actividades: a) un análisis de los requerimientos para el desarrollo de la herramienta, b) elaboración de un prototipo de la herramienta automatizada, c) programación de la herramienta la cual validará que la información capturada y clasificada por protocolo de aplicación, cumple con un modelo de regresión lineal, y posterior a este chequeo se genere el modelo en tiempo real que describe el consumo de ancho de banda del enlace WAN de la red, permitiendo al administrador tomar decisiones acertadas en relación a los futuros requerimientos de ancho de banda para las diferentes aplicaciones y d) validación de los resultados generados por la herramienta. La investigación se realizó bajo la modalidad de proyecto especial. La misma se desarrolló, para permitir que el modelo de predicción de tráfico que se generó en investigaciones anteriores a través de un trabajo manual de tipo estático y poco flexible, solvente el problema de rigidez de dicho modelo, debido que la herramienta generó de forma dinámica (en tiempo real) el modelo, lo que permite que sea flexible a cualquier organización donde se implemente.

Palabras Clave: Modelar, Tiempo real, Ancho de banda, Redes WAN, Protocolos.

INTRODUCCIÓN

A la hora de gestionar una red de comunicaciones de medio y gran tamaño, se necesitan de herramientas de gestión automatizadas que permitan mantener un control continuo de todas las redes. Este proceso consiste en la definición de alarmas e identificación automática de problemas que notifiquen al gestor de manera inmediata las desviaciones del comportamiento habitual de la red. De esta forma, se puede aislar el problema y actuar en consecuencia.

Una plataforma de análisis y monitorización de tráfico es una herramienta de gran utilidad para los gestores de red, ya que les permite estudiar las características del tráfico cursado por las redes, ayudando a conocer el uso que se está dando a las infraestructuras. Actualmente las redes de área extensa están evolucionando a escenarios de alta velocidad en los que se hace necesario una monitorización de tráfico con alta fiabilidad.

En esta investigación se desarrolló una herramienta automatizada que permite modelar en tiempo real el consumo de ancho de banda del enlace WAN de una red. Su realización está motivada por el trabajo de grado de Guedez, quien generó un modelo de predicción de tráfico haciendo uso de la técnica de regresión lineal múltiple, donde la generación del modelo se efectuó en forma estática.

Este estudio se ubicó en la modalidad de proyectos especiales, sustentado en una investigación de campo, y está conformado de la siguiente forma:

El Capítulo I está comprendido por: planteamiento del problema, justificación, los objetivos de la investigación (General y Específicos), el alcance y las limitaciones del estudio.

El Capítulo II describe el Marco Teórico, en donde se expone los antecedentes de la investigación así como toda la información bibliográfica que conforman las bases teóricas que sustentan el estudio.

El Capítulo III correspondiente al Marco Metodológico especifica: tipo de investigación, fases de la metodología de la investigación.

El Capítulo IV corresponde al análisis de los resultados, en donde se especifica la elaboración de la herramienta, se efectúan pruebas de la herramienta y se realiza análisis y comparación de resultados.

El Capítulo V plantea una serie de Conclusiones y Recomendaciones sobre la investigación.

El Capítulo VI corresponde a la descripción de la herramienta.

Finalmente se presenta las referencias bibliográficas y los anexos.

CAPITULO I

EL PROBLEMA

Planteamiento del Problema

En la actualidad, Internet se ha consolidado como la red telemática más extendida y utilizada a nivel mundial. El crecimiento en lo que respecta a infraestructuras así como al número de usuarios conectados ha sido vertiginoso en los últimos tiempos (Galan, 2003).

La popularización de Internet ha traído consigo la alteración intrínseca del tráfico de la red, no sólo en lo que a volumen respecta, sino también en cuanto al tipo de aplicaciones utilizadas y, por consiguiente, a la naturaleza del tráfico que éstas generan. Así, hace tan sólo unos años los principales servicios eran el correo electrónico (e-mail), la transferencia de ficheros (ftp), las noticias (news) y la conexión a ordenadores remotos (telnet). Hoy en día, sin embargo, el tráfico WWW (World Wide Web) es el que predomina sobre el resto, a la vez que se aprecia un sensible crecimiento de otras aplicaciones, como las de audio y vídeo en tiempo real. La variabilidad y la impredecibilidad son probablemente las dos características más comúnmente aceptadas a la hora de describir el comportamiento del tráfico de Internet. Este conocimiento incierto, que dificulta cualquier intento serio de obtener un modelo adecuado a la hora de dimensionar los recursos de la red, lo es aún más cuando nos enfrentamos al empleo de nuevas tecnologías de banda ancha (Álvarez-Campana, 1998).

Esta continua evolución en el tráfico que circula por Internet hace necesario contar con herramientas adecuadas y flexibles que permitan conocer las características de dicho tráfico.

En este sentido existen herramientas en el mercado, tales como analizadores de protocolos y monitores de red, que solo arrojan información sobre el tráfico actual pero no dicen nada en relación a la predicción de requerimientos de ancho de banda

futuros. Por esta razón, se hace necesario realizar un estudio del tráfico que pasa por los enlaces de comunicación (tipo de servicios usados por los clientes, ancho de banda que consume el tráfico generado por ellos, comportamiento de estas tasas de consumo) para luego calcular, de alguna manera, los requerimientos de ancho de banda que podrían requerir las aplicaciones en el futuro (Guedez, 2005).

Motivado por esto, Guedez (2005) efectuó una investigación donde generó un modelo del consumo del ancho de banda basado en la regresión lineal múltiple, como base futura para la predicción de tráfico entrante en el enlace de Banda Ancha ADSL de la Universidad Fermín Toro en su sede de Cabudare, Estado Lara. Esto le permitió conocer el consumo de ancho de banda y predecir el grado de saturación del enlace WAN, con un nivel de confianza del 95%, donde las variables de protocolo que influyen significativamente en la variación del ancho de banda fueron HTTP y DNS, siendo HTTP la de mayor influencia. A su vez determinó una limitación de la aplicabilidad de este modelo en otras redes donde podría existir una relación directa entre el servicio WEB y el DNS, implicando una colinealidad entre estas variables que impedirían la aplicación de un modelo de regresión lineal múltiples. En el caso de la UFT, desde el punto de vista del enlace WAN, no se realizó el estudio de colinealidad, puesto que, el servicio WEB (protocolo HTTP) estaba desligado del servicio de directorio (Protocolo DNS), debido a que la organización no maneja un DNS remoto, sino que el servicio de directorio (DNS) está configurado dentro de la red LAN, permitiendo de esta forma la independencia entre estas variables.

El trabajo de Guedez implicó varias etapas: a) realizó capturas de los paquetes de tráfico que transitan por la red, apoyado en una herramienta de análisis y monitoreo de tráfico llamada EtherReal, b) determinó cuáles eran los diferentes protocolos de la capa de aplicación de la pila TCP/IP que transitan por el enlace ADSL, c) clasificó las tramas por protocolo, d) realizó la tabulación de los datos, e) realizó los cálculos de promedios de consumo diario por hora, de protocolo y de ancho de banda. Además calculó el tráfico promedio de protocolo por hora durante la semana y el consumo de ancho promedio en la misma semana y f) efectuó un análisis a través del paquete estadístico SPSS para Windows versión 11.5 en función de los datos tabulados y los

promedios encontrados, el cual muestra el ajuste de un modelo de regresión lineal múltiple que describe la interrelación entre el ancho de banda (BW) y cinco variables independientes estudiadas (HTTP, DNS, POP3, FTP, SMTP). Es de hacer notar que el modelo del consumo de ancho de banda del enlace WAN que generó Guedez (2005), fue un trabajo manual (estático) haciendo de esta manera que el modelo generado sea poco flexible al momento de adaptarlo a otra organización o cuando suceda un cambio en el comportamiento de la red..

Por tal motivo, se propone crear una herramienta que permita generar el modelo en tiempo real (dinámico), permitiéndole adaptarse a los cambios de comportamiento que puedan ocurrir en la red de la UFT u otra organización.

En este sentido, se ha decidido dividir el trabajo en dos etapas: a) la captura y la clasificación de los datos (protocolos a nivel de la capa de aplicación), la cual será atacado por el maestrante Rey González y b) desarrollo de la herramienta para generar el modelo de forma dinámica (en tiempo real), la cual será atacada en esta investigación.

En consecuencia se plantea la siguiente interrogante:

¿Es importante para cualquier organización contar con una herramienta que permita modelar en tiempo real el consumo de ancho de banda del acceso al enlace WAN?

Objetivos de la Investigación

Objetivo General

Desarrollar una herramienta automatizada que permita modelar en tiempo real el consumo de ancho de banda del enlace WAN de una red.

Objetivos Específicos

1. Analizar los requerimientos para el desarrollo de la herramienta.
2. Elaborar un prototipo de la herramienta automatizada.
3. Programar la herramienta para validar que la información capturada y clasificada por protocolo de aplicación cumple con un modelo de regresión lineal múltiple, y además generar el modelo en tiempo real.
4. Validar los resultados generados por la herramienta.

Justificación e Importancia

Las redes de computadoras son vitales para los negocios, ya que permiten compartir y acceder a la información, además de permitir una amplia comunicación entre las personas. Día a día el avance de la tecnología hace que las redes de computadoras tengan mayor velocidad, convirtiéndose en redes más complejas. Todo esto conlleva a que administrarlas sea todo un reto.

Para tener una buena administración de la red no importando la complejidad de la misma, es necesario contar con una documentación que nos explique sus características principales, de tal manera que la administración se facilite.

Existen herramientas que permiten al administrador realizar de manera eficiente su trabajo, tales como analizadores del tráfico y monitores de la red, pero que no permiten medir en forma precisa y clara cual es el grado de influencia de un determinado protocolo sobre el BW consumido en la red, ¿Qué pasa si aumenta o disminuye el tráfico generado por un determinado protocolo? Jugar con estas variables le permitiría al administrador

determinar cuáles protocolos restringir, cuáles liberar. También darse cuenta en caso de no tener tal flexibilidad, de cuanto BW debe contratar para satisfacer los requerimientos futuros de los usuarios.

Por esto, Guedez (2005), propuso una modelo de predicción de tráfico basado en la técnica de regresión lineal múltiple que permite saber el comportamiento futuro del BW ante posibles variaciones del tráfico producido por alguno de los protocolos usados en la red , facilitando una buena gestión de los recursos. Para llegar a la conclusión del estudio, el investigador realizó un proceso de análisis y depuración de los datos, para posteriormente tabularlo y generar estadísticas, lo que ocasionó un trabajo largo, exhaustivo, tedioso. Además, cuando el comportamiento de la red cambia, se tiene que realizar nuevamente el trabajo originando pérdida de tiempo y esfuerzo. En vista de esto, se puede notar lo importante de realizar este estudio creando una herramienta que genere el modelo en tiempo real aportando mayor flexibilidad en presencia de nuevos datos de entrada, y adaptarse a los cambios de comportamiento que puedan ocurrir en la red de cualquier organización.

Por otra parte, con ayuda de generadores de tráfico, esta herramienta puede ayudar a ilustrar el comportamiento y consumo de ancho de banda (BW) con fines educativos (entrenamiento de administradores).

Alcances y Limitaciones

Alcance

El actual estudio se orientó al desarrollo de una herramienta automatizada para generar un modelo en tiempo real el consumo de ancho de banda del enlace WAN de una red, basado en la técnica de regresión lineal múltiple.

Limitaciones

Se asumió que los datos de entrada para la herramienta a desarrollar ya están clasificados según la capa de aplicación de la pila TCP/IP, los cuales van a representar los diferentes servicios que van utilizar los usuarios de una determinada red.

CAPÍTULO II

MARCO TEÓRICO

Antecedentes de la Investigación

Debido a transformación tecnológica y de alteración en los patrones de tráfico en Internet, se presentan dificultades extraordinarias a la hora de dimensionar y gestionar los recursos de la red. El hecho de que el tráfico de Internet responda a patrones complejos, no bien conocidos, impide disponer de modelos de tráfico fiables. Esta complejidad radica principalmente en la gran diversidad de servicios y perfiles de usuario, que al superponerse en la Red se manifiestan en la agregación de innumerables flujos de tráfico con características muy diferentes (Alvarez-Campana, 1998)

Por tal motivo, una gran parte de los estudios acerca del comportamiento del tráfico en redes de banda ancha, han sido fundamentalmente de carácter teórico. Sin afán de menospreciar este tipo de estudios, si cabe al menos plantearse la validez práctica de los resultados partiendo del conocimiento incierto que se tiene sobre la naturaleza del tráfico IP. En este sentido, la disponibilidad de medidas detalladas sobre escenarios reales, de modo que puedan contrastarse resultados teóricos y prácticos, puede ser el elemento clave que permita profundizar en la caracterización del tráfico de la redes (Alvarez-Campana, 1998). A continuación se presentan ciertos estudios relacionado con el análisis y predicción del tráfico de redes que hacen usos de herramientas y diversos modelos.

Galan y otros (2000) propusieron una plataforma de monitorización y análisis de tráfico para redes IP, desarrollada en el marco del proyecto MIRA. Una plataforma de análisis y monitorización de tráfico es una herramienta de gran utilidad para los operadores, ya que les permite estudiar las características del tráfico cursado por sus redes, ayudándoles a conocer el uso que se está dando a sus infraestructuras. De esta forma, MIRA es una herramienta avanzada que incorpora novedosas características,

como el análisis automático de contenidos (no únicamente cabeceras de protocolos), posibilidades de despliegue distribuido, detección de ataques de seguridad y soporte para IPv6, entre otras. Además, es versátil y soporta distintas tecnologías de subred, como Ethernet o ATM.

Galan expone que, MIRA es un sistema de captura y monitorización avanzada de tráfico altamente configurable y flexible, adecuado para su utilización tanto en redes de área local como en redes de gran cobertura. La posibilidad de ser desplegado siguiendo una arquitectura distribuida escalable permite obtener un análisis homogéneo del tráfico agregado sobre un conjunto heterogéneo de subredes.

Además de su capacidad de distribución, MIRA posee otras características que la dotan de un valor que permite el análisis de contenido mas allá de las cabeceras de los protocolos de comunicaciones, que es la clave que ha permitido dotar a la plataforma de dos novedosas funcionalidades: la detección de contenidos de tipo lúdico como es el audio MP3 y la detección de incidentes de seguridad (ya sea en modo pasivo o con la posibilidad de respuestas activas en tiempo real).

Villadangos y otros (2001) presentan una arquitectura de estimador del ancho de banda basada en una red neuronal que permite predecir el ancho de banda requerido para asegurar el servicio sus clientes. Se presenta los resultados de su aplicación para el caso del enlace de entrada y salida de la Universidad Pública de Navarra.

El sistema se basa en el uso de un estimador de la carga de la red que permite predecir con horas de anticipación los bytes totales transferidos por los usuarios. El estimador consiste en un filtro FIR realizado mediante una red neuronal.

Los datos para entrenar la red neuronal se obtienen de la captura de todos los paquetes que circulan por el punto de acceso (TR1) de la Universidad Pública de Navarra con RedIris, la troncal que une a las universidades españolas con el resto de Internet. El tráfico analizado se puede considerar representativo del tráfico que maneja un ISP, ya que la universidad tenía un alto número de usuarios que hacen uso de muy diversos servicios de Internet.

Además Villadongos y otros (2001) la utilidad de las redes neuronales como estimadores de tráfico en redes de datos, las cuales se caracterizan por tener una alta

variabilidad en el patrón del tráfico que circula por ellas. La estimación del ancho de banda permite prever con suficiente antelación el ancho de banda que debe contratar el proveedor de servicios para dar servicio a sus usuarios. Esto permite que los proveedores puedan contratar un ancho de banda básico y activar líneas de respaldo cuando la demanda de ancho de banda se prevea muy alta. Esto permitiría reducir costes de contratación a los proveedores de servicios.

Villadongos y otros (2001) mostró la posibilidad de integrar de manera natural en el estimador de entradas de gran contenido semántico como son el día de la semana y la hora del día además de los valores del ancho de banda requerido por los usuarios en horas anteriores. Estas entradas no se pueden integrar en la realización de estimadores clásicos y, sin embargo, son de gran utilidad para la realización de las previsiones de un operador de red porque tienen en cuenta datos importantes como la estacionalidad del tráfico.

Villadongos y otros (2001) expresan que las redes neuronales muestran su capacidad para predecir el tráfico que circula por un enlace de comunicaciones, aunque se debe buscar una arquitectura de red u otro tipo de sistemas para ajustar la estimación a la carga real del sistema e intentar la predicción con mayor antelación.

Borray y otros (2001), exponen cómo se debe realizar el modelado de redes de telecomunicaciones de acceso fijo para tráfico caracterizado por procesos estocásticos de Poisson y procesos auto-similares, haciendo una diferencia entre estos dos tipos para la solución del modelo. Se caracterizan los elementos de red por medio de la teoría de colas tradicional y a partir de estas adaptarlas a procesos de entrada con la propiedad de dependencia a largo plazo (autosimilares). Igualmente se muestran los procedimientos de generación de tráfico como base para la caracterización de la red utilizando procesos de arribo de tráfico como entrada.

Para este estudio se tomaron en cuenta ciertos parámetros, el cálculo de variables como el tráfico en la red, retardos, porcentaje de pérdidas de información y utilización de la red son parte fundamental en la planeación de la red y establecen los objetivos de calidad sobre los cuales ésta se mantendrá.

Borray y sus colaboradores (2001), realizaron un estudio relacionado con el modelado matemático del tráfico de redes de telecomunicaciones de acceso fijo, en el cual exponen que, el objetivo es estimar cuantitativamente el comportamiento del tráfico real mediante la obtención de parámetros, que entregan la información necesaria para caracterizar el tráfico y al adecuarlos, determinar los parámetros de entrada a una simulación. Para obtener el valor de dichos parámetros se necesita la ayuda de los estimadores, los cuales nos entregan una medida aproximada de la realidad. Dependiendo del tipo de estimador se pueden obtener diferentes grados de confiabilidad. Los criterios de selección de los estimadores a utilizar surgen de la necesidad de obtener el parámetro Hurst a partir de la medición de las propiedades que presenta el tráfico auto-similar, además de determinar en que instante de tiempo se presentan dichas características en el tráfico, para que la estimación sea la más cercana a la realidad. Igualmente, es necesario hacer un análisis independiente de la cantidad de tráfico o de las variaciones de las amplitudes. Esto se logra mediante el análisis en frecuencia.

Por último, para caracterizar el tráfico en las redes es necesario conocer la variación de la demanda durante largos intervalos de tiempo (perfil de tráfico), por ejemplo un día. Esto permite determinar el lapso de tiempo en el cual el tráfico presente baja variabilidad pero sea máxima la demanda para que la estimación sea de una alta confiabilidad y lograr el correcto dimensionamiento o determinación de rendimiento de los componentes de la red a estudiar.

En esta investigación los autores concluyeron que la determinación del valor del parámetro Hurst en el tráfico es indispensable para realizar el modelado, ya que determina el comportamiento auto-similar que presenta el tráfico de datos. La intensidad del tráfico de datos es superior con respecto al tráfico telefónico. Se necesita mayor capacidad en los equipos de redes de datos. Los modelos son de gran flexibilidad y permiten la adaptación de nuevos modelos, con herramientas matemáticas más robustas (que pueden ser expuestas en estudios posteriores) pero conservando los conceptos expuestos. Aplicando los modelos se logra predecir o

describir el comportamiento de una red y establecer una relación Costo – Eficiencia en el diseño de la red.

Magaña y otros (2001), exponen que por lo general las empresas proveedores de servicios de Internet (ISP), en el reparto del ancho de banda del acceso a Internet pueden seguir dos filosofías: (a) garantizar a cada usuario un ancho de banda mínimo, el contratado por sus usuarios, o (b) limitar el ancho de banda máximo a cada usuario. Todo ello con el fin de proveer calidad de servicio. En el primer caso, se tiene que los usuarios como mínimo usan el ancho de banda contratado. El ancho de banda sobrante, se puede repartir entre los usuarios activos. Por tanto, los usuarios pueden estar consumiendo un ancho de banda mayor que el contratado. En el segundo caso, se trata de que cada usuario únicamente use el ancho de banda que ha contratado. En caso de que el ancho de banda del acceso de Internet sea mayor que todo el ancho de banda de los usuarios activos, el ancho de banda sobrante no se reparte entre ellos. Por tanto, no se aprovecha todo lo posible el ancho de banda del acceso a Internet en caso de que algún usuario requiera poco. En este trabajo se exponen los resultados experimentales de una plataforma de planificación de tráfico. Los resultados han sido obtenidos de una red experimental con dos máquinas transmisoras y una receptora, los cuales se conectan mediante un router con características de planificación sobre plataforma LINUX. El objeto del sistema es probar las posibilidades del sistema de planificación para limitar el ancho de banda por usuario y por servicio.

El planificador hace uso de un esquema de planificación basado en que cada usuario es atendido por una cola, determinada por su dirección IP. Por tanto, todos los paquetes de sus conexiones en curso se almacenan en dicha cola para la cual, según el diseño, se garantiza una tasa mínima de servicio. Este tipo de planificación proporciona el ancho de banda mínimo que se le asigna al usuario. Sin embargo, no se limita a ese ancho de banda, sino que en caso de que exista ancho de banda disponible lo reparte entre los usuarios que están transmitiendo. Esto deriva en situaciones en las que los usuarios están utilizando mayor ancho de banda que el contratado, lo cual redundará en un mejor servicio de cara al usuario, ya que los retardos en sus comunicaciones serán menores.

Puesto que el planificador no tiene como tarea principal el ser un limitador de tasas de salida, se le ha añadido a la salida de cada clase un limitador de tasa TBF (Token Bucket Filter). Permitiendo así, garantizar una tasa mínima al conjunto de conexiones en curso.

Barakat y otros (2005) diseñaron un modelo de tráfico que puede ser usado en router o por administradores de redes para ayudarles en el diseño y manejo de la red. Ellos optaron por un modelo de tráfico basado en nivel de flujo. En este modelo un flujo es una noción muy genérica, puede ser una conexión TCP o un flujo de UDP, o puede ser una dirección de destino prefijada. La definición de un flujo es deliberadamente mantenida en forma muy general, lo cual permite utilizar el modelo en diferentes aplicaciones y a diferentes mecanismos de transportes. Haciendo el modelo específico para un cierto tipo de transporte, se pueden esperar mejores resultados.

El flujo de datos llega al enlace principal en tiempos aleatorios, se transporta un volumen aleatorio de datos, y permanece activo por periodos aleatorios. Dada la información sobre el flujo, el modelo computa la tasa total de datos observada en el enlace. El interés se dirige a capturar las dinámicas de la tasa total de datos en cortos periodos de tiempo (cientos de milisegundos). Esta dinámica puede ser completamente caracterizada mediante el uso de simples herramientas matemáticas, principalmente los procesos shot-noise. La principal contribución del modelo es el cómputo de simples expresiones para importantes medidas del tráfico en el troncal (backbone) de la red, tales como su promedio, su varianza y su función de auto correlación.

El modelo puede ser de utilidad para manejar y dimensionar redes IP. El conocer el promedio y la varianza del tráfico permite a un ISP la provisión de los enlaces principales de manera de evitar la saturación.

En el estudio anteriormente mencionado, los autores concluyeron que la función de auto correlación del tráfico puede ser usada con propósitos predictivos para sus futuros valores. Una aplicación interesante es el uso de la predicción en cortos periodos de tiempo, para optimizar el enrutamiento de paquetes y balanceo de las

cargas. El modelo también puede ser usado para fijar el impacto de los cambios de tráfico en el enlace realizados en el resto de la Internet, tales como la adición de un nuevo cliente, una nueva aplicación o un nuevo mecanismo de transporte. El ISP puede planificar el aprovisionamiento de su backbone de manera de absorber los cambios resultantes de tráfico antes de que estos cambios tengan lugar.

En el trabajo de Guedez (2005), la investigación se orientó a proponer un modelo para el modelado del tráfico en los enlaces de Banda Ancha de la Universidad Fermín Toro. Estos enlaces usan la tecnología de línea de abonados digital asimétrica (ADSL) para el acceso a Internet, proporcionado por la empresa Compañía Anónima Nacional Teléfonos de Venezuela CANTV, en el cual la velocidad de transmisión en ambos sentidos (subida y bajada) es distinta. El autor solo abordó el enlace de bajada por ser de interés para el administrador, ya que esta es la que más consume ancho de banda. Guedez, propuso crear un modelo basado en regresión lineal múltiple, técnica que permite estimar el valor de una variable a partir de valores de otras variables conocidas siempre que no exista relación de dependencia entre éstas.

En esta investigación, el análisis del tráfico efectuado al enlace entrante permitió constatar que el consumo de ancho de banda diario mantuvo valores cercanos al contratado. Al discriminar los protocolos de la capa de aplicación del protocolo TCP/IP, se encontró que el protocolo HTTP fue el que más consumió ancho de banda y que el resto de los protocolos tuvieron muy poco consumo y, por lo tanto, poca influencia en el consumo del ancho de banda. Además, el modelo propuesto para la estimación del ancho de banda en la red de la UFT, tuvo estadísticamente un nivel de confianza del 95%, el cual permitió calcular en forma anticipada el efecto de poner en práctica una nueva aplicación o política de uso de la red sobre el ancho de banda sin poner en peligro el funcionamiento de la red.

Es de hacer notar, que el trabajo que efectuó el investigador fue largo, exhaustivo, tedioso y propenso a errores, ocasionando poca flexibilidad al modelo lo que no permite adaptarse a cualquier organización, el modelo generado es estático debido a que si ocurre un cambio en la red, se debe realizar nuevamente el estudio para generar

el modelo de predicción, en vista de esto, Guedez (2005) en su trabajo recomienda desarrollar el modelo en tiempo real.

De los antecedentes de investigaciones planteados anteriormente, no se encontró un estudio que hiciera uso de una herramienta automatizada que permita modelar en tiempo real el consumo de ancho de banda de la red basándose en el análisis de regresión múltiple.

Bases Teóricas

Descargar una página Web o enviar un mensaje de e-mail implica transmitir varios paquetes a través de Internet, de allí que la posibilidad de disponer de un modelo de predicción de tráfico podría caracterizar de manera adecuada el comportamiento de las aplicaciones que emplean los usuarios en un entorno de trabajo. La predicción de tráfico es un aspecto básico para una adecuada planificación de los recursos en los sistemas de comunicaciones. Se hace necesario respaldar el trabajo de grado de manera teórica, partiendo de la conceptualización del protocolo TCP/IP para ilustrar los diferentes protocolos que existen a nivel de aplicación con los cuales se van a trabajar. Se caracterizan las redes de área amplia (WAN) donde se exponen los conceptos de las diferentes tecnologías utilizadas para acceder sus enlaces WAN. Finalmente, se expuso la teoría relacionada con la técnica de regresión lineal múltiple (técnica estadística para generar el modelo de predicción) y algunas características principales de analizadores y monitores de red.

Protocolo TCP/IP

Según Gordo Saez (1998), TCP/IP es el protocolo común utilizado por todos los ordenadores conectados a Internet, de manera que éstos puedan comunicarse entre sí. Hay que tener en cuenta que en Internet se encuentran conectados ordenadores de

clases muy diferentes y con hardware y software incompatibles en muchos casos, además de todos los medios y formas posibles de conexión. Aquí se encuentra una de las grandes ventajas del TCP/IP, pues este protocolo se encargará de que la comunicación entre todos sea posible. TCP/IP es compatible con cualquier sistema operativo y con cualquier tipo de hardware.

TCP/IP no es un único protocolo, sino que es en realidad lo que se conoce con este nombre es un conjunto de protocolos que cubren los distintos niveles del modelo OSI. Los dos protocolos más importantes son el TCP (*Transmission Control Protocol*) y el IP (*Internet Protocol*), que son los que dan nombre al conjunto. En Internet se diferencian cuatro niveles o capas en las que se agrupan los protocolos, y que se relacionan con los niveles OSI de la siguiente manera:

- **Aplicación:** Se corresponde con los niveles OSI de aplicación, presentación y sesión. Aquí se incluyen protocolos destinados a proporcionar servicios, tales como correo electrónico (SMTP), transferencia de ficheros (FTP), conexión remota (TELNET) y otros más recientes como el protocolo HTTP (*Hypertext Transfer Protocol*).
- **Transporte:** Coincide con el nivel de transporte del modelo OSI. Los protocolos de este nivel, tales como TCP y UDP, se encargan de manejar los datos y proporcionar la fiabilidad necesaria en el transporte de los mismos.
- **Internet:** Es el nivel de red del modelo OSI. Incluye al protocolo IP, que se encarga de enviar los paquetes de información a sus destinos correspondientes. Es utilizado con esta finalidad por los protocolos del nivel de transporte.
- **Enlace:** Los niveles OSI correspondientes son el de enlace y el nivel físico. Los protocolos que pertenecen a este nivel son los encargados de la transmisión a través del medio físico al que se encuentra conectado cada *host*, como puede ser una línea punto a punto o una red *Ethernet*.

El TCP/IP necesita funcionar sobre algún tipo de red o de medio físico que proporcione sus propios protocolos para el nivel de enlace de Internet. Por este

motivo hay que tener en cuenta que los protocolos utilizados en este nivel pueden ser muy diversos y no forman parte del conjunto TCP/IP. Sin embargo, esto no debe ser problemático puesto que una de las funciones y ventajas principales del TCP/IP es proporcionar una abstracción del medio de forma que sea posible el intercambio de información entre medios diferentes y tecnologías que inicialmente son incompatibles.

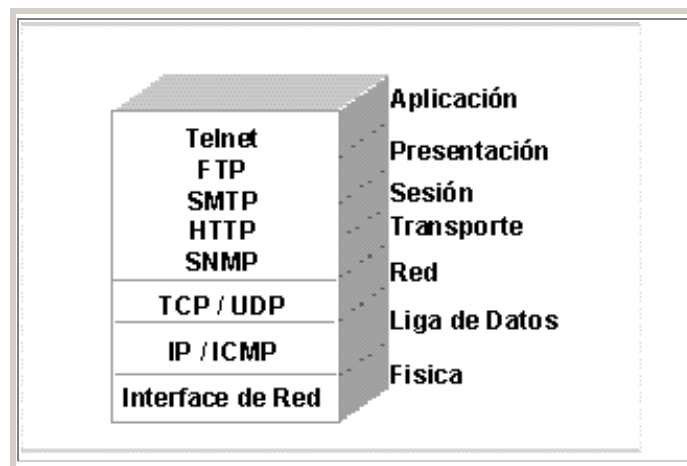


Figura 1. Modelo de Capas de TCP/IP

Fuente: <http://www.uca.edu.sv/investigacion/tutoriales/tcp-ip.html>

Protocolos de la capa de Aplicación

SMTP: El significado de las siglas de **SMTP** es **Protocolo Simple de Transmisión de Correo** (*Simple Mail Transfer Protocol*). Este protocolo está descrito en la **RFC 821** y es el estándar de Internet para el intercambio de correo electrónico. SMTP es un protocolo independiente del subsistema de transmisión usado. Necesita que el subsistema de transmisión ponga a su disposición un canal de transmisión fiable y con entrega ordenada, con lo cual el uso del protocolo TCP en la capa de transporte es lo adecuado. Para que dos sistemas intercambien correo mediante el protocolo SMTP, no es necesario que exista una conexión interactiva, ya que este protocolo usa métodos de almacenamiento y reenvío de mensajes.

SNMP o Protocolo Simple de Gestión de Redes: es el protocolo de gestión de red más importante y usado en la actualidad. El protocolo SNMP se utiliza principalmente para monitorizar y controlar el estado de los recursos conectados a las redes IP, en especial equipos de intercomunicación (conmutadores, enrutadores, módems), aunque se puede usar en cualquier tipo de equipo que permita ejecutar un agente SNMP. Forma parte del conjunto de protocolos TCP/IP y está definido en la capa de aplicación del mismo. SNMP busca la sencillez y es por ello que en la capa de transporte está soportado por el protocolo UDP (caracterizado por su rapidez). Este protocolo permite a un servidor TCP/IP que ejecuta una aplicación SNMP, interrogar a otros nodos para estadísticas y condiciones de error de la red. Los otros servidores, que proporcionan agentes SNMP responden a estas preguntas y le permiten a un solo servidor recoger estadísticas de muchos nodos de la red.

DNS: En el grupo de protocolos TCP-IP se encuentran los protocolos de resolución de nombres por direcciones IP. Estos protocolos permiten a las aplicaciones tener acceso a los servicios de un computador a través del uso de un nombre. Para ello debe existir un mecanismo que permita la resolución y asociación de una dirección IP por un nombre. El mecanismo de asociación consiste en una base de datos donde se encuentran las asociaciones de una dirección IP con su nombre respectivo. Y el mecanismo de resolución consiste en identificar cuál es la dirección IP asociada a un nombre. De esta manera los computadores de la red pueden ser accedidos a través de un nombre en vez de su dirección IP.

El protocolo DNS traduce un nombre de Internet, como por ejemplo "www.cisco.com", en una dirección IP. Muchas aplicaciones se apoyan en los servicios de directorio que provee el DNS para esta tarea. Todos los navegadores de web, los programas de correo electrónico y los programas de transferencia de archivos utilizan nombres para identificar a los sistemas remotos. El protocolo DNS permite que estos clientes soliciten a los servidores DNS de la red la traducción de dichos nombres a direcciones IP. Las aplicaciones pueden entonces utilizar dichas direcciones para enviar sus mensajes. Sin este servicio de directorios de búsqueda, la Internet sería casi imposible de usar.

Un Sistema de Nombres de Dominio está compuesto de una base de datos distribuida de nombres. Los nombres en la base de datos DNS generan una estructura lógica en forma de árbol conocida como espacios de nombre de dominio. Cada nodo o dominio en el espacio de nombres de dominio es nombrado y puede contener subdominios. Los dominios y subdominios están agrupados en zonas que permiten la administración distribuida del espacio de nombre. El nombre de dominio identifica la posición del dominio en el árbol lógico de DNS en relación a su dominio padre, separando cada rama del árbol con un punto ".".

POP3: El significado de las siglas POP es Protocolo de Oficina de Correos (*Post Office Protocol*). Permite a los clientes de correo electrónico recuperar los mensajes de los servidores remotos y guardarlos en las máquinas locales. La mayoría de los clientes de correo que utilizan el protocolo POP se configuran automáticamente para eliminar el mensaje del servidor de correo después de transferirlo correctamente al sistema del cliente, aunque esto se puede cambiar.

La ventaja principal que tiene este protocolo es que carpetas, mensajes, se guardan en nuestro ordenador, lo que permite leer el correo recibido sin estar conectado a la red. Además, al leer los mensajes y bajarlos al ordenador, se libera espacio en el buzón del Host, por lo cual tiene menos probabilidades que por descuido se llene el buzón y no se pueda recibir más mensajes. Es el más extendido (prácticamente todos los programas de correo lo soportan) y es el ideal para conectarse siempre desde un mismo ordenador.

TELNET: es el nombre de un protocolo (y del programa informático que implementa el cliente) que sirve para acceder mediante una red a otra máquina, para manejarla como si estuviéramos sentados delante de ella. Para que la conexión funcione, como en todos los servicios de Internet, la máquina a la que se accedía debe tener un programa especial que reciba y gestione las conexiones. El puerto que se utiliza generalmente es el 23. Sólo sirve para acceder en modo terminal, es decir, sin gráficos, pero fue una herramienta muy útil para arreglar fallos a distancia, sin necesidad de estar físicamente en el mismo sitio que la máquina que los tenía. También se usaba para consultar datos a distancia, como datos personales en

máquinas accesibles por red, información bibliográfica.

FTP: significa Protocolo de Transferencia de Archivos y es el ideal para transferir grandes bloques de datos por la red. Por lo general llamado FTP es una utilidad para manejar archivos a través de máquinas sin tener que establecer una sesión remota con Telnet. El FTP permite transferir archivos en uno y otro sentido, administrar directorios y tener acceso al correo electrónico. El FTP no está diseñado para permitir el acceso a otra máquina para ejecutar programas, pero es la mejor utilidad para transferencias de archivos.

TFTP: es un protocolo para transferir ficheros entre distintas máquinas conectadas a través de una red de comunicaciones. Se implementa sobre un servicio de comunicaciones no fiable y no orientado a conexión. Consiste fundamentalmente en la lectura o escritura por parte de un cliente de/a un fichero de un servidor. Es mucho menos complejo que FTP, es decir, soporta menos funciones, el código es más pequeño, consume menos memoria y como consecuencia es más rápido. Sin embargo, es menos confiable que FTP y no provee mecanismos de seguridad o autenticidad.

HTTP: es el protocolo de la Web (WWW), usado en cada transacción. Las letras significan Hyper Text Transfer Protocol, es decir, protocolo de transferencia de hipertexto. El hipertexto es el contenido de las páginas Web, y el protocolo de transferencia es el sistema mediante el cual se envían las peticiones de acceder a una página Web, y la respuesta de esa Web, remitiendo la información que se verá en pantalla. HTTP es un protocolo sin estado, es decir, que no guarda ninguna información sobre conexiones anteriores. Al finalizar la transacción todos los datos se pierden.

Servicio	Protocolo TCP/IP
World Wide Web	HTTP
Transferencia de archivos	FTP, TFTP
Compartir archivos	NFS
Correo electrónico de Internet	SMTP, POP3, IMAP
Administración remota	Telnet
Servicios de directorio (Internet)	DNS, LDAP
Configuración automática de dirección de red	DHCP
Administración de red	SNMP

Figura 2. Servicio basado en TCP/IP

Fuente: Cisco System

Redes de Área Amplia (WAN)

Cuando una LAN crece y expande la cantidad de computadoras y usuarios en diversas ubicaciones o localidades, se convierte en una red de área amplia. Por lo tanto, la única cosa que hace diferente a una LAN de una WAN es su cobertura geográfica.

Según Nicara (2002), una WAN se extiende sobre un área geográfica amplia, a veces un país o un continente; contiene una colección de máquinas dedicadas a ejecutar programas de usuario (aplicaciones), estas máquinas se llaman Hosts. Los hosts están conectados por una subred de comunicación. El trabajo de una subred es conducir mensajes de un host a otro.

Según Cisco, una WAN es una red de comunicación de datos que opera más allá del alcance geográfico de una LAN. Una de las diferencias primordiales entre una WAN y una LAN es que una empresa u organización debe suscribirse a un proveedor de servicio WAN externo para utilizar los servicios de red de una operadora de servicios WAN. Una WAN utiliza enlaces de datos suministrados por los servicios de una operadora para acceder a Internet y conectar los sitios de una organización entre sí, con sitios de otras organizaciones, con servicios externos y con usuarios remotos. Las WAN generalmente transportan varios tipos de tráfico, tales como voz, datos y vídeo. Los servicios telefónicos y de datos son los servicios WAN de uso más

generalizado.

Los dispositivos de las instalaciones del suscriptor se conocen como equipo terminal del abonado (CPE). El suscriptor es dueño de un CPE o alquila un CPE del proveedor de servicios. Un cable de cobre o fibra conecta el CPE a la central telefónica del proveedor de servicio más cercano. Este cableado muchas veces se llama bucle local, o última milla. Una llamada marcada se conecta de forma local a otros bucles locales o de forma no local a través de un troncal a un centro primario. Luego se dirige a un centro de sección y luego a un centro de operación internacional o regional a medida que la llamada viaja a su destino.

Para que el bucle local transporte datos, se necesita un dispositivo como un módem que prepare los datos para su transmisión. Los dispositivos que colocan los datos en el bucle local se llaman equipos de terminación de circuito de datos, o equipos de comunicación de datos (DCE). Los dispositivos del cliente que transmiten datos al DCE se llaman equipo terminal de datos (DTE). El propósito principal del DCE es suministrar una interfaz para el DTE al enlace de comunicación en la nube WAN.

Los enlaces WAN vienen en varias velocidades medidos en bits por segundo (bps), kilobits por segundo (kbps o 1000 bps), megabits por segundo (Mbps o 1000 kbps) o gigabits por segundo (Gbps o 1000 Mbps) Los valores de bps por lo general son de full duplex. Esto significa que una línea E1 puede transportar 2 Mbps, o T1 puede transportar 1,5 Mbps en cada dirección de manera simultánea.

Tipo de línea	Estándar de señal	Capacidad de la velocidad de transmisión
56	DS0	56 Kbps
64	DS0	64 Kbps
T1	DS1	1.544 Mbps
E1	ZM	2.048 Mbps
E3	M3	34.064 Mbps
J1	Y1	2.048 Mbps
T3	DS3	44.736 Mbps
OC-1	SONET	51.84 Mbps
OC-3	SONET	155.54 Mbps
OC-9	SONET	466.56 Mbps
OC-12	SONET	622.08 Mbps
OC-18	SONET	933.12 Mbps
OC-24	SONET	1244.16 Mbps
OC-36	SONET	1866.24 Mbps
OC-48	SONET	2488.32 Mbps

Figura 3. Tipos y ancho de banda de la línea WAN

Fuente: Cisco System

Las WAN transportan varios tipos de tráfico, tales como datos, voz y video. El diseño seleccionado debe ofrecer capacidad adecuada y tiempos de tránsito que cumplan con de las necesidades de las empresas. Entre las especificaciones, el diseño debe tener en cuenta la topología de las conexiones entre varias ubicaciones, la naturaleza de aquellas conexiones y la capacidad del ancho de banda.

Como las WAN son simplemente un grupo de interconexiones entre los routers basados en las LAN, no hay servicios en la WAN. Las tecnologías WAN funcionan en las tres capas inferiores del modelo de referencia OSI. Los routers determinan el destino de los datos a partir de los encabezados de capa de red y transfieren los paquetes a la conexión de enlace de datos indicada para su envío en la conexión física

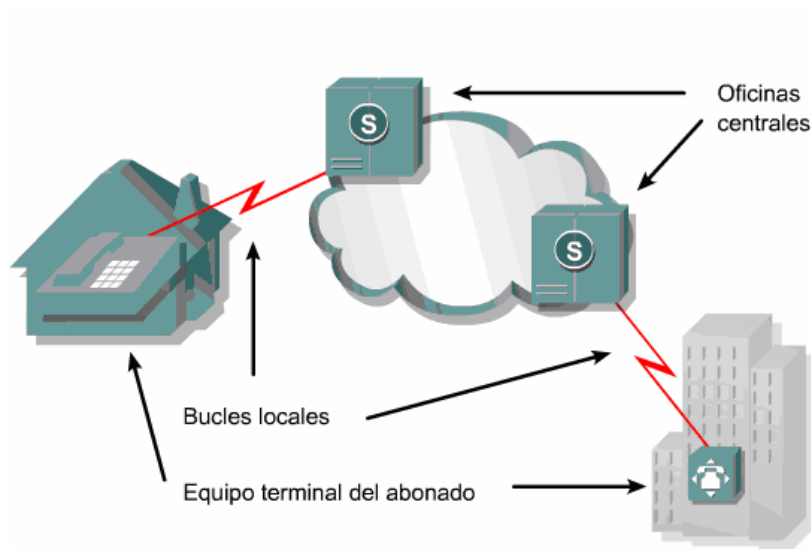


Figura 4. Tecnología Wan

Fuente: Cisco System

Al diseñar la WAN, es necesario saber qué clase de tráfico de datos se debe transportar, su origen y su destino. Las WAN transportan una variedad de tipos de tráfico que requieren diferentes anchos de banda, latencia y fluctuación.

Tráfico	Latencia	Fluctuación de fase	Ancho de banda
Voz	Baja	Baja	Medio
Datos de transacción (por ejemplo, SNA)	Medio	Medio	Medio
Mensajería (e-mail)	Alta	Alta	Alta
Transferencia de archivos	Alta	Alta	Alta
Datos en lote	Alta	Alta	Alta
Administración de red	Alta	Alta	Baja
Videoconferencia	Baja	Baja	Alta

Algunos tipos de tráfico de WAN con tolerancia a la latencia y fluctuación de fase, junto con los requisitos de ancho de banda

Figura 5. Comparación de los tipos de gráficos en la WAN

Fuente: Cisco System

Para cada par de puntos finales y para cada tipo de tráfico, se necesita información sobre las distintas características del tráfico. Determinar esto puede requerir de un estudio exhaustivo y la consulta a los usuarios de la red. El diseño con frecuencia implica actualizar, expandir o modificar una WAN ya existente. Muchos de los datos necesarios pueden provenir de estadísticas de administración de redes existentes.

Conocer los varios puntos finales permite la selección de una topología o diseño para la WAN. Las cuestiones geográficas y las necesidades, como la disponibilidad, tendrán influencia en la topología. Un gran requisito de disponibilidad requiere de enlaces adicionales que ofrezcan rutas de datos alternativos para la redundancia y el equilibrio de las cargas.

Cuando se hayan elegido los puntos finales y los enlaces, es posible estimar el ancho de banda necesario. El tráfico en los enlaces puede tener distintos requisitos de latencia y fluctuación. Una vez que se determine la disponibilidad de ancho de banda, se deben elegir las tecnologías de enlace adecuadas.

Las WAN utilizan el modelo de referencia OSI, pero se enfocan principalmente en las Capas 1 y 2. Los estándares WAN, por lo general, describen tanto los métodos de envío de la capa física como los requisitos de la capa de enlace de datos, incluyendo el direccionamiento físico, el control de flujo y el encapsulamiento. Hay varias autoridades reconocidas que definen y administran los estándares WAN.

Los protocolos de capa física describen cómo proporcionar las conexiones eléctricas, mecánicas, operativas y funcionales a los servicios brindados por un proveedor de servicios de comunicaciones.

Los protocolos de la capa de enlace de datos definen cómo se encapsulan los datos para su transmisión a lugares remotos, y los mecanismos de transferencia de las tramas resultantes. Se utiliza una variedad de tecnologías, tales como ISDN, Frame Relay o el Modo de Transferencia Asíncrona (ATM). Estos protocolos utilizan los mismos mecanismos de entramado, control de enlace de datos de alto nivel (HDLC), un estándar ISO o uno de sus subgrupos o variantes.

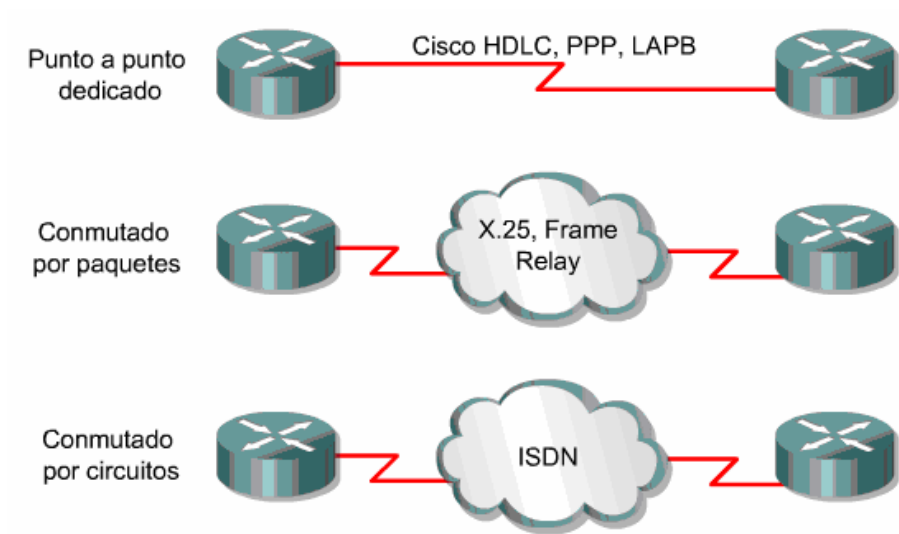


Figura 6. La capa de enlace de datos WAN

Fuente: Cisco System

Los datos de la capa de red se envían a la capa de enlace de datos para su transmisión en un enlace físico, que normalmente es de punto a punto en una conexión WAN. La capa de enlace de datos crea una trama alrededor de los datos de capa de red de modo que se apliquen los controles y verificaciones necesarios. Cada tipo de conexión WAN utiliza un protocolo de Capa 2 para encapsular el tráfico mientras atraviesa el enlace WAN. Para asegurarse de que se esté utilizando el protocolo de encapsulamiento correcto, se debe configurar el tipo de encapsulamiento de Capa 2 utilizado en cada interfaz serial del router. El protocolo de encapsulamiento que se debe usar depende de la tecnología WAN y del equipo. La mayoría del entramado se basa en el estándar HDLC. El entramado HDLC garantiza una entrega confiable de datos en líneas poco confiables e incluye mecanismos de señalización para el control de flujo y errores.

Protocolo	Uso
Procedimiento de acceso al enlace balanceado (LAPB)	X.25
Procedimiento de acceso al enlace en el canal D (LAPD).	Canal D ISDN
Trama de procedimiento de acceso al enlace (LAPF)	Frame Relay
Control de enlace de datos de alto nivel (HDLC)	Valor por defecto de Cisco
Protocolo punto a punto (PPP)	Conexiones de marcación telefónica

Figura 7. Protocolo de Enlace de datos de las WAN

Fuente: Cisco System

Red Conmutada por Paquetes y por Circuitos

Las redes conmutadas por paquetes se desarrollaron para compensar el gasto de las redes conmutadas por circuitos públicas y suministrar una tecnología WAN más económica.

Cuando un suscriptor realiza una llamada telefónica, el número marcado se utiliza para establecer switches en las centrales a lo largo de la ruta de la llamada de modo que haya un circuito continuo entre quien hace la llamada y quien recibe la llamada. Debido a la operación de conmutación usada para establecer el circuito, el sistema telefónico se conoce como **red conmutada por circuito**. Si los módems reemplazan a los teléfonos, entonces el circuito conmutado puede transportar datos de computador. Varias conversaciones comparten la ruta interna que sigue el circuito entre los centrales. Se utiliza la multiplexión por división de tiempo (TDM) para dar a cada conversación una parte de la conexión por turno. TDM garantiza que una conexión de capacidad fija esté disponible al suscriptor.

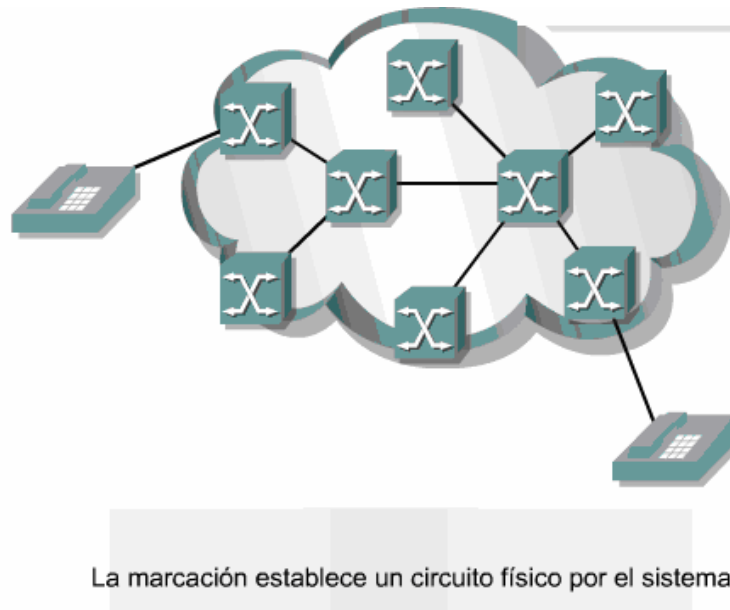


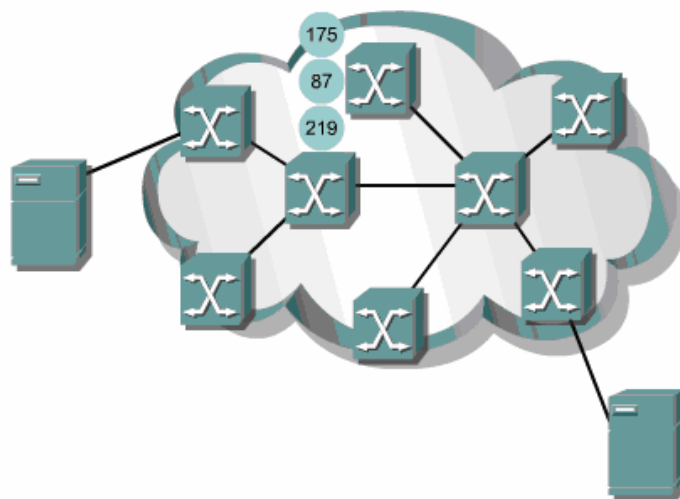
Figura 8. Conmutación de Circuitos

Fuente: Cisco System

Si el circuito transporta datos de computador, es posible que el uso de esta capacidad fija no sea eficiente. Por ejemplo, si se utiliza el circuito para tener acceso a Internet, habrá "ráfagas" de actividad en el circuito mientras se transfiere una página Web. Entonces, es posible que le siga un período sin actividad mientras el usuario lee la página y luego otra ráfaga de actividad mientras se transfiere la página siguiente. Esta variación en el uso entre máximo y nada es típica del tráfico informático de red. Como el suscriptor tiene uso exclusivo de la capacidad fija asignada, los circuitos conmutados, en general, son una forma cara de transferir datos.

Una alternativa es asignar la capacidad al tráfico solo cuando es necesario y compartir la capacidad disponible entre varios usuarios. Con una conexión conmutada por circuito, los bits de datos puestos en el circuito se transmiten de forma automática al extremo más lejano porque el circuito ya está establecido. Si es necesario compartir el circuito, tiene que haber un mecanismo para rotular los bits de modo que el sistema sepa dónde transmitirlos. Es difícil rotular bits individuales, por lo tanto, se juntan en grupos llamados celdas, tramas o paquetes. Los paquetes se transfieren de central a

central para su envío a través de la red del proveedor. Las redes que implementan este sistema se llaman **redes conmutadas por paquetes**.



Los datos rotulados se pasan de switch a switch. Es posible que tenga que esperar su turno en un enlace.

Figura 9. Conmutación de Paquetes

Fuente: Cisco System

Los enlaces que conectan estos switches en la red del proveedor pertenecen a un suscriptor individual durante la transferencia de datos, de modo que muchos suscriptores pueden compartir el enlace. Los costos pueden ser significativamente menores que en la conexión conmutada por circuitos. Los datos en redes conmutadas por paquetes están sujetos a demoras impredecibles cuando paquetes individuales esperan que los switches transmitan los paquetes de otros suscriptores.

Los switches de una red conmutada por paquetes determinan, según la información de direccionamiento en cada paquete, cuál es el siguiente enlace por el que se debe enviar el paquete. Hay dos maneras de determinar este enlace: orientada a conexión o sin conexión. Los sistemas sin conexión, tal como Internet, transmiten toda la información de direccionamiento en cada paquete. Cada switch debe evaluar la dirección para determinar dónde enviar el paquete. Los sistemas orientados a

conexión predeterminan la ruta del paquete y cada paquete necesita llevar sólo un identificador. El switch determina la ruta a seguir buscando el identificador en las tablas que tiene en su memoria. Este grupo de entradas en las tablas identifica una ruta o circuito particular a través del sistema. Si este circuito está físicamente disponible sólo mientras el paquete esté pasando por él, se llama Circuito virtual (VC).

Las entradas de la tabla que constituyen el VC se pueden establecer enviando peticiones de conexión a través de la red. En este caso, el circuito resultante se llama Circuito virtual conmutado (SVC). Los datos a transmitir en un SVC deben esperar hasta que se hayan establecido las entradas de la tabla. Una vez establecido, el SVC puede permanecer en operación durante horas, días o semanas. Cuando se requiere que un circuito esté siempre disponible, se establece un Circuito virtual permanente (PVC). Los switch cargan las entradas de la tabla durante el arranque, de modo que el PVC está siempre disponible.

La conmutación de circuitos establece una conexión dedicada para voz y datos entre el emisor y el receptor. Antes de que comience la conmutación, es necesario establecer la conexión configurando los switch.

Para conectar una red conmutada por paquetes, el suscriptor necesita un bucle local a la ubicación más cercana donde el proveedor ofrece el servicio. Esto se llama punto de presencia (POP) del servicio. Por lo general ésta es una línea alquilada dedicada. Esta línea es mucho más corta que una línea alquilada conectada directamente a las diferentes ubicaciones del suscriptor y muchas veces transporta VC. Como que es poco probable que todos los VC enfrenten la máxima demanda al mismo tiempo, la capacidad de una línea alquilada puede ser menor a la de la suma de los VC individuales. Ejemplos de conexiones conmutadas por paquetes o celdas son: Frame Relay, X.25, ATM

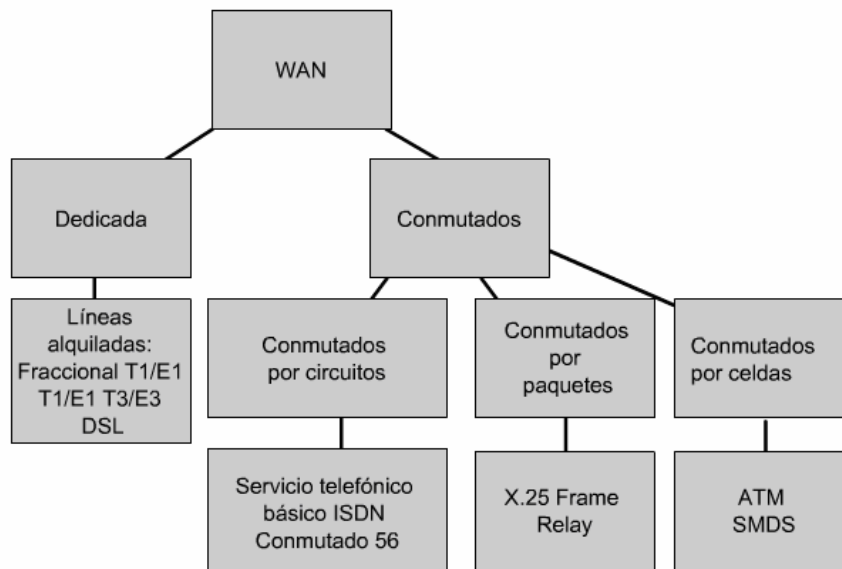


Figura 10. Opciones de enlace WAN

Fuente: Cisco System

X.25: Debido al costo de las líneas alquiladas, los proveedores de telecomunicaciones introdujeron las redes conmutadas por paquetes utilizando líneas compartidas para reducir los costos. La primera de estas redes conmutadas por paquetes se estandarizó como el grupo de protocolos X.25. X.25 ofrece una capacidad variable y compartida de baja velocidad de transmisión que puede ser conmutada o permanente.

X.25 es un protocolo de capa de red y los suscriptores disponen de una dirección en la red. Los circuitos virtuales se establecen a través de la red con paquetes de petición de llamadas a la dirección destino. Un número de canal identifica la SVC resultante. Los paquetes de datos rotulados con el número del canal se envían a la dirección correspondiente. Varios canales pueden estar activos en una sola conexión.

Los suscriptores se conectan a la red X.25 con una línea alquilada o con una conexión de acceso telefónico. Además, las redes X.25 pueden tener canales preestablecidos entre los suscriptores que proveen un PVC.

X.25 puede resultar muy económica porque las tarifas se calculan con base en la cantidad de datos enviados y no el tiempo de conexión ni la distancia. Los datos se pueden enviar a cualquier velocidad igual o menor a la capacidad de conexión. Esto ofrece más flexibilidad. Las redes X.25 por lo general tienen poca capacidad, con un máximo de 48 kbps. Además, los paquetes de datos están sujetos a las demoras típicas de las redes compartidas.

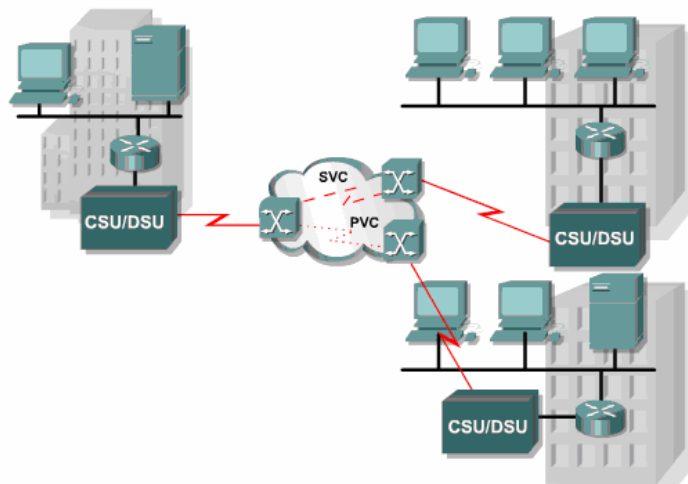


Figura 11. WAN con X.25

Fuente: Cisco System

Frame Relay: Con la creciente demanda de mayor ancho de banda y menor latencia en la conmutación de paquetes, los proveedores de comunicaciones introdujeron el Frame Relay. Aunque la configuración de la red parece similar a la de X.25, la velocidad de transmisión de datos disponible es por lo general de hasta 4 Mbps y algunos proveedores ofrecen aún mayores velocidades.

Frame Relay difiere de X.25 en muchos aspectos. El más importante es que es un protocolo mucho más sencillo que funciona a nivel de la capa de enlace de datos y no en la capa de red.

Frame Relay no realiza ningún control de errores o flujo. El resultado de la administración simplificada de las tramas es una reducción en la latencia, y las

medidas tomadas para evitar la acumulación de tramas en los switches intermedios ayudan a reducir las fluctuaciones de fase.

La mayoría de las conexiones de Frame Relay son PVC y no SVC. La conexión al extremo de la red con frecuencia es una línea alquilada, pero algunos proveedores ofrecen conexiones telefónicas utilizando líneas ISDN. El canal D ISDN se utiliza para configurar una SVC en uno o más canales B. Las tarifas de Frame Relay se calculan con base en la capacidad del puerto de conexión al extremo de la red. Otros factores son la capacidad acordada y la velocidad de información suscripta (CIR) de los distintos PVC a través del puerto.

Frame Relay ofrece una conectividad permanente, compartida, de ancho de banda mediano, que envía tanto tráfico de voz como de datos. Frame Relay es ideal para conectar las LAN de una empresa. El router de la LAN necesita sólo una interfaz, aún cuando se estén usando varios VC. La línea alquilada corta que va al extremo de la red Frame Relay permite que las conexiones sean económicas entre LAN muy dispersas.

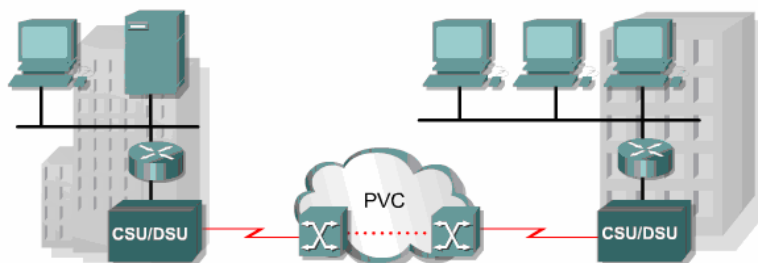


Figura 12. Frame Relay

Fuente: Cisco System

ATM: Los proveedores de comunicaciones vieron la necesidad de una tecnología de red compartida permanente que ofreciera muy poca latencia y fluctuación a anchos de banda mucho más altos. Su solución fue el Modo de Transferencia Asíncrona (ATM). ATM tiene una velocidad de transmisión de datos superior a los 155 Mbps.

Al igual que las otras tecnologías compartidas, como X.25 y Frame Relay, los diagramas de las WAN ATM se ven igual.

La tecnología ATM es capaz de transferir voz, video y datos a través de redes privadas y públicas. Tiene una arquitectura basada en celdas más bien que una basada en tramas. Las celdas ATM tienen siempre una longitud fija de 53 bytes. La celda ATM de 53 bytes contiene un encabezado ATM de 5 bytes seguido de 48 bytes de carga ATM. Las celdas pequeñas de longitud fija son adecuadas para la transmisión de tráfico de voz y video porque este tráfico no tolera demoras. El tráfico de video y voz no tiene que esperar que se transmita un paquete de datos más grande.

La celda ATM de 53 bytes es menos eficiente que las tramas y paquetes más grandes de Frame Relay y X.25. Además, la celda ATM tiene un encabezado de por lo menos 5 bytes por cada 48-bytes de datos. Cuando la celda está transportando paquetes de capa de red segmentados, la carga general será mayor porque el switch ATM tiene que poder reagrupar los paquetes en el destino. Una línea ATM típica necesita casi un 20% más de ancho de banda que Frame Relay para transportar el mismo volumen de datos de capa de red.

ATM ofrece tanto los PVC como los SVC, aunque los PVC son más comunes en las WAN.

Como las otras tecnologías compartidas, ATM permite varios circuitos virtuales en una sola conexión de línea alquilada al extremo de red.

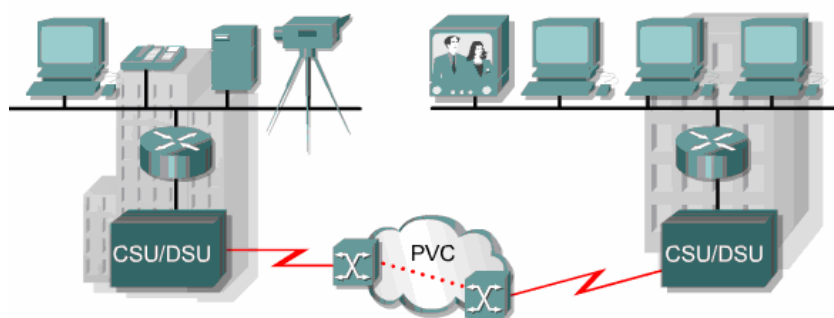
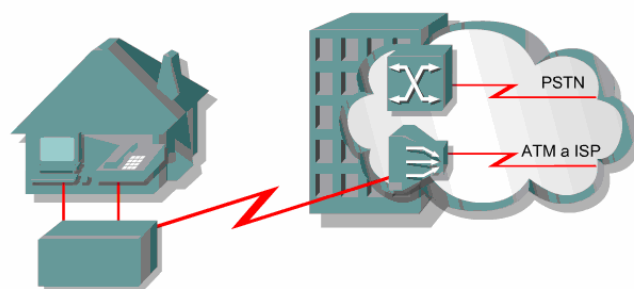


Figura 13. ATM

Fuente: Cisco System

DSL: La tecnología de línea Digital del suscriptor (DSL) es una tecnología de banda ancha que utiliza líneas telefónicas de par trenzado para transportar datos de alto ancho de banda para dar servicio a los suscriptores. El servicio DSL se considera de banda ancha, en contraste con el servicio de banda base típico de las LAN. Banda ancha se refiere a la técnica que utiliza varias frecuencias dentro del mismo medio físico para transmitir datos. El término xDSL se refiere a un número de formas similares, aunque en competencia, de tecnologías DSL: DSL Asimétrico (ADSL), DSL simétrico (SDSL), DSL de alta velocidad de bits (HDSL), ISDN (como) DSL (IDSL), DSL para consumidores (CDSL), también llamado DSL-lite o G.lite.

La tecnología DSL permite que el proveedor de servicios ofrezca a los clientes servicios de red de alta velocidad, utilizando las líneas de cobre de bucle local instaladas. La tecnología DSL permite que la línea de bucle local se utilice para realizar conexiones telefónicas de voz normales y conexiones permanentes para tener conectividad de red al instante. Las líneas del suscriptor DSL múltiples se pueden multiplexar a un enlace de alta capacidad al usar el Multiplexor de acceso DSL (DSLAM) en el sitio del proveedor. Los DSLAM incorporan la tecnología TDM para juntar muchas líneas del suscriptor a un solo medio más pequeño, en general una conexión T3/DS3. Las tecnologías DSL están utilizando técnicas de codificación y modulación complejas para lograr velocidades de transmisión de datos de hasta 8.192 Mbps.



El bucle local conecta el divisor de señal al DSLAM. Voz y datos usan intervalos de frecuencia distintos.

Figura 14. Tecnología DSL

Fuente: Cisco System

Conexión telefónica y analógica: Cuando se necesitan transferencias de datos de bajo volumen e intermitentes, los módems y las líneas telefónicas analógicas ofrecen conexiones conmutadas dedicadas y de baja capacidad.

La telefonía convencional utiliza cables de cobre, llamados bucle local, para conectar el equipo telefónico a las instalaciones del suscriptor a la red telefónica pública conmutada (PSTN). La señal en el bucle local durante una llamada es una señal electrónica en constante cambio, que es la traducción de la voz del suscriptor.

El bucle local no es adecuado para el transporte directo de datos informáticos binarios, pero el módem puede enviar datos de computador a través de la red telefónica de voz. El módem modula los datos binarios en una señal analógica en el origen y, en el destino, demodula la señal analógica a datos binarios.

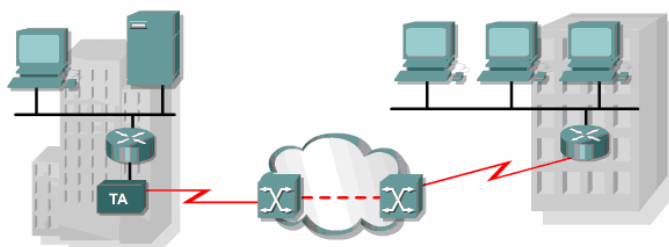
Las características físicas del bucle local y su conexión a PSTN limitan la velocidad de la señal. El límite superior está cercano 33 kbps. Es posible aumentar la velocidad a 56 kbps si la señal viene directamente por una conexión digital.

ISDN: Las conexiones internas o troncales de PSTN evolucionaron y pasaron de llevar señales de multiplexión por división de frecuencia, a llevar señales digitales de multiplexión por división de tiempo (TDM). El próximo paso evidente es permitir que el bucle local lleve las señales digitales que resultan en conexiones conmutadas de mayor capacidad.

La red digital de servicios integrados (ISDN) convierte el bucle local en una conexión digital TDM. La conexión utiliza canales portadores de 64 kbps (B) para transportar voz y datos, y una señal, canal delta (D), para la configuración de llamadas y otros propósitos.

La interfaz de acceso básico (BRI) ISDN está destinada al uso doméstico y a las pequeñas empresas y provee dos canales B de 64 kbps y un canal D de 16 kbps. Para las instalaciones más grandes, está disponible la interfaz de acceso principal (PRI) ISDN. En América del Norte, PRI ofrece veintitrés canales B de 64 kbps y un canal D de 64 kbps, para un total de velocidad de transmisión de hasta 1,544 Mbps. Esto incluye algo de carga adicional para la sincronización. En Europa, Australia, y otras partes del mundo, PRI ISDN ofrece treinta canales B y un canal D para un total de

velocidad de transmisión de hasta 2,048 Mbps, incluyendo la carga de sincronización. En América del Norte, PRI corresponde a una conexión T1. La velocidad de PRI internacional corresponde a una conexión E1.



Los routers pueden tener interfaces ISDN o conectarse a un adaptador de terminal.

Figura 15. WAN con ISDN

Fuente: Cisco System

Cuando se requieren conexiones dedicadas permanentes, se utilizan líneas alquiladas con capacidades de hasta 2.5 Gbps.

Un enlace punto a punto ofrece rutas de comunicación WAN preestablecidas desde las instalaciones del cliente a través de la red hasta un destino remoto. Las líneas punto a punto se alquilan por lo general a una operadora de servicios de telecomunicaciones y se denominan líneas alquiladas. Se pueden conseguir líneas alquiladas con distintas capacidades. Estos circuitos dedicados se cotizan, en general, según el ancho de banda necesario y la distancia entre los dos puntos conectados.

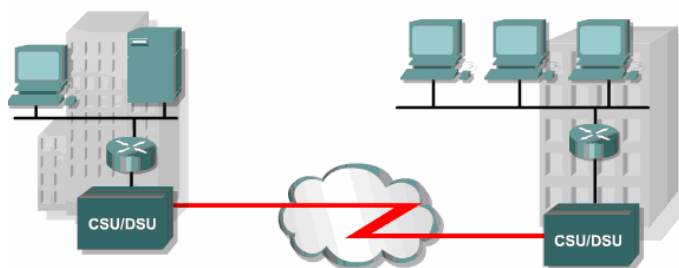


Figura 16. WAN con una línea Alquilada

Fuente: Cisco System

Regresión Lineal

Según Santos (2002), el análisis de regresión tiene por objetivo estimar el valor promedio de una variable, variable dependiente, con base en los valores de una o más variables adicionales, variables explicativas. En este tipo de análisis, la variable dependiente es estocástica mientras que las variables explicativas son no estocásticas en su mayor parte. El análisis de regresión ha cobrado popularidad debido al gran número de paquetes estadísticos que lo incluyen y por ser un “proceso robusto que se adapta a un sinfín de aplicaciones científicas y ejecutivas que permite la toma de decisiones”.

El análisis de regresión trata con la descripción y evaluación de las relaciones entre una variable determinada (llamada dependiente o explicada o endógena) y una o más variables adicionales (llamadas independientes, explicativas o exógenas).

Cuando se tiene solo una variable exógena se habla del modelo de regresión lineal simple. Si se tiene más de una variable exógena se habla del modelo de regresión lineal múltiple

Según Arsham (2002), el análisis de regresión tiene tres objetivos: predecir, modelar, y la caracterización. ¿Cuál debería ser el orden lógico en el cual se aborden estos tres objetivos de forma tal que uno de ellos guíe y/ o justifique los otros objetivos?. Obviamente, esto dependerá de cual es el objetivo principal. Algunas veces se necesita modelar con el objetivo de realizar mejores predicciones. Por lo tanto, el orden lógico es obvio. Algunas veces simplemente se necesita explicar los hechos, por lo tanto el modelar sería la clave, a pesar de la muestra, la predicción podría ser utilizada para probar el modelo. Con frecuencia, modelar y predecir utilizan procesos iterativos de los cuales no existe ningún “orden lógico” en el sentido más amplio. Se podría modelar para obtener predicciones, lo cual posibilita mayor control, sin embargo, las iteraciones son fáciles de aparecer, y existen algunas aproximaciones para controlar los problemas.

Muchos problemas surgen cuando se describe cómo las variables están relacionadas. El más simple de todos los modelos que describe la relación entre dos

variables es un modelo lineal, o de línea recta. La regresión lineal es siempre lineal en los coeficientes que son estimados, y no necesariamente lineal en las variables.

El método más simple de dibujar un modelo lineal es “calcular visualmente” una línea a través de los datos sobre un diagrama, pero un modelo más elegante sería el método convencional de mínimos cuadrados, el cual encuentra la línea al reducir al mínimo la suma de las distancias verticales entre los puntos observados y la línea ajustada. Es importante recalcar, que ajustando la “mejor” línea de acuerdo a la vista es muy difícil, especialmente cuando hay mucha variabilidad residual en los datos.

Según Walpole y Myers (1992), en la práctica, con mucha frecuencia es necesario resolver problemas que implican conjuntos de variables, cuando se sabe que existe alguna relación inherente entre ellas. Puede ser interesante desarrollar un método de predicción. El aspecto estadístico del problema consiste entonces en lograr la mejor estimación de la relación entre las variables.

Para la mayoría de las aplicaciones, existe una clara distinción entre las variables en cuanto a su papel dentro del proceso experimental. Muy a menudo se tiene una sola variable dependiente o respuesta Y , la cual no se controla en el experimento. Esta respuesta depende de una o más variables independientes o de regresión, como son x_1, x_2, \dots, x_k , las cuales se miden con un error despreciable y en realidad, en la generalidad de los casos se controlan en el experimento. Así, las variables independientes no son aleatorias y, por lo tanto, no tienen propiedades distribucionales.

Según Cole (2004), el objeto de un análisis de regresión es investigar la relación estadística que existe entre una variable dependiente (Y) y una o más variables independientes ($x_1, x_2, x_3 \dots$). Para poder realizar esta investigación, se debe postular una relación funcional entre las variables. Debido a su simplicidad analítica, la forma funcional que más se utiliza en la práctica es la relación lineal. Cuando solo existe una variable independiente, esto se reduce a una línea recta:

$$\hat{Y} = b_0 + b_1 X$$

donde los coeficientes b_0 y b_1 son parámetros que definen la posición e inclinación de la recta. El símbolo especial \hat{Y} es utilizado para representar el valor de Y calculado por la recta. El valor real de Y rara vez coincide exactamente con el valor calculado, por lo que es importante hacer esta distinción.

El parámetro b_0 , conocido como la “ordenada en el origen,” indica cuánto es Y cuando $X = 0$. El parámetro b_1 , conocido como la “pendiente,” indica cuánto aumenta Y por cada aumento de una unidad en X . El problema consiste en obtener estimaciones de estos coeficientes a partir de una muestra de observaciones sobre las variables Y y X . En el análisis de regresión, estas estimaciones se obtienen por medio del método de mínimos cuadrados.

Estimación de la Recta de Regresión y del Coeficiente de Determinación

Para estimar los coeficientes por medio de mínimos cuadrados, se utilizan las siguientes fórmulas:

$$b_1 = \frac{\sum XY - \bar{y} \sum X}{\sum X^2 - \bar{x} \sum X}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Coeficiente de Determinación (R^2)

Según Arsham (2002), otra medida de la cercanía de los puntos a la línea de regresión es el Coeficiente de Determinación, el cuál es la cantidad de la desviación al cuadrado en Y , la cual es explicada por los puntos en la menor línea de regresión de los cuadrados.

Una pregunta importante que se plantea en el análisis de regresión es la siguiente: ¿Qué porcentaje de la variación total en Y se debe a la variación en X ? En otras palabras, ¿cuál es la proporción de la variación total en Y que puede ser “explicada”

por la variación en X? El estadístico que mide esta proporción o porcentaje se denomina coeficiente de determinación:

$$R^2 = 1 - \frac{\sum e^2}{\sum (Y - \bar{y})^2}$$

Regresión Múltiple

Según Arsham (2002), los objetivos en un problema de regresión múltiple son esencialmente iguales que para una regresión simple. Mientras que los objetivos siguen siendo iguales, mientras más predictores tenemos, los cálculos y las interpretaciones son más complicadas. Con la regresión múltiple, se puede utilizar más de un predictor. Esto siempre es mejor, sin embargo, ser parsimonioso, es decir, utilizar tan pocas variables como predictores sean necesarios para conseguir un pronóstico razonablemente exacto.

En el caso más general de la regresión múltiple, existen dos o más variables independientes:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots$$

La estimación de los coeficientes de una regresión múltiple es un cálculo bastante complicado y laborioso, por lo que se requiere del empleo de programas de computación especializados. Sin embargo, la interpretación de los coeficientes es similar al caso de la regresión simple: el coeficiente de cada variable independiente mide el efecto separado que esta variable tiene sobre la variable dependiente. El coeficiente de determinación, por otro lado, mide el porcentaje de la variación total en Y que es explicado por la variación *conjunta* de las variables independientes.

Según Rodríguez y Escanciano (2001), se necesita una hipótesis adicional, que no exista multicolinealidad, es decir, ninguna de las variables explicativas X_i es

combinación lineal de las restantes. Cuando existe multicolinealidad, no sólo se puede simplificar el modelo, sino que además, en caso de no simplificarlo, no se podría utilizar para obtener estimaciones. Lo que se busca es el plano $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots$ que mejor se adapte a los puntos que se obtienen. En numerosas ocasiones se presenta una multicolinealidad aproximada, lo cual también presenta problemas

Requisitos y Limitaciones

Según Barón y Téllez (2000), hay ciertos requerimientos para poder utilizar la técnica de regresión lineal múltiple:

Linealidad: se supone que las variables repuestas depende linealmente de las variables explicativas. Si la respuesta no aparenta ser lineal, se debe introducir en el modelo componentes no lineales (como incluir transformaciones no lineales de las variables independientes en el modelo). Otro tipo de respuesta no lineal es la interacción. Para ello se ha de incluir en el modelo términos de interacción, que equivalen a introducir nuevas variables explicativas que en realidad son el producto de dos o más de las independientes.

Normalidad y equidistribución de los residuos: se llaman residuos las diferencias entre los valores calculados por el modelo y los realmente observados en la variable dependiente. Para tener un buen modelo de regresión no es suficiente con que los residuos sean pequeños. La validez del modelo requiere que los mismos se distribuyan de modo normal y con la misma dispersión para cada combinación de valores de las variables independientes.

Número de variables independientes: Se puede estar tentados en incluir en el modelo cualquier cosa que tengamos en una base de datos, con la esperanza de cuantas variables incluyamos, más posibilidades hay de que funcione el modelo. Si ocurre esto, se está en el riesgo de cometer error de tipo I. Otra razón, es que si se espera ajustar unas pocas observaciones usando muchas variables, muy probablemente se consiga una aproximación muy artificial, y además muy sensible a los valores observados. La inclusión de una nueva observación puede cambiar

completamente el valor de los coeficientes del modelo. Esto se traduce que al realizar el contraste ocurre todo lo contrario a lo que se desea del modelo: todas las variables independientes del modelo serán consideradas no significativas. Una regla que se suele recomendar es la de incluir al menos 20 observaciones por cada variable independiente que se estima a priori, números inferiores nos lleva probablemente a no poder obtener conclusiones y errores de tipo II.

Colinealidad: Si dos variables independientes están estrechamente relacionadas y ambas son incluidas en el modelo, muy posiblemente ninguna de las dos sea considerada significativa, aunque si se incluye una solo de ellas, si. Hay diferentes técnicas para detectar la colinealidad pero que requieren profundizar, una de las técnicas muy simple es examinar los coeficientes del modelo para ver si se vuelven inestables al introducir una nueva variable; si es así posiblemente hay colinealidad entre la nueva variable y las anteriores.

La colinealidad, es uno de los problemas más desesperantes se puede encontrar en un análisis de regresión. Anteriormente se hizo referencia a la estimación de los coeficientes, si en un modelo de Regresión Lineal Múltiple, alguna variable independiente es combinación lineal de otras, el modelo es irresoluble, debido a que, en ese caso, la matriz $X'X$ es singular, es decir, su determinante es cero y no se puede invertir.

A este fenómeno se le denomina colinealidad. Que una variable X_1 sea combinación lineal de otra X_2 , significa que ambas están relacionadas por la expresión $X_1 = b_1 + b_2X_2$, siendo b_1 y b_2 constantes, por lo tanto el coeficiente de correlación entre ambas variables será 1.

Del mismo modo, que una variable X_1 sea combinación lineal de otras X_2, \dots, X_i con $i > 2$, significa que dichas variables están relacionadas por la expresión $X_1 = b_1 + b_2 X_2 + \dots + b_i X_i$, siendo b_1, \dots, b_i constantes y por tanto, el coeficiente de correlación múltiple $R_{X_1|X_2, \dots, X_i}$ también será 1.

Otro modo, por tanto, de definir la colinealidad es decir que existe colinealidad cuando alguno de los coeficientes de correlación simple o múltiple entre algunas de las variables independientes es 1, es decir, cuando algunas variables independientes

están correlacionadas entre sí.

En la práctica, esta colinealidad exacta raras veces ocurre, pero sí surge con cierta frecuencia la llamada casi-colinealidad, o por extensión, simplemente colinealidad en que alguna variable es "casi" combinación lineal de otra u otras, o dicho de otro modo, algunos coeficientes de correlación simple o múltiple entre las variables independientes están cercanos a 1, aunque no llegan a dicho valor.

En este caso la matriz $X'X$ es casi-singular, es decir su determinante no es cero pero es muy pequeño. Como para invertir una matriz hay que dividir por su determinante, en esta situación surgen problemas de precisión en la estimación de los coeficientes, ya que los algoritmos de inversión de matrices pierden precisión al tener que dividir por un número muy pequeño, siendo además inestables.

Por consiguiente, a la hora de plantear modelos de Regresión Lineal Múltiple, conviene estudiar previamente la existencia de casi-colinealidad (la colinealidad exacta no es necesario estudiarla previamente, ya que todos los algoritmos la detectan, de hecho no pueden acabar la estimación). Como medida de la misma hay varios estadísticos propuestos, los más sencillos son los coeficientes de determinación de cada variable independiente con todas las demás

Dentro los síntomas que nos pueden alertar del problema de la existencia de la colinealidad son:

- 1) Alta correlación entre dos variables.
- 2) Si las pruebas de nulidad de los coeficientes de regresión nos llevan a eliminar variables relevantes.
- 3) Signo de un coeficientes opuesto al esperado.
- 4) Varianza de los estimadores anormalmente grande.
- 5) Encontrar un coeficiente de regresión múltiple entre la regresora y las demás muy altos.
- 6) Intervalos de confianza grandes para los regresores que representan a variables importantes.
- 7) Si al eliminar una variable, las varianzas anormalmente grandes de los regresores, disminuyen drásticamente.

Estos síntomas son meramente orientativos, siendo muy recomendable hacer el diagnóstico al efectuar una regresión múltiple.

Regresión No Lineal

La regresión lineal no siempre da buenos resultados, porque a veces la relación entre Y y X no es lineal sino que exhibe algún grado de curvatura. La estimación directa de los parámetros de funciones no-lineales es un proceso bastante complicado. No obstante, a veces se pueden aplicar las técnicas de regresión lineal por medio de transformaciones de las variables originales.

Una función no-lineal que tiene muchas aplicaciones es la *función exponencial*:

$$Y = AX^b$$

donde A y b son constantes desconocidas. Si aplicamos logaritmos, esta función también puede ser expresada como:

$$\text{Log}(Y) = \log(A) + b.\log(X)$$

Analizadores de Protocolos

Según García (2002), un analizador de protocolos es una herramienta tradicional en el desarrollo y depuración de protocolos y aplicaciones de red. Un analizador de protocolos es un programa que permite al ordenador capturar tramas de la red para, posteriormente o en tiempo real, proceder a su análisis. Por analizar se entiende que el programa es capaz de reconocer que la trama capturada transporta información asociada a un protocolo concreto (por ejemplo a TCP, a ICMP, entre otros) y muestra al usuario la información convenientemente decodificada. De esta manera el usuario puede, de forma cómoda, ver qué es lo que está circulando por la red.

Según Cheng (2005), muchas veces dentro de una red ocurren problemas bien sea tráfico, ataques de virus, gusanos. Estos problemas son difíciles de detectar sin la ayuda de herramientas que permitan ver el tráfico que está pasando por la red. Muchas de estas herramientas son software que se instalan sobre una computadora normal, hay otros que son equipos especializados que permiten ver mucho más. Para poder usar estas herramientas instalados sobre una computadora, es necesario tener una tarjeta de red con capacidad de cambiarse al modo promiscuo (que escucha todo, se puede decir que todas las soporta). En el caso de una red local (LAN) alamburada, el equipo debe estar conectado hacia un concentrador o hub (y los demás equipos también), o sino, en el caso de un conmutador o switch, debe estar conectado en el puerto que tenga capacidad de PORT MIRRORING (no todos los switches los trae).

En el mercado existen diferentes analizadores de la red, pero uno de los más importantes es *Sniffer* pues nos permite analizar la red en las diferentes capas de la misma, auxiliándose de un hardware especial con el cual es posible ver la información a un nivel más bajo. Esto sólo un experto en el área es capaz de analizarlo. Dentro de las muchas herramientas que contiene *Sniffer*, cabe resaltar a un sistema experto, el cual nos muestra algunos síntomas de la red, es decir una vez que se analizó la red, el producto es capaz de decirnos algunos problemas de red, por ejemplo un tiempo de respuesta muy alto.

Para observar y analizar el comportamiento de los protocolos de red es preciso disponer de una herramienta capaz de monitorear el tráfico en la red y mostrarlo en una forma legible. Las herramientas que capturan y muestran el tráfico existente en una interfaz de red se denominan *analizadores de protocolos de red*, *analizadores de paquetes*, "*packet sniffers*" o simplemente "*sniffers*" (del inglés sniff, olfatear).

Para visualizar el tráfico los analizadores de protocolo se coloca la tarjeta de red en *modo promiscuo*, una modalidad en la cual es capturado todo el tráfico visible para la tarjeta de red. En una red Ethernet una interfaz de red en modo promiscuo puede ver todo el tráfico generado por todos los equipos que comparten el mismo conjunto de cables y concentradores (hubs). El modo promiscuo implica riesgos evidentes de seguridad, por lo que su uso suele limitarse al supervisor.

Ethereal: es un analizador de protocolos con interfaz gráfica capaz de reconocer muchos protocolos distintos. Permite tanto revisar los paquetes de datos en una red activa como desde un archivo de captura previamente generado; es capaz de comprender diversos formatos de archivo propios de otros programas de captura, en particular el clásico tcpdump.

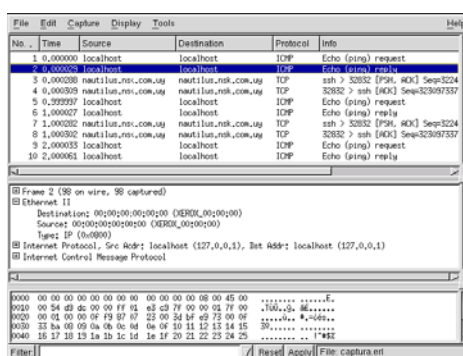


Figura 17. Ventana principal de Ethereal luego de una captura.

Fuente: <http://iie.fing.edu.uy/ense/assign/redatos>

Iris Network Traffic Analyzer: Este es un programa que realiza funciones más avanzadas que Ethereal, tales como creación de archivos de filtros, estadísticas avanzadas, reconstrucción de sesiones, captura de paquetes, entre otras. Se puede adquirir en la página del fabricante

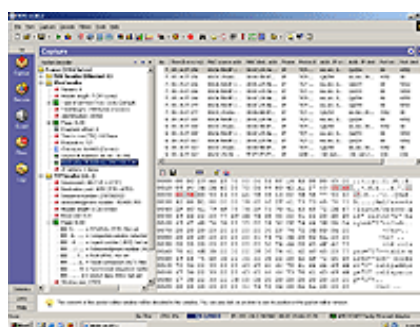


Figura 18. Iris Analyzer

Fuente: <http://ruben.cheng-ca.com/es/knowledge/network/trafficanalysis.htm>

Etherpeek: Es un programa comercial permite llevar un gráfico mucho más avanzado, filtros avanzados, mediciones del uso de la red, estadísticas, gráfico de conexiones y mucho más.

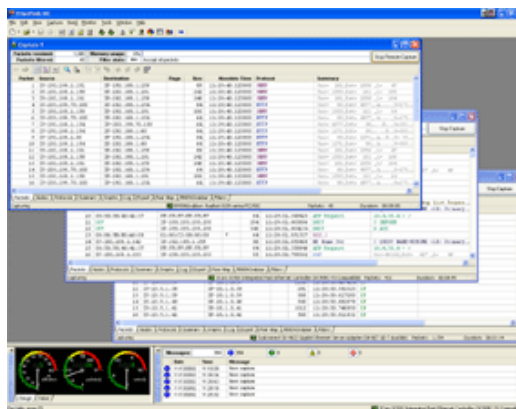


Figura 19. Etherpeek

Fuente: <http://ruben.cheng-ca.com/es/knowledge/network/trafficanalysis.htm>

Network Top (NTOPT): Ntop es una sonda de red que muestra el uso de la red discriminando protocolos, puertos y aplicaciones. Está basada en la librería de captura de paquetes “pcap” y bajo sistemas UNIX se le conoce como TCPDump.

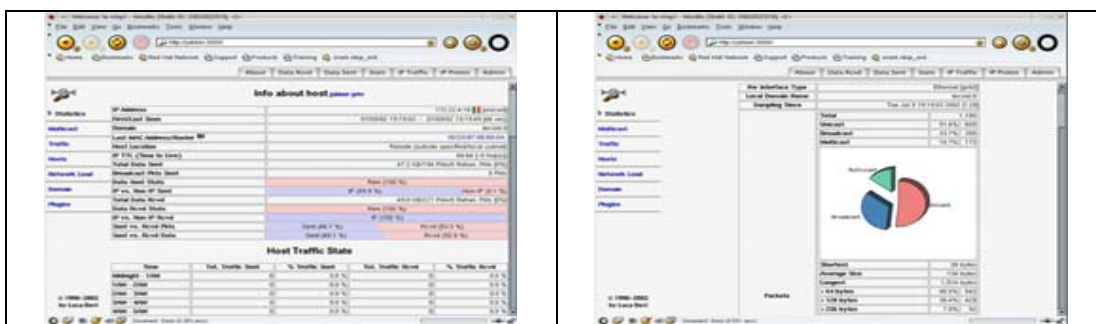


Figura 20. NTOPT

Fuente: www.ntop.org

Monitores de Tráfico

Las redes de cómputo de las organizaciones, se vuelven cada vez más complejas y la exigencia de la operación es cada vez más demandante. Las redes, cada vez mas, soportan aplicaciones y servicios estratégicos de las organizaciones. Por lo cual el análisis y monitoreo de redes se ha convertido en una labor cada vez mas importante y de carácter pro-activo para evitar problemas.

Según Cheng (2005), dentro de una red, sea grande o pequeña, es importante llevar un monitor de tráfico, esta permitiría ayudar a detectar fallas como también medir el uso de la red para poder hacer previsiones y tomar medidas. Nos permiten visualizar los componentes de nuestra red, saber si los equipos de comunicaciones se encuentran activos.

Según Maraboli (1997), un monitor es un instrumento que entrega datos de algún tipo (numérico, visual, auditivo) de un proceso o fenómeno. De aquí que el término puede ser utilizado para monitores de computador o de vigilancia, hasta monitores de incendio y de intrusos (alarmas). El último caso, además de ser un monitor normal, entrega un dato específico cuando alguna variable ha sobrepasado algún umbral o situación prefijada.

Un monitor de tráfico en redes de computadoras es un instrumento que entrega datos acerca de la red en la cual está conectada. Estos datos pueden ser entregados en diversas formas, dependiendo del fin con el cual el monitor fue diseñado.

A continuación se presenta una clasificación de los diversos tipos de monitores de red que existen en el mercado como en Internet. Cabe destacar que la clasificación no es estricta, por lo que un monitor puede cumplir múltiples características. Los monitores de red se pueden clasificar según los criterios de objetivo, reporte, intrusividad, operación y protocolos que miden

Objetivo

1) **Monitor de Estados (variables):** El objetivo es el de “avisar” situaciones de emergencias como “caídas” de equipos o el sobrepaso de un umbral de alguna variable fijada por el administrador. Por ejemplo: la NO-Respuesta de un Router o la superación del umbral de 40% en el porcentaje de colisiones. Para ello se ocupa

generalmente el estándar SNMP que está orientado a monitorear y configurar el equipamiento físico de una red (bridges, routers, hubs y Workstations).

Dentro de los Monitores SNMP Comerciales se destacan HP OpenView, SUNNET Manager y Cisco Works. Cabe mencionar que el SNMP no sólo sirve para monitorear y/o configurar variables de los equipamientos de la red física sino que mostrar el tráfico histórico en forma gráfica acerca de lo se ha "traficado" por los equipos de la red, como routers, switches, etc.

2) **Monitor de Tráfico:** el objetivo es el de registrar el tráfico de las redes, ya sea con fines estadísticos o de detección de congestión. Este tipo de monitor también puede tener "alarmas" y por ende, enviar una señal al administrador cuando una variable monitoreada haya excedido de un umbral prefijado (considerado como alarmante).

Reporte

1) **Histórico:** este tipo de monitor entrega informes o resúmenes de la actividad histórica de la red, ya sea por: hora, día, semana o mes. La actividad mencionada puede corresponder a tráfico, alarmas ocurridas, etc. El reporte generalmente consiste en archivos de tipo texto, dado que éste puede ser ingresado a otro programa para transformarlo a gráficos. Los últimos monitores de este tipo entregan sus reportes en formato HTML (WWW) y generan automáticamente los gráficos usando programas propios o del sistema. Un monitor histórico debe entregar la mayor cantidad de información posible, ya sea en forma de texto o de gráfico, ya que la información entregada puede ser ingresada a una base de datos.

2) **Tiempo Real:** Un monitor de este tipo entrega datos de muy reciente ocurrencia, desde 1 segundo hasta 10 minutos, con el objeto de detectar y corregir los problemas cuanto antes. En algunos casos, como en las "alarmas" de umbrales, las variables son de tiempo corto y por lo tanto son inmediatamente avisadas. En otros casos, si la variable tiene un tiempo de "ejecución", como la transmisión de un paquete, se debe esperar la transmisión completa antes de poder tomar acciones. Por estas razones, un monitor de tiempo real no sólo abarca a los monitores

"instantáneos", sino también a los que deben monitorear variables que requieren de un tiempo determinado de ejecución.

Intrusividad

1) **Intrusivo**: Es aquel monitor que interviene en el proceso o fenómeno a monitorear, vale decir, actúa como agente activo. En este caso, el monitor afecta la medición, haciéndola no confiable o no representativa del proceso. Por ejemplo, un monitor de tráfico local se considera intrusivo si ocupa la red para hacer mediciones. Sin embargo, si un monitor ocupa la red para obtener datos, no significa que su medición sea intrusiva, ya que esto depende del proceso que desee medir. Por ejemplo: si el monitor debe medir el tráfico que es cursado por un router y consulta a éste último por esa información (usando la red local), su medición no es intrusiva.

2) **No Intrusivo**: Un monitor es no intrusivo si no interviene el proceso o fenómeno a monitorear, vale decir, si actúa como agente pasivo. Por ejemplo: un monitor de tráfico local no es intrusivo si no utiliza la red para medir su tráfico local. Si el monitor es un monitor del estado de un proceso en UNIX, el monitor en cuestión no debe interactuar con tal proceso, porque afectaría su ejecución normal.

3) **Operación**: Esta clasificación se refiere a la operación del monitor, en cuanto a dónde despliega sus datos y dónde es configurado por un administrador

4) **Local**: si la operación del monitor es local, entonces significa que muestra los datos en el mismo lugar de donde los obtiene. En el caso de una alarma de una casa, la alarma es configurada por el dueño en la casa misma y la alarma suena en la misma casa en caso de la entrada de un intruso. En el caso de un monitor de tráfico, esto significa que los datos se despliegan en el mismo computador que obtiene los datos.

5) **Remota**: la operación de este tipo de monitores se realiza en forma remota. En el caso de la alarma de casa, la operación se realiza desde una empresa dedicada. En caso de la entrada de un intruso, la alarma avisa en las dependencias de la Empresa y no en la casa en cuestión. En el caso de un monitor de tráfico, esto significa que los datos son desplegados en otro computador distinto al que obtiene los datos. Lo anterior requiere ocupar la red para comunicar estos 2 computadores, por lo

que dependiendo del tipo de medición, será también clasificado como intrusivo o no intrusivo.

Protocolos: Esta clasificación permite destacar los tipos de protocolos de redes con que el monitor puede trabajar.

La clasificación contempla mencionar la compatibilidad con los siguientes protocolos de red: Ethernet en 4 versiones (Versión II, Novell RAW, IEEE 802.3 y SNAP.TCP/IP), estándar para aplicaciones Internet ó Intranet IPX, estándar para redes NovellNETbeui (NETbios), estándar para redes Microsoft.

A continuación se presentan algunas herramientas de monitoreo de tráfico que están actualmente en el mercado.

MRTG - Multi Router Traffic Grapher: este es un programa de Código Abierto, que se puede implementar sobre Sistemas Operativos UNIX (más común) y Windows para realizar graficaciones del tráfico con un tiempo de muestreo de mínimo 5 minutos en el modo normal usando diferentes fuentes de datos, pero generalmente se utiliza el SNMP. Dado que cada ciclo de muestreo, se genera un gráfico de cada monitor, esto trae como desventajas una carga al sistema que en la práctica hace limitada su implementación a gran escala, y además dado que su gráfico es un muestreo de 5 minutos, es difícil detectar alguna falla durante los lapsos entre cada muestreo.

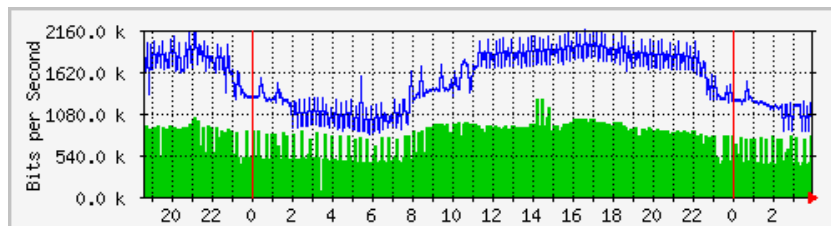


Figura 21. MRTG

Fuente: <http://ruben.cheng-ca.com/es/knowledge/network/trafficmon.htm>

RRD Tools - Round Robin Database Tools: Este como tal no es un programa de monitoreo de tráfico, sino más bien un programa de Código Abierto, que en conjunto con el programa MRTG, permite tener mayor capacidad de monitoreo al descargar la

tarea de generar los gráficos cada 5 minutos y además almacenar los muestreo dentro de una base de datos (exportable a XML) lo cual lo hace más eficiente, y permite que el muestreo se puede llevar a cabo con un mínimo de 1 minuto. Adicionalmente trae los módulos necesarios para tomar esa información y generar los gráficos. Generalmente se utiliza un programa interfaz en conjunto con el RRD Tools para generar los gráficos de los cuales hay varios.

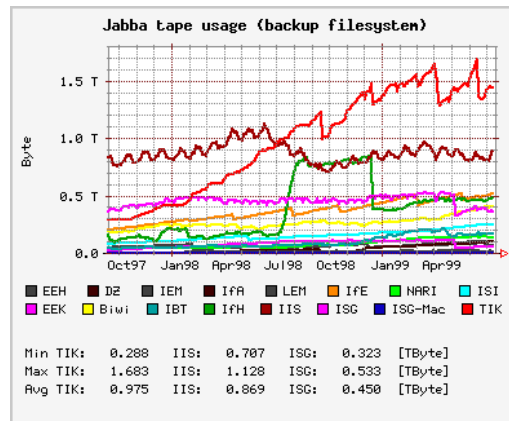


Figura 22. RRD Tools

Fuente: <http://ruben.cheng-ca.com/es/knowledge/network/trafficmon.htm>

Routers2.cgi: Este es un programa de Código Abierto, que funciona como un CGI usando PERL y que utiliza los datos generados del RRD Tools y la configuración de MRTG para generar los gráficos de monitoreo cuando se necesite visualizarlo y de manera organizada y personalizada.

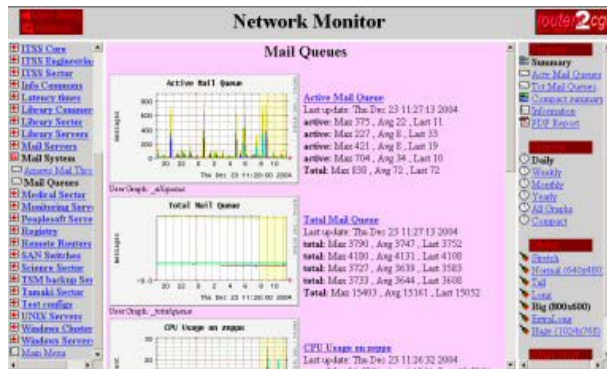


Figura 23. Routers2.cgi

Fuente: <http://ruben.cheng-ca.com/es/knowledge/network/trafficmon.htm>

STG - SNMP Traffic Grapher: es programa gratuito que corre desde Windows 98, y fue desarrollado como un complemento al MRTG. Realiza gráficos de dos colores (generalmente interpretado como entrada/salida) a través de contadores en agentes SNMP en tiempo real con un mínimo de tiempo de muestreo desde 100 milisegundos, aunque en la práctica, es útil con un muestreo entre 500 ms a 5000 milisegundos (5 segundo), además de poder guardar las muestras en un archivo de CSV y procesarlo con otro programa, como también permite guardar la configuración del monitor como un archivo y poder utilizarlo después. Este es un programa básico que debe tener siempre ya que ayuda a detectar fallas rápidamente sin esperar por los gráficos de MRTG.

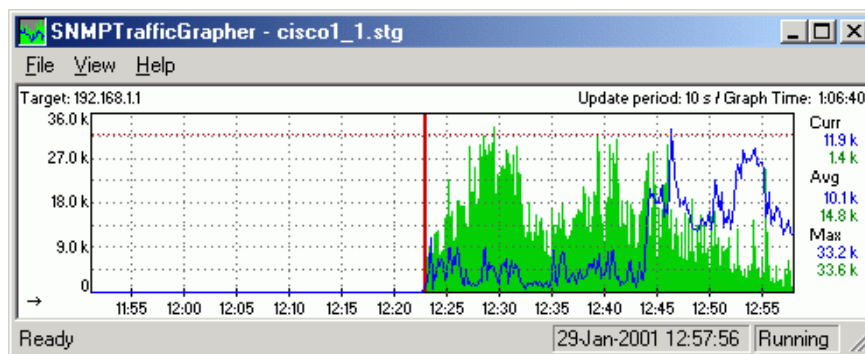


Figura 24. STG

Fuente: <http://ruben.cheng-ca.com/es/knowledge/network/trafficmon.htm>

PRTG Traffic Grapher: Es programa comercial (con una versión gratuita limitada) que sirve como complemento al STG pero con mayores facilidades y opciones tales como un servidor web para mostrar los gráficos a través de un navegador, gráficos históricos, etc.

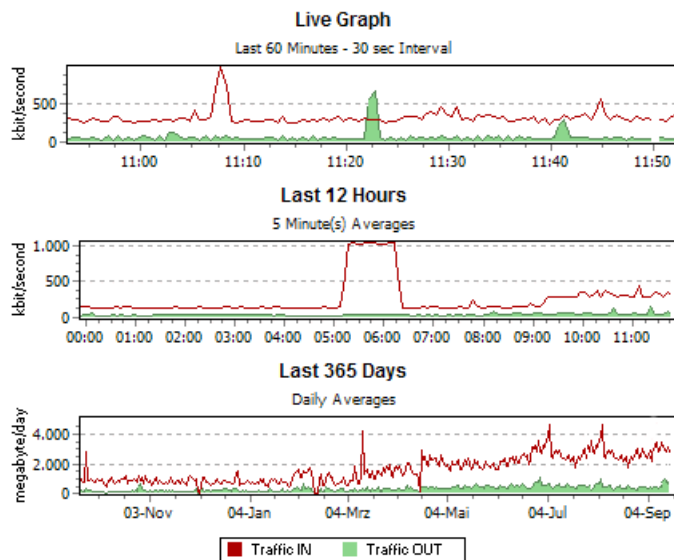


Figura 25. PRTG Traffic Grapher

Fuente: <http://www.paessler.com/prtg>

Sistema de Variables

Las variables de esta investigación fueron clasificadas en: Independientes y Dependientes dado que se trabajó con un modelo de tipo causal. Como se muestra en la Tabla 1, las variables independientes se derivan del Protocolo TCP/IP en la capa de Aplicación, definida como los protocolos destinados a prestar servicios. En este sentido las variables independientes están conformadas por los siguientes protocolos en sus dimensiones: DNS, POP3, SMTP, SNMP, FTP, TFTP, HTTP. Operacionalmente se definen estas variables por su correspondiente medida en kbps.

La Variable Dependiente se definió nominalmente como “Tráfico”. Según Cisco System (2002), corresponde a la cantidad de información que puede fluir desde un lugar hacia otro en un período de tiempo determinado. Operacionalmente fue medida considerando las dimensiones, ancho de banda y el indicador representado por paquetes o kbps

Tabla 1. Operacionalización de las Variables

Variabes	Dimensiones	Indicadores	Items
Protocolos	DNS	Sistema de Nombre de Dominio	7
	FTP	Transferencia de Archivos	
	HTTP	Transferencia de Hipertexto	
	POP3	Correo Entrante	
	SMTP	Correo Saliente	
	SNMP	Administración de Red	
	TFTP	Transferencia Trivial de Archivos	
Tráfico	Ancho de Banda	Número de kbps	1,2,3,4,5,6,7

En el modelo de regresión que se usó para la predicción de tráfico los protocolos mencionados en la Tabla 1, actuaron como las variables independientes y tendrán causa efecto en la variable dependiente “Tráfico”.

CAPÍTULO III

MARCO METODOLÓGICO

Tipo de Investigación

Este estudio se refiere al desarrollo de una herramienta automatizada que permita modelar en tiempo real el consumo de ancho de banda del enlace WAN de una red. Está basado en una investigación de campo con carácter descriptivo, siguiendo la modalidad de un proyecto especial. Según el manual de la Universidad Centroccidental Lisandro Alvarado (2002), se entiende por Investigación de Campo “la aplicación del método científico en el tratamiento de un sistema de variables y sus relaciones, las cuales conducen a conclusiones y al enriquecimiento de un campo del conocimiento o disciplina inherente a la especialidad, con la sustentación de los experimentos y observaciones realizadas” (p.4).

En relación al tipo descriptivo, según Grajales (2000), “los estudios descriptivos buscan desarrollar una imagen o fiel representación (descripción) del fenómeno estudiado a partir de sus características. Describir en este caso es sinónimo de medir. Miden variables o conceptos con el fin de especificar las propiedades importantes de comunidades, personas, grupos o fenómeno bajo análisis. El énfasis está en el estudio independiente de cada característica, es posible que de alguna manera se integren la mediciones de dos o más características con el fin de determinar cómo es o cómo se manifiesta el fenómeno. Pero en ningún momento se pretende establecer la forma de relación entre estas características. En algunos casos los resultados pueden ser usados para predecir”. De esto se puede notar la investigación a realizar entra dentro de este tipo, debido a que se generará un modelo en tiempo real que informe el consumo de ancho de banda de una red.

El Manual de presentación de trabajos de grado y tesis doctoral de la Universidad Pedagógica Experimental Libertador (2002), define como proyecto especial a:

trabajos que llevan a creaciones tangibles, susceptibles de ser utilizadas como soluciones a problemas demostrados, o que respondan a necesidades e intereses de tipo cultural. Se incluyen en esta categoría los trabajos de elaboración de libros de texto y de materiales de apoyo educativo, el desarrollo de software, prototipos y de productos tecnológicos en general, así como también los de creación literaria y artística (p.8).

De esta forma, se propone una herramienta automatizada haciendo uso de la técnica de regresión lineal múltiple, para generar un modelo en tiempo real del consumo de ancho de banda del enlace WAN de una red.

Población y Muestra

Para Sabino (2000), la población es “el número relativamente alto de unidades que coinciden con ciertas especificaciones y que serán prácticamente imposible, por razones de tiempo y de costos, y porque no es en realidad imprescindible, examinar cada una de las unidades que lo compone” (p.122).

Con base a la población tomada por Guedez (2005) en su investigación se realizará esta investigación, la población consistió en la cantidad de paquetes que circulan en el enlace ADSL de la UFT por un período de cinco (5) días hábiles, lo cual refleja el comportamiento de la red en un período normal de funcionamiento de las aplicaciones que usan los servicios de la red.

Muestra

Según Sabino (2000), la muestra es “una parte del todo que llamamos universo y que sirve para representarlo, es un conjunto de unidades, una porción del total, que representa la conducta del universo en su conjunto” (p.122).

Con base a la muestra tomada por Guedez (2005) en su trabajo de investigación, permitió el desarrollo del presente estudio; la muestra que se tomó representa la cantidad de paquetes que circulan en el enlace WAN de la organización durante cinco (5) días hábiles, comprendidos desde el Lunes 02 de Mayo de 2005 hasta el viernes

06 de Mayo de 2005 en un horario comprendido de 8:00 a.m a 12 m y de 3:00 a.m a 7:00 p.m.

Por tal motivo, en este estudio se asumió que la muestra tomada como referencia de la investigación de Guedez (2005), es la totalidad de los paquetes definidos en la población y esta es representativa para la investigación, debido a que consiste en 40 observaciones siendo suficiente para una análisis de regresión lineal múltiple.

Fases del Estudio

A continuación se describe las distintas fases para el desarrollo de la presente investigación:

Fase 1. Diagnóstica

La finalidad de esta fase consistió en la revisión de la investigación de Guedez (2005), para así determinar la debilidad que presentó su modelo de predicción de tráfico para el acceso a redes de banda ancha de la UFT, encontrando que para generar el modelo tuvo que realizar un trabajo largo, exhaustivo, propenso a errores, y que si el comportamiento de la red de la UFT hubiese cambiado en algún momento el investigador tendría que haber realizado nuevamente el estudio para generar el modelo de predicción, lo que lo hace poco flexible y engorroso por la rigidez del mismo. Por tal motivo, se está en la necesidad de desarrollar una herramienta automatizada que genere el modelo en tiempo real, lo que permite que el modelo se adapte a los cambios de comportamiento que pueda ocurrir en le red de cualquier organización.

Fase 2. Documentación Bibliográfica

Esta fase tuvo como finalidad la búsqueda de información en diferentes fuentes (bibliográfica, documentales, electrónicas, etc), para así desarrollar los antecedentes y las bases teóricas de la investigación, los cuales permiten fundamentar el estudio.

Fase 3. Elaboración y Evaluación de la herramienta

En esta fase se realizó la programación de la herramienta que validará que la información capturada y clasificada por protocolo de aplicación cumpla con un modelo de regresión lineal múltiple y posteriormente generar el modelo en tiempo real para el consumo de ancho de banda del enlace WAN de la red.

Una vez realizada la validación se procederá a chequear los resultados que genera la herramienta desarrollada. Consistió de varias pruebas con diferentes datos para verificar que la generación del modelo es flexible y se adapta a cualquier organización.

Fase 4. Elaboración de Conclusiones y Recomendaciones

Con base a la evaluación de la herramienta, en esta fase se presentaron las conclusiones, los aportes y las recomendaciones del trabajo para futuras investigaciones.

CAPÍTULO IV

Análisis de Resultados

Elaboración de la Herramienta

Esta fase de la implementación de la herramienta propuesta respondió al tercer (3er) objetivo específico de la investigación, el cual fue: Programar la herramienta para validar que la información capturada y clasificada por protocolo de aplicación cumple con un modelo de regresión lineal múltiple, y además generar el modelo en tiempo real.

La herramienta automatizada para el modelaje en tiempo real del consumo de ancho de banda del enlace WAN, cuenta con módulos para generar el modelo y graficar la relación de los diferentes protocolos a nivel de la capa de aplicación con respecto al ancho de banda (BW). Para la generación del modelo que representa el ancho de banda de la red, se utilizó la técnica de Análisis de Regresión Lineal Múltiple, en el que se relaciona la variable que se desea explicar BW, con las k variables explicativas (x_1, x_2, \dots, x_k) que representan los diferentes protocolos (DNS, FTP, HTTP, POP3, SMTP, SNMP, TFTP).

Para el desarrollo de la aplicación se usaron los siguientes insumos:

Hardware

- Un equipo Pentium III.
- Procesador 500 Mhz.
- Disco Duro de 8 G.B.
- 128 MB de RAM.
- Monito SVGA.

Software

- Sistema Operativo Windows XP

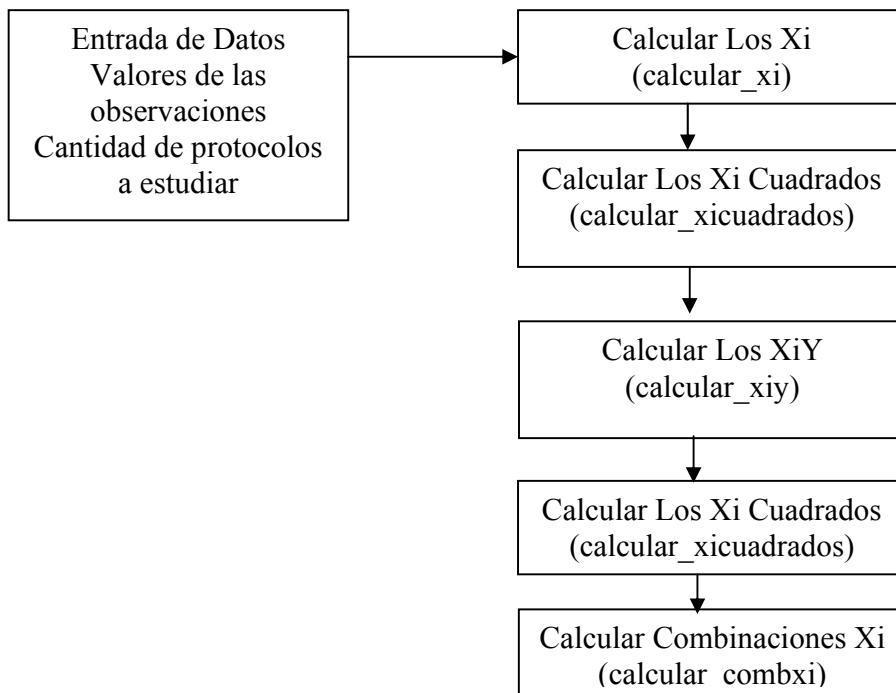
Lenguaje de programación utilizado

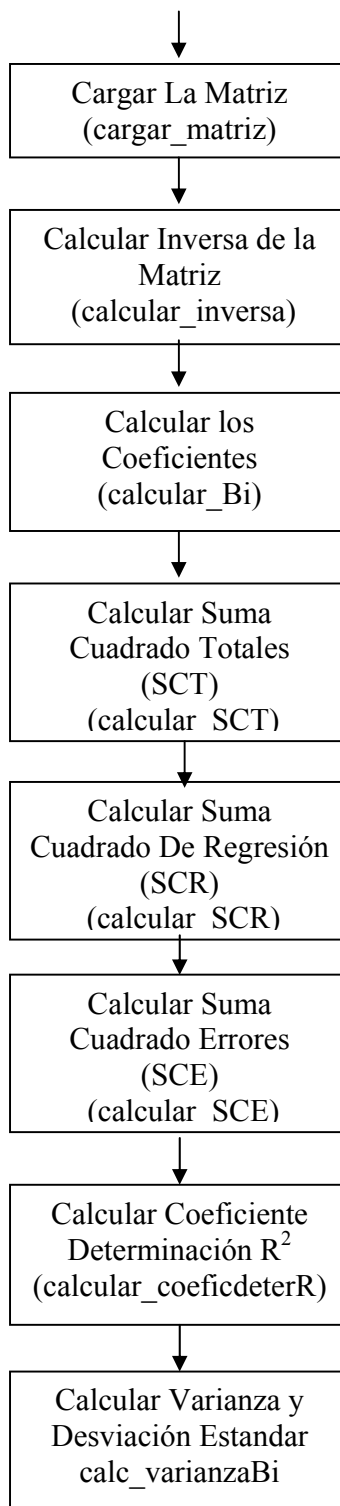
- Visual Basic 6.0

La herramienta desarrollada consta de una pantalla principal en donde tiene tres opciones de menú: 1) Una opción que permite al usuario entrar a la pantalla de configuración y allí establecer la cantidad y escoger los protocolos que desea estudiar, 2) Un opción de captura, que permite al usuario visualizar la pantalla de captura de los datos, que una vez cargados se genera el modelo en forma automática, 3) una opción que permite generar gráficos que representan la relación de BW con los protocolos estudiados.

Es de hacer notar que una vez capturado los datos se puede generar el modelo, siempre y cuando los datos disponibles (n-muestras) sean mayores que las $k+1$ variables (protocolos); es decir $n > k + 1$. Si la cantidad de observaciones cumplen con $n > k + 1$ se permite generar el modelo.

A continuación se presenta un esquema del desarrollo de esta aplicación para generar el modelo de Regresión Lineal Múltiple:





Para el desarrollo de los módulos Calcular los Xi (calcular_xi), calcular los Xi² (calcular_xicuadrados), calcular los XiY (calcular_xiy) y calcular las combinaciones de Xi (calcular_combxi), se efectuó la sumatoria de todos los Xi a partir de los datos de entrada, aplicando las siguientes formulas:

$$\sum_{i=1}^n y_i \quad \sum_{i=1}^n y_i^2 \quad (\text{Sumatoria de todos los } Y_i)$$

$$\sum_{i=1}^n x_{1i} y_i \dots \dots \dots \sum_{i=1}^n x_{ki} y_i \quad (\text{Sumatoria de los } x_i y_i)$$

$$\sum_{i=1}^n x_{1i} \quad \sum_{i=1}^n x_{2i} \dots \sum_{i=1}^n x_{ki} \quad (\text{Sumatoria de todas las Variables } X_i)$$

$$\sum_{i=1}^n x_{1i}^2 \quad \sum_{i=1}^n x_{2i}^2 \quad \dots \quad \sum_{i=1}^n x_{ki}^2 \quad (\text{Sumatoria de todos los cuadrados de } X_i)$$

$$\sum_{i=1}^n x_{i1} x_{i2} \quad \sum_{i=1}^n x_{i1} x_{i3} \quad \dots \quad \sum_{i=1}^n x_{i1} x_{ki} \quad (\text{Sumatoria de todas las combinaciones de } X_i)$$

Una vez realizada las sumatorias planteadas anteriormente, se procede a cargar la matriz (cargar matriz). Es de hacer notar que las matrices en todo este desarrollo de la herramienta fueron cargadas en forma dinámica, para que se adaptara a los protocolos que desea estudiar el usuario.

Luego se procede a realizar un cálculo matricial, cargando la matriz de la siguiente manera:

$$\begin{pmatrix}
 n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \dots & \sum_{i=1}^n x_{ki} \\
 \sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \dots & \sum_{i=1}^n x_{i1}x_{ki} \\
 \sum_{i=1}^n x_{ki} & \sum_{i=1}^n x_{i1}x_{ki} & \sum_{i=1}^n x_{ki}x_{2i} & \dots & \sum_{i=1}^n x_{ki}^2
 \end{pmatrix}$$

Una vez cargada la matriz con sus respectivos valores, se realiza el cálculo de la inversa de la matriz (calcular_inversa) con el método de la matriz idéntica adyacente. Como cualquier método puede generar un error si la matriz no posee inversa. Una vez calculada la matriz inversa, se procede a determinar los coeficientes de regresión (calcular_Bi). Para obtener estos se efectúa la siguiente relación:

$$\hat{B} = (X^t X)^{-1} X^t y$$

Debido a la formula planteada anteriormente, se procede a realizar una multiplicación de matrices para así obtener cada uno de los Bi (B₀, B₁,...B_k).

Luego de haber determinado los coeficientes, se procede a estudiar la bondad de ajuste del modelo. Para ello se debe calcular el Coeficiente de Determinación R² que representa la variabilidad explicada por el modelo con respecto a la variabilidad total. Que viene dada por:

$$R^2 = \frac{SCR}{SCT} \quad \text{o} \quad R^2 = 1 - \frac{SCE}{SCT}$$

Debido a esto, antes de encontrar el coeficiente de determinación se deben realizar ciertos cálculos (SCR, SCT, SCE).

En el módulo Calcular Suma de los Cuadrados Totales (Calcular_SCT) se determinó la media de los Yi, que es la sumatoria de los Yi entre la cantidad de

muestra que se utilizan para el estudio, posterior a esto se efectuó un cálculo matricial de la misma como sigue:

$$SCT = \sum_{i=1}^n (y_i - \bar{y})^2 = (y_1 - \bar{y} \ y_2 - \bar{y} \ \dots \ y_n - \bar{y}) \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \dots \\ y_n - \bar{y} \end{pmatrix}$$

Después de determinar SCT, se realizó el cálculo de SCR (calcular_SCR). Para ello se aplicó la siguiente fórmula:

$$SCR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = (\hat{y}_1 - \bar{y} \ \hat{y}_2 - \bar{y} \ \dots \ \hat{y}_n - \bar{y}) \begin{pmatrix} \hat{y}_1 - \bar{y} \\ \hat{y}_2 - \bar{y} \\ \dots \\ \hat{y}_n - \bar{y} \end{pmatrix}$$

Luego, en el módulo calcular Suma cuadrados de los errores (Calcular_SCE), se empleó la siguiente fórmula aplicando el cálculo matricial que sigue:

$$SCE = \sum_{i=1}^n e_i^2 = (e_1 \ e_2 \ \dots \ e_n) \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}$$

Una vez calculado SCT, SCR, SCE se procede a calcular el coeficiente de determinación R^2 sustituyendo los valores en la fórmula:

$$R^2 = \frac{SCR}{SCT}$$

Una vez calculado R^2 , se efectúa el módulo calcular varianza y desviación estándar de cada B_i (calc_varianzaBi). Para ello se realiza un cálculo previo que es determinar la varianza de los errores que viene dado así: $S^2 = SCE / n - k - 1$ donde n es el número de muestra y k la cantidad de variables que se están estudiando.

Para determinar la varianza de cada B_i se toma los elementos de la diagonal principal de la matriz inversa. Una vez determinada la varianza se procedió a calcular la desviación estándar de los B_i que se obtiene de la siguiente manera:

$$S_{B_i} = \sqrt{Var(B_i)}$$

La finalidad de calcular las desviaciones estándar es para determinar los intervalos de confianza de los parámetros del modelo con un nivel de confianza del 95%, siguiendo una distribución T-Student con $n - k - 1$ grados de libertad, que se establece de la siguiente manera:

$$\left[\hat{B}_i - t_{\alpha/2, n-k-1} S_{B_i}, \hat{B}_i + t_{\alpha/2, n-k-1} S_{B_i} \right]$$

Los valores de T-Student consiste en una tabla donde se intersectan los grados de libertad y el nivel de confianza α . Para manipular estos valores, se diseñó una tabla llamada *valorest* que almacena los valores indicados de la tabla T-Student.

Prueba de la herramienta

El efectuar pruebas de la herramienta corresponde al último objetivo específico que se planteó en esta investigación. Consiste en validar los resultados. Para esto se tomó como muestra las observaciones de Guedez (2005) empleadas en su trabajo de grado, las muestras tomadas por el investigador, las realizó durante cinco (5) días hábiles consecutivos desde la 8:00 a.m – 12:00 m y de 3:00 p.m – 7:00 p.m, obteniendo ocho observaciones por días lo que da un total de 40 observaciones tomada durante los cinco días.

A continuación se presenta las observaciones tomadas del trabajo de Guedez (2005):

CAPTURA DE PAQUETES DEL DIA LUNES (MEDIDO EN Kbps)

Protocolo	HTTP	FTP	DNS	POP3	SMTP	BW (Kbps)
Hora						
08am - 9am	498,92	21,78	0,217	26,64		634,557
9am - 10 am	613,12	16,96	0,243		10,8	696,123
10am - 11am	567,36	9,56	0,288	8,18		675,87
11am - 12m	529,6		0,238	12,37		635,26
3 pm - 4 pm	462,32	26,72	0,198	9,84	14,64	557,42
4 pm - 5pm	614,32		0,222	19,6	13,36	724,69
5 pm - 6 pm	624,64		0,214			732,96
6 pm - 7 pm	604,12	36,8	0,216	12,48	7,2	679,76

CAPTURA DE PAQUETES DEL DIA MARTES (MEDIDO EN Kbps)

Protocolo	HTTP	FTP	DNS	POP3	SMTP	BW (Kbps)
Hora						
08am - 9am	571,42	42,08	0,414	17,84	9,84	676,23
9am - 10 am	587,76	26,33	0,334			669,09
10am - 11am	467,19		0,156	9,34	5,56	555,09
11am - 12m	599,84		0,145	4,56		635,26
3 pm - 4 pm	525,92	16,32	0,112	10,4	3,44	633,94
4 pm - 5pm	584,08		0,202			669,8
5 pm - 6 pm	610,2	15,67	0,156	7,88	7,13	702,11
6 pm - 7 pm	598,24	8,56	0,178	9,23		694,76

CAPTURA DE PAQUETES DEL DIA MIERCOLES (MEDIDO EN Kbps)

Protocolo	HTTP	FTP	DNS	POP3	SMTP	BW (Kbps)
Hora						
08am - 9am	621,68	36,78	0,187	9,84	1,68	709,68
9am - 10 am	660,56	1,28	0,101	8,72	0,88	715,23
10am - 11am	571,44	10,34	0,056			673,29
11am - 12m	547,04		0,0023	5,67	4,23	646,89
3 pm - 4 pm	545,92		0,0345	12,23	1,34	626,45
4 pm - 5pm	656,32	0,88	0,107	0,976	1,76	763,29
5 pm - 6 pm	563,76		0,278	15,73	10,68	690,12
6 pm - 7 pm	591,6	38,32	0,243	5,67	3,23	713,45

CAPTURA DE PAQUETES DEL DIA JUEVES (MEDIDO EN Kbps)

Protocolo	HTTP	FTP	DNS	POP3	SMTP	BW (Kbps)
Hora						
08am - 9am	575,92	27,6	0,012	2,96	0,8	665,34
9am - 10 am	579,12	6,96	0,168	19,12	3,6	645,23
10am - 11am	653,44	34,34	0			701,23
11am - 12m	516,6		0	12,11	8,23	578,45
3 pm - 4 pm	560,88	12,34	0	16,23	1,34	643,45
4 pm - 5pm	617,12	7,816	0,115	7,28	1,76	689,12
5 pm - 6 pm	659,8		0,104	5,65		719,35
6 pm - 7 pm	641,23	23,45	0,234	4,67	3,44	723,15

CAPTURA DE PAQUETES DEL DIA VIERNES (MEDIDO EN Kbps)

Protocolo	HTTP	FTP	DNS	POP3	SMTP	BW (Kbps)
Hora						
08am - 9am	408,4	11,67	0,123	6,78	4,23	516,4
9am - 10 am	476,24	5,89	0,101	12,23		545,23
10am - 11am	492,16	39,21	0,056			611,89
11am - 12m	334,4	1,89	0,058	5,67	3,21	441,23
3 pm - 4 pm	527,28		0,091			587,33
4 pm - 5pm	622,12	10,45	0	4,23	1,56	682,12
5 pm - 6 pm	579,32	12,89	0			654,67
6 pm - 7 pm	539,16	5,23	0,101	7,8	4,78	616,89

Con base a las observaciones anteriores, se puede notar que son cinco (5) protocolos que se van a estudiar (DNS, FTP, HTTP, POP3, SMTP). Para efecto de comenzar la captura de datos por medio de la herramienta, se debe configurar los protocolos a estudiar por medio de la opción configuración, que se puede visualizar a continuación:

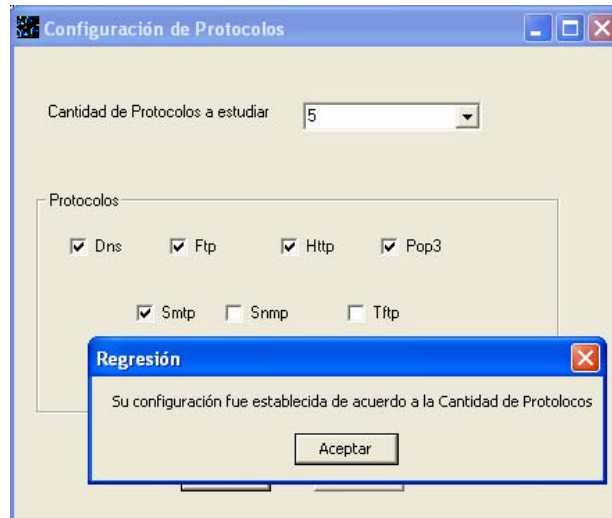


Figura 26. Configuración de los Protocolos a Estudiar (DNS, FTP, HTTP, POP3, SMTP)

Una vez establecida la configuración, se procede a la captura de los datos, en donde se introducen las 40 observaciones. A continuación se puede notar la carga de los valores de cada protocolo con el ancho de banda:

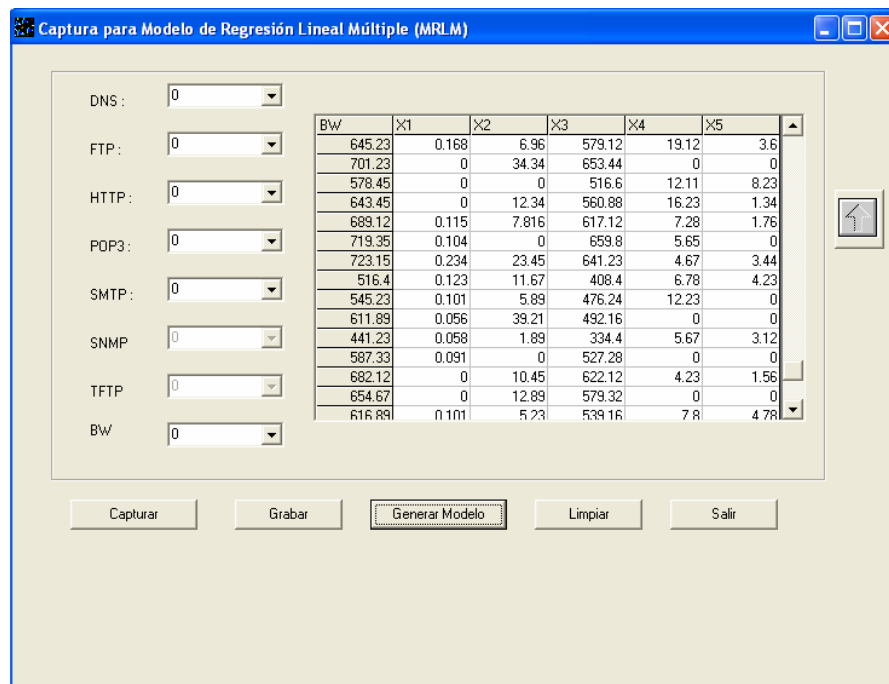


Figura 27. Captura de las 40 Observaciones

Una vez realizada la captura de los datos, se puede generar el modelo automáticamente, mostrando los resultados de la siguiente forma:

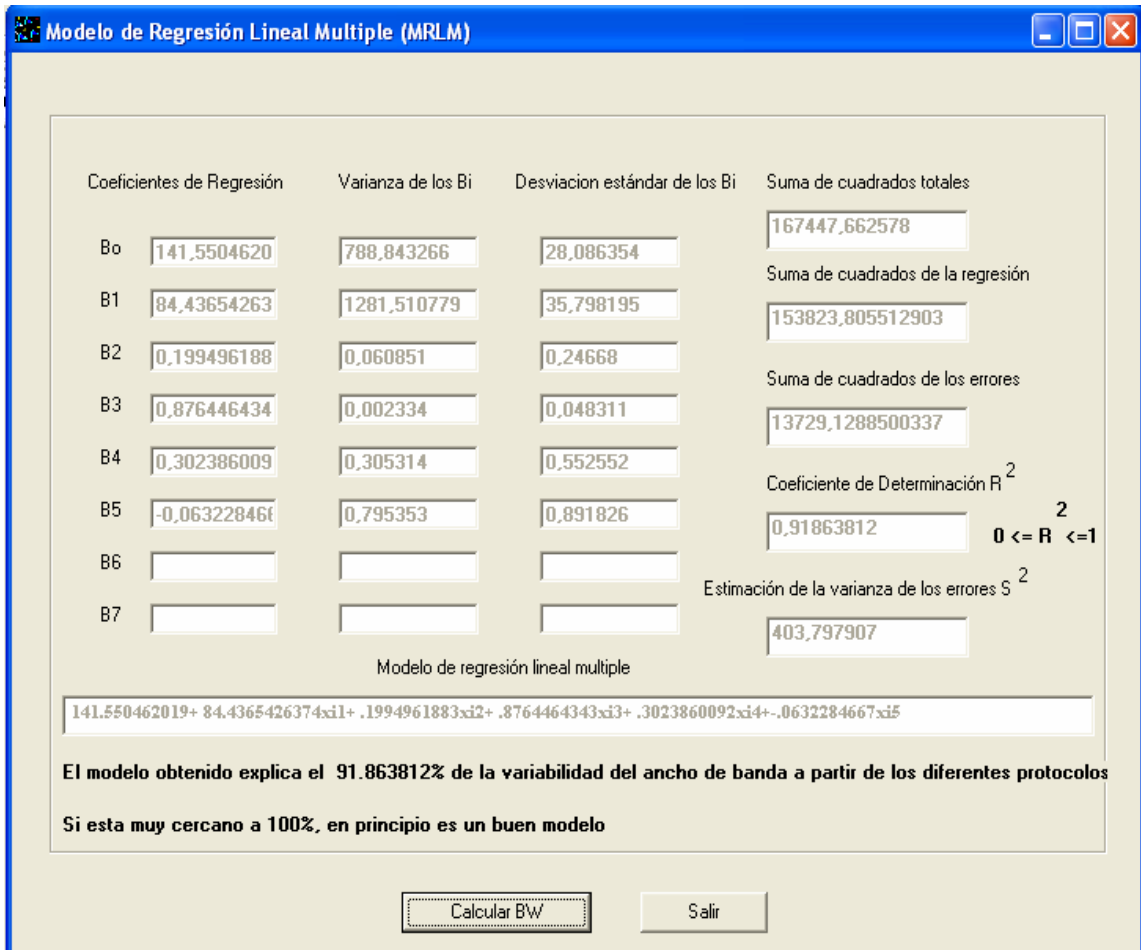


Figura 28. Modelo Generado en base a las 40 Observaciones

Se puede observar que se muestra información, en donde se determinan todos los B_i que corresponden a cada uno de los protocolos involucrados en el estudio, en donde el resultado correspondiente es el siguiente $B_0 = 141.550462019$, $B_1 = 84.43654263$, $B_2 = 0.199496188$, $B_3 = 0.876446434$, $B_4 = 0.302386009$ y $B_5 = -0.0632284667$. Esto significa que el modelo quedaría de forma genérica como se especifica a continuación:

$$BW = 141.550462019 + 84.43654263 X_1 + 0.199496188 X_2 + 0.876446434 X_3 + 0.302386009 X_4 - 0.0632284667 X_5$$

Donde cada X_i corresponde a cada uno de los protocolos correspondiente, sustituyendo cada protocolo en los X_i la forma del modelo para determinar el consumo de ancho de banda del acceso al enlace WAN es el siguiente:

$$BW = 141.550462019 + 84.43654263 * DNS + 0.199496188 * FTP + 0.876446434 * HTTP + 0.302386009 * POP3 - 0.0632284667 * SMTP$$

Además, en el resultado de la generación del modelo automáticamente, se muestra el coeficiente de determinación R^2 , este valor indica la bondad de ajuste del modelo, donde aquí basándose en los datos analizados, el ancho de banda se puede explicar a través de el modelo de regresión en forma adecuada, debido a que R^2 es 0.91863812, lo que indica que el modelo explica el 91.86 % de las variaciones del ancho de banda, dado que esta muy cercano al 100% , y en principio es un buen modelo.

Una vez estimado el modelo de regresión, es importante determinar los intervalos de confianza de cada uno de los protocolos que se están estudiando, es por ello, que puede observarse que también se genera el cálculo de la estimación de la varianza de los errores S^2 , la varianza y la desviación estándar de cada protocolo, la finalidad de generar esta información es que estos resultados nos van a permitir la determinación de los intervalos de confianza. Se muestra a continuación que los resultados son los siguientes:

Tabla 2. Varianza de los Errores del Modelo, Varianza y Desviación estándar de los protocolos (DNS, FTP, HTTP, POP3, SMTP)

Coefficientes de Regresión	Varianza	Desviación Estándar	Varianza de los Errores S^2
B_0	788.843226	28.086354	403.797907
B_1 (DNS)	1281.510779	35.798195	
B_2 (FTP)	0.060851	0.24668	
B_3 (HTTP)	0.002334	0.048311	
B_4 (POP3)	0.305314	0.552552	
B_5 (SMTP)	0.795353	0.891826	

A partir de los valores en la tabla 2, se puede construir las estimaciones de los intervalos de confianza para los coeficientes de Regresión, que se determinaron a partir de una distribución t de Student con $n-k-1$ grados de libertad, con un nivel de confianza del 95%, tomando los valores de la distribución de Student con $n-k-1 = 40 - 5 - 1 = 34$ grados de libertad el valor crítico correspondiente para $\alpha/2 = 0.025$ siendo este valor de 2.042, obteniéndose los siguientes resultados:

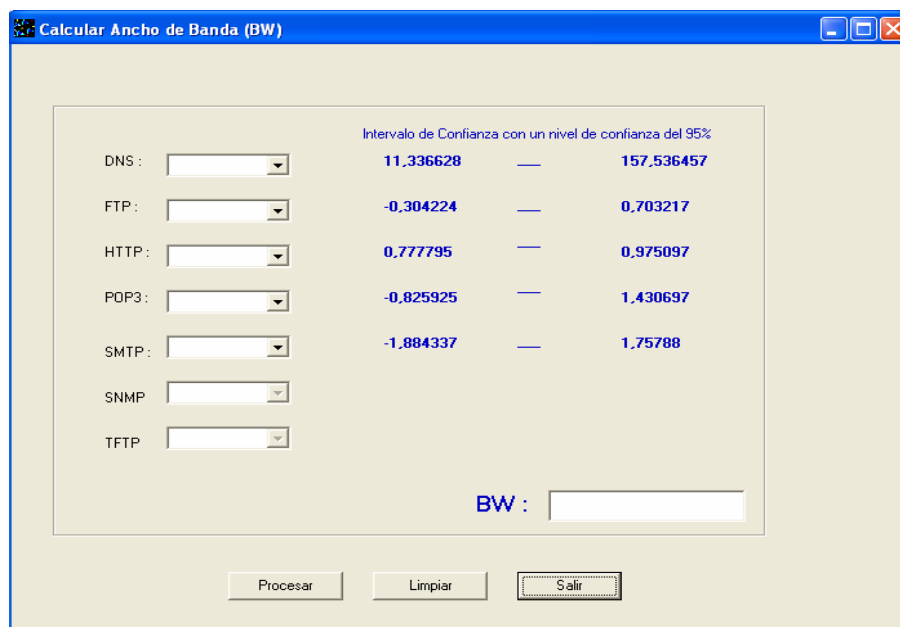


Figura 29. Intervalos de Confianza de cada Protocolo

Tabla 3. Valores de los Intervalos de Confianza para cada protocolo (DNS, FTP, HTTP, POP3, SMTP)

Coefficientes de Regresión	Valor Mínimo	Valor Máximo
B ₁ (DNS)	11.33628	157.536457
B ₂ (FTP)	-0.304224	0.703217
B ₃ (HTTP)	0.777795	0.975097
B ₄ (POP3)	-0.855925	1.4306697
B ₅ (SMTP)	-1.884337	1.75788

Los valores que se presentan en la Tabla 3, representan los valores máximos y mínimos que pueden tomar los respectivos protocolos en un enlace de situación de no saturación con un 95% de confianza. Es decir cuando las variables en estudio oscilan entre los valores máximos y mínimos especificados, el enlace no presenta saturación. Una vez determinado estos intervalos de confianza de forma automática, se puede generar el ancho de banda BW introduciendo valores que estén dentro de los valores especificados de cada protocolo. A continuación se puede observar el resultado del ancho de banda:

Intervalo de Confianza con un nivel de confianza del 95%

DNS :	12	11,336628	—	157,536457
FTP :	0,5	-0,304224	—	0,703217
HTTP :	0,5	0,777795	—	0,975097
POP3 :	1,2	-0,825925	—	1,430697
SMTP :	1,3	-1,884337	—	1,75788
SNMP :				
TFTP :				

BW : 1155,60761118

Procesar Limpiar Salir

Figura 30. Calculo del Ancho de Banda (BW)

Luego de haber obtenido el modelo de regresión lineal múltiple que describe la interrelación entre el ancho de banda (BW) y las cinco variables independientes estudiadas que corresponden a los cinco (5) protocolos. Se generaron una serie de graficas correspondiente a cada día con los datos analizados. Para la generación de las gráficas, el sistema cuenta con una opción Graficar, que a su vez permite escoger la fecha en la que se desea efectuar el gráfico, una vez indicada la fecha como se muestra en la figura 31, se procede a generar el gráfico correspondiente a esa fecha, y así sucesivamente se realiza el mismo proceso de escogencia de la fecha para generar el resto de las gráficas.

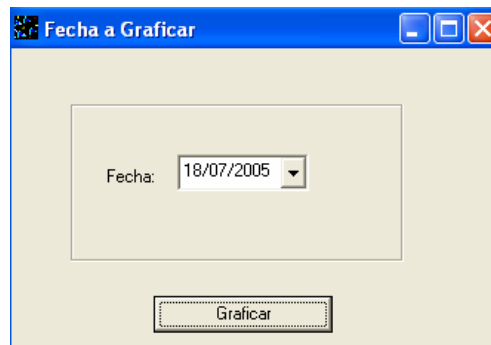


Figura 31. Escoger la fecha a Graficar

A continuación se muestra las cinco (5) gráficas, que representa una por día

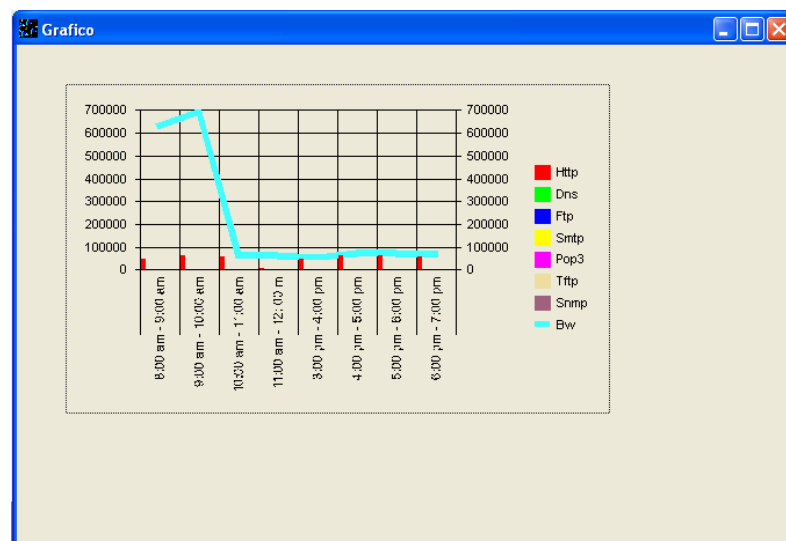


Gráfico 1. Correspondientes a los datos analizados el día Lunes

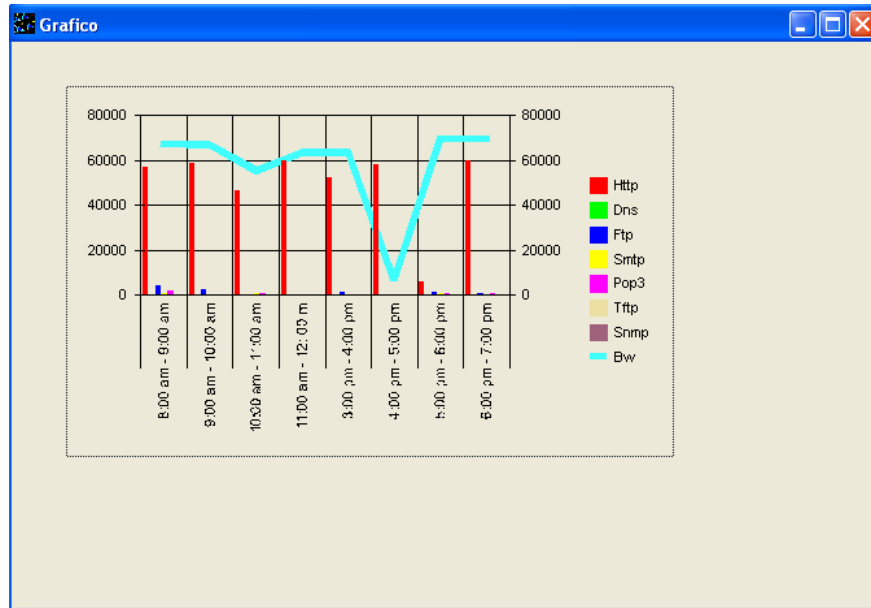


Gráfico 2. Correspondientes a los datos analizados el día Martes

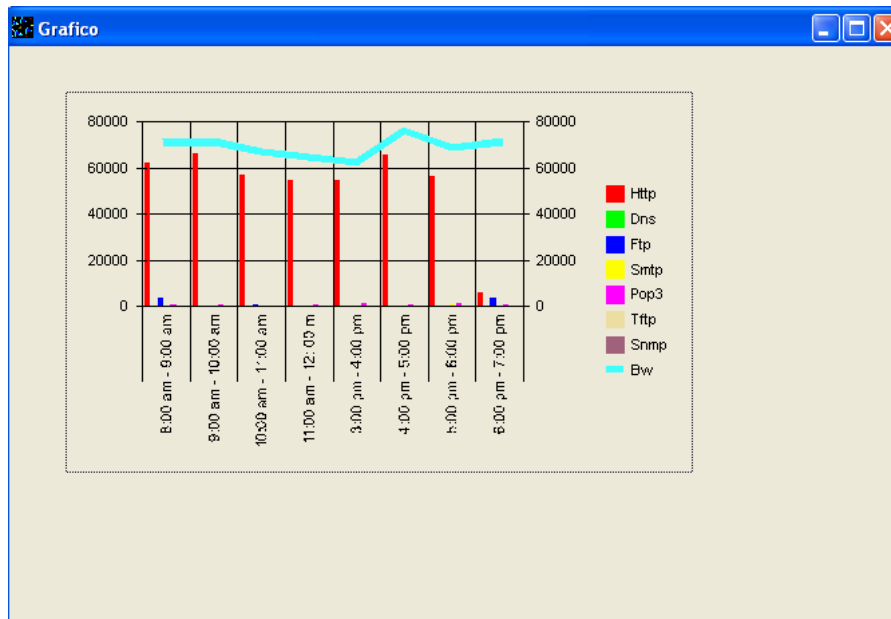


Gráfico 3. Correspondientes a los datos analizados el día Miércoles

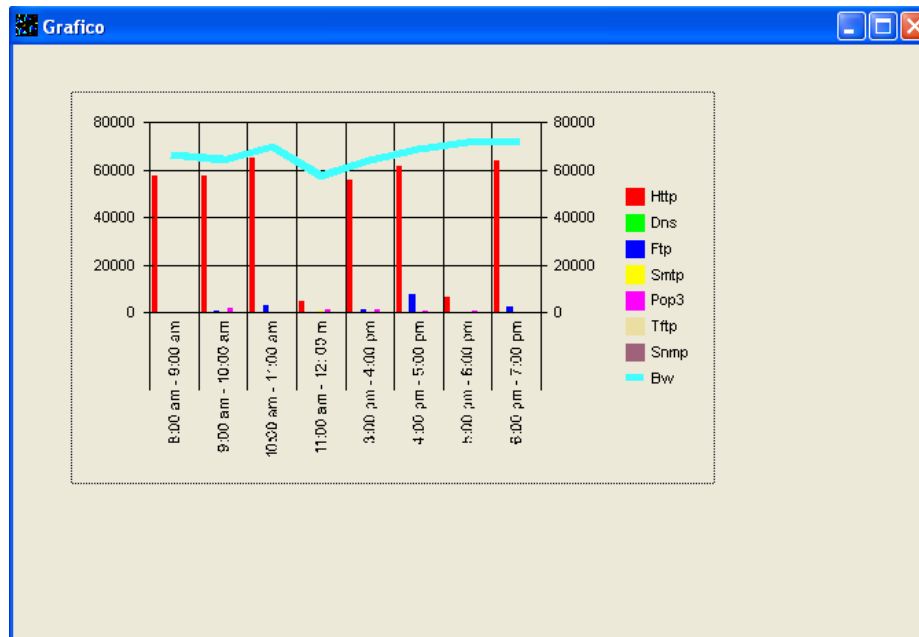


Gráfico 4. Correspondientes a los datos analizados el día Jueves

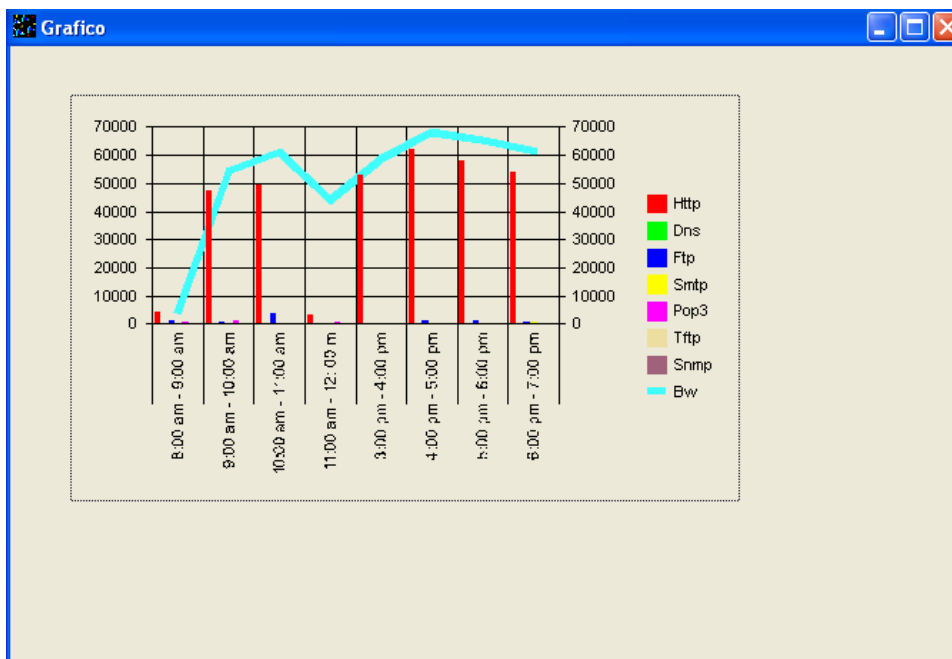


Gráfico 5. Correspondientes a los datos analizados el día Viernes

Se pueden observar en los gráficos anteriores (Gráfico 1 al Gráfico 5) que el protocolo que consume más ancho de banda es el HTTP, siendo el resto de los

protocolos (DNS, FTP, POP3, SMTP) de poca influencia en el ancho de banda del enlace WAN.

Análisis y Comparación de Resultados

Con la realización de la prueba, basada en la cuarenta (40) observaciones tomadas por Guedez (2005) en su trabajo de grado, se pudo generar el modelo automáticamente que describe la relación BW con los cinco (5) protocolos que se estudiaron (DNS, FTP, HTTP, POP3, SMTP), determinando los coeficientes de regresión equivalentes a los obtenidos por Guedez (2005) en su investigación, el modelo generado explica que en principio es un buen modelo porque tiene un coeficiente de relación R^2 de 0.916 equivalente al 0.92 (el obtenido por Guedez (2005)), la herramienta indica de manera informativa lo importante que el valor de R^2 para el estudio ya que este explica la variabilidad del ancho de banda (BW) con respecto al protocolo.

Con respecto a los intervalos de confianza de los protocolos, difiere los valores obtenidos en la investigación de Guedez (2005) con respecto a esta, ya que se generan valores muy distintos, aquí para la obtención de estos intervalos se empleó la T-Student determinándolos con un nivel de confiabilidad al 95%.

Referente a la generación de gráficas por días, que representan la relación de BW con respecto a los protocolos generados, al comparar las gráficas obtenidas con las gráficas presentadas en el estudio anterior, estas son equivalentes, de tal forma que se llega a la conclusión que determinó Guedez (2005) que el protocolo que más consume más ancho de banda en el acceso al enlace WAN es el HTTP, y el resto de los protocolos tienen muy poco consumo, y por tanto, poca influencia en el ancho de banda BW, lo que demuestra que la herramienta genera resultados esperados a los ya estudiados.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

Este trabajo logró cumplir con todos los objetivos propuesto de la investigación realizada, dando así solución al problema estudiado mediante el diseño y desarrollo de una herramienta automatizada que permite modelar en tiempo real el consumo de ancho de banda del acceso al acceso WAN de una red, tal como fue planteado en el objetivo general del proyecto.

1. La herramienta diseñada, permite generar en forma automática un modelo de Regresión Lineal Múltiple, que describe la interrelación entre el Ancho de Banda (BW) con las variables independiente que son los protocolos que se desean estudiar.
2. La herramienta presenta flexibilidad, debido que permite configurar la cantidad de protocolos que se desea estudiar y escogerlos, lo que implica que se adapte a cualquier cambio que pueda ocurrir, a su vez permite la generación dinámica del modelo ya que se realiza automáticamente.
3. Con la realización de la prueba, basada en la cuarenta (40) observaciones tomadas por Guedez (2005) en su trabajo de grado, se pudo generar el modelo automáticamente que describe la relación BW con los cinco (5) protocolos que se estudiaron (DNS, FTP, HTTP, POP3, SMTP), determinando los coeficientes de regresión equivalentes a los obtenidos por Guedez (2005) en su investigación, el modelo generado explica que en principio es un buen modelo porque tiene un coeficiente de relación R^2 de 0.916 equivalente al 0.92 (el obtenido por Guedez (2005)), lo que indica que la herramienta arroja resultados apropiados.

4. Con los datos analizados de las observaciones de Guedez (2005), se obtienen gráficas por días, que representan la relación de BW con respecto a los protocolos generados, estas gráficas generadas al compararlas con el estudio anterior son equivalentes, lo que demuestra que la herramienta genera resultados esperados a los ya estudiados.
5. La herramienta permite calcular intervalos con un nivel de confiabilidad del 95%, proporcionando rango de valores dentro del cual no se presente saturación al acceso del enlace WAN.
6. La herramienta valida que el número de observaciones (n), sea mayor que el número de variables independientes (k), es decir $n > k + 1$.
7. La herramienta contempla el control de colinealidad, ya que al existir una combinación lineal de una variable independiente (en este caso los protocolos en estudio) con respecto a otra, la matriz que se carga (X^tX) al calcular su inversa $(X^tX)^{-1}$ no se puede invertir, debido a que existen condiciones matemáticas que establecen que: si los elementos de una fila (columna) de una matriz cuadrada son combinación lineal de las filas (columnas) restantes, es decir son el resultado de sumar los elementos de otras filas (columnas) multiplicadas por números reales, su determinante es cero (0); y una matriz cuadrada tiene inversa cuando su determinante es distinto de cero (0). En vista de las condiciones anteriores, la herramienta valida la colinealidad, porque al presentarse esa situación se indica que no se puede invertir la matriz y por tanto no se puede generar el modelo.
8. La herramienta no contempla el estudio de si una variable X_i (protocolos) no es una variable explicativa del comportamiento de la variable Y (BW), y prescindir de ella sino contribuye en el modelo generado.

Recomendaciones

1. Mejorar la herramienta logrando que la captura de los datos se efectúe en tiempo real, permitiendo de esta manera que la captura de los datos no sea tediosa al momento de tener una gran cantidad de datos de entradas.
2. La herramienta puede ser mejorada en relación a la forma de capturar los datos, esto es que los datos estén tabulados en una hoja de Excel o en una tabla por fecha y hora, y de esta forma importar los datos a la aplicación y luego generar el modelo automáticamente.
3. Implementar la herramienta en diferentes escenarios, para observar su funcionalidad

CAPÍTULO VI

SOFTWARE DE IMPLEMENTACION QUE GENERA EL CONSUMO DE ANCHO DE BANDA AL ACCESO DEL ENLACE WAN DE UNA RED

De acuerdo con los objetivos planteados en este trabajo de grado, se presenta a continuación una breve explicación del software elaborado para modelar en tiempo real el consumo de ancho de banda del enlace WAN de una red.

En la Figura 32 se muestra el menú principal de la herramienta que consta de cuatro (4) opciones: Configuración, Modelo, Graficar y Salir

La opción Configuración permite al usuario acceder a la pantalla de configurar protocolos, para establecer y elegir los protocolos a estudiar.

En la opción Modelo se despliega una opción Capturar Datos, que permite la introducción de los datos para posteriormente generar el modelo.

La opción Graficar permite al usuario entrar a la pantalla donde se introduce la fecha correspondiente a los valores de las variables introducidas en la pantalla Capturar Datos.

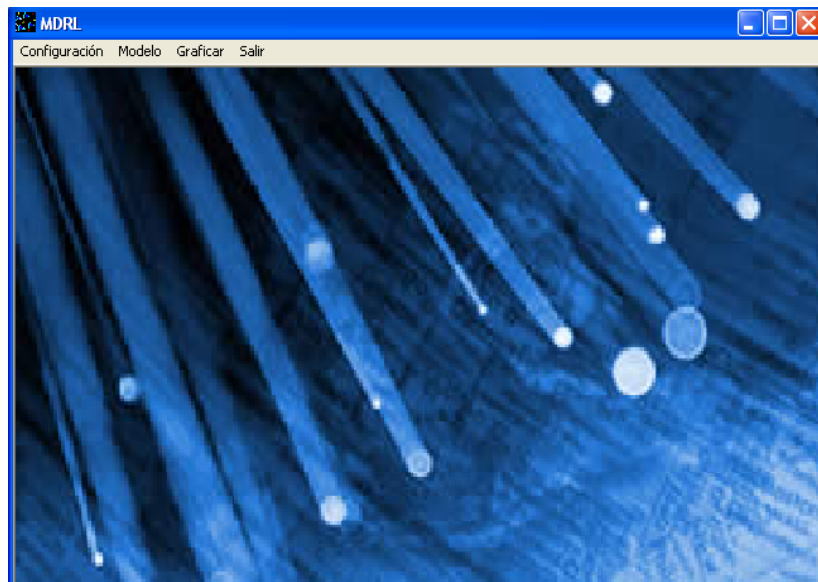


Figura 32. Menú Principal de la herramienta Modelo de Regresión Lineal Múltiple

Al hacer clic en la opción Configuración de la Figura 32, se visualiza la pantalla que se muestra en la Figura 33, esta permite al usuario indicar la cantidad de protocolos a estudiar y a su vez permite la selección de los mismos.

En la pantalla de Configuración de protocolos, tiene dos botones: 1) Aceptar que permite al usuario establecer la configuración, 2) la opción Salir que permite al usuario regresar al menú principal.

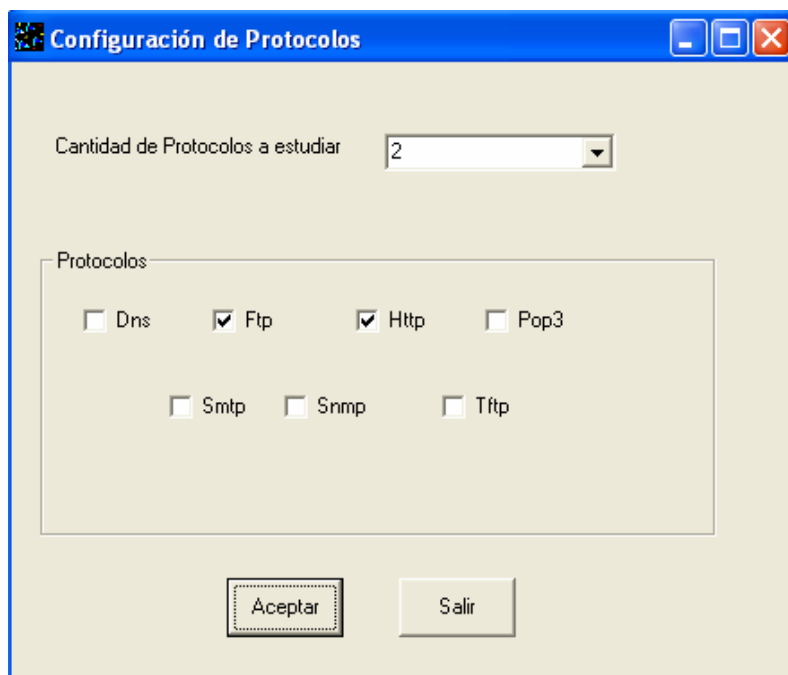


Figura 33. Pantalla de Configuración de Protocolos

Al seleccionar la opción Modelo del menú principal Figura 32, esta despliega una opción llamada Capturar Datos, al hacer clic en la opción, se visualiza la pantalla que se muestra en la Figura 34, la cual permite introducir los valores por protocolos, ya sea por teclado o archivo de texto.

Esta pantalla consta de varios botones:

La opción Capturar, que al hacer clic muestra una pantalla Figura 35 que va a permitir seleccionar el directorio en donde se encuentra los archivos de textos que contienen los valores de los protocolos.

La opción Grabar permite al usuario guardar los valores introducidos en cada uno de los protocolos. El sistema le solicitará la fecha y la hora asociada a la información a guardar. Una vez guardada la información, esta se va reflejando en la matriz de valores, que le da la facilidad al usuario de visualizar los datos que ha introducido.

La opción Generar Modelo permite al usuario visualizar la pantalla Figura 36, la cual muestra el modelo generado automáticamente.

La opción Limpiar permite al usuario colocar en cero todos los valores de las variables y blanquear la matriz de valores.

La opción salir permite al usuario regresar al menú principal.

BW	X1	X2
24.6	1	11
33.0	3	13
36.6	4	13
39.8	4	14
28.6	2	12

Figura 34. Pantalla de Captura de Datos para el Modelo de Regresión Lineal Múltiple

En la Figura 35 se visualiza la pantalla Directorio que se muestra al hacer clic en la opción Capturar de la Pantalla de captura de datos.

La opción Aceptar permite mostrar otra pantalla Figura 37 donde se muestra los valores que contiene el archivo de texto seleccionado.

La opción Salir permite al usuario regresar a la pantalla de Captura de Datos para el Modelo de Regresión Lineal Múltiple.

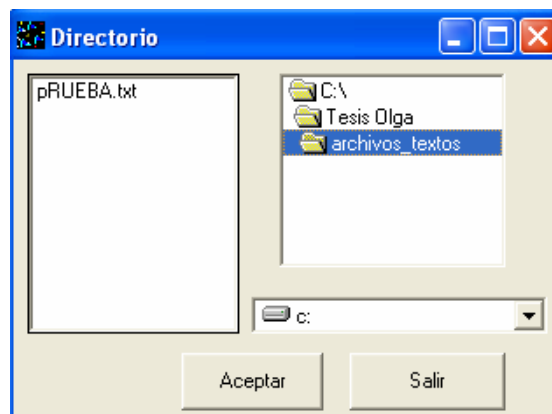


Figura 35. Pantalla de Selección de Directorio

En la Figura 36 esta pantalla permite al usuario visualizar el modelo de regresión lineal múltiple generado, esta contiene varias opciones:

La opción Calcular Bw va a otra pantalla Figura 38, donde se muestra los valores de los intervalos de confianza al 95% que no saturaran al enlace.

La opción Salir permite al usuario regresar a la pantalla de Captura de Datos para el Modelo de Regresión Lineal Múltiple.

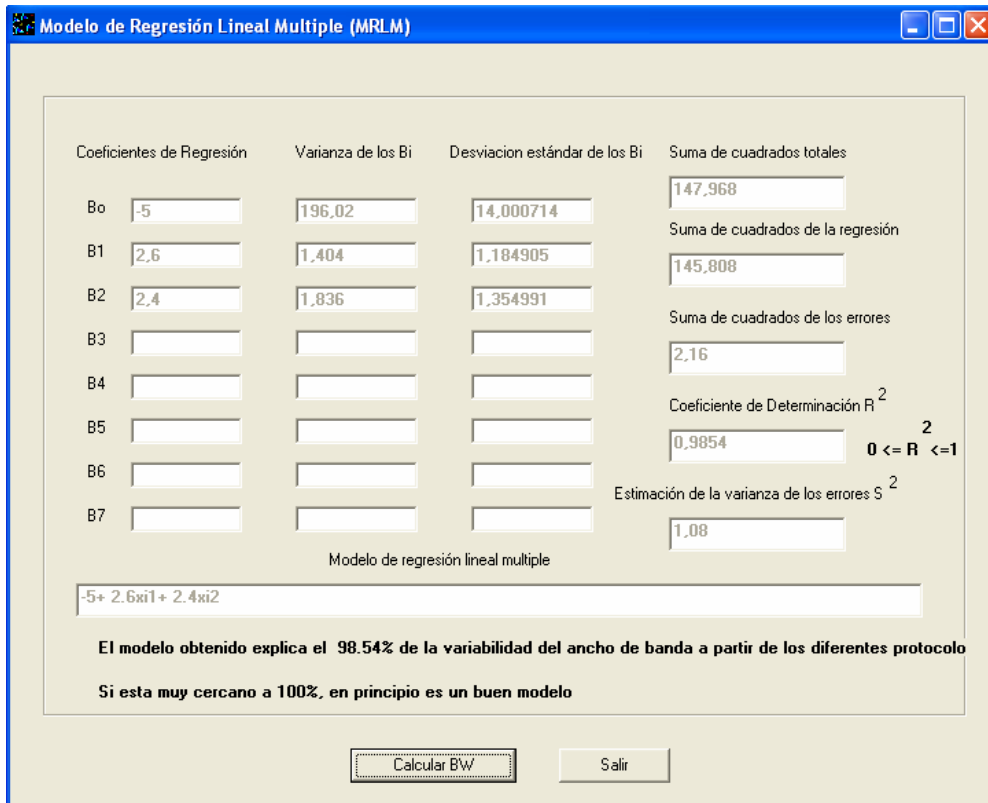


Figura 36. Pantalla del Modelo de Regresión Lineal Múltiple

La figura 37 esta pantalla muestra al usuario los valores que contiene el archivo de texto seleccionado.

La opción cambiar archivo de texto permite al usuario ir a la pantalla de directorio Figura 35 para cambiar el archivo en caso de que no sea el que deseaba.

La opción Aceptar permite al usuario tomar los datos del archivo de texto y vaciar los valores en la pantalla de captura de datos Figura 34, en cada protocolo correspondiente.



Figura 37. Pantalla que muestra el Archivo de Texto

En la Figura 38, esta pantalla le permite al usuario introducir los valores de los protocolos, mostrándole los intervalos de confianza respectivamente.

La opción Procesar permite calcular el ancho de banda.

La opción Limpiar permite al usuario colocar en cero todos los valores de los protocolos y del ancho de banda.

La opción Salir permite al usuario regresar a la pantalla del Modelo de Regresión Lineal Múltiple Figura 36.

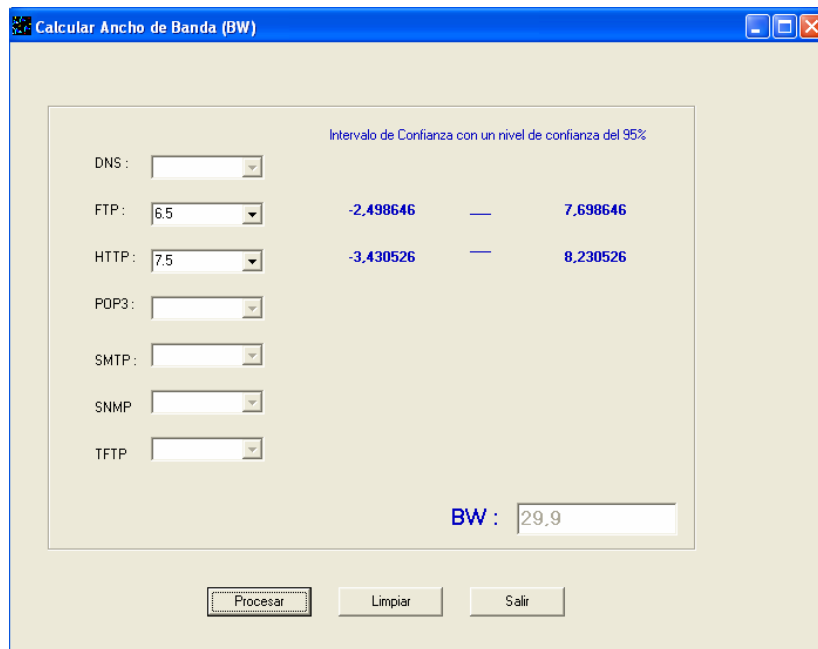


Figura 38. Pantalla para Calcular el Ancho de Banda

Al seleccionar la opción graficar de la Figura 32, se obtiene la pantalla de la Figura 39, esta permite al usuario introducir la fecha en la que desea graficar los protocolos estudiados, una vez que introduce la fecha, al hacer clic en la opción graficar se genera la gráfica correspondiente a esa fecha Figura 40.

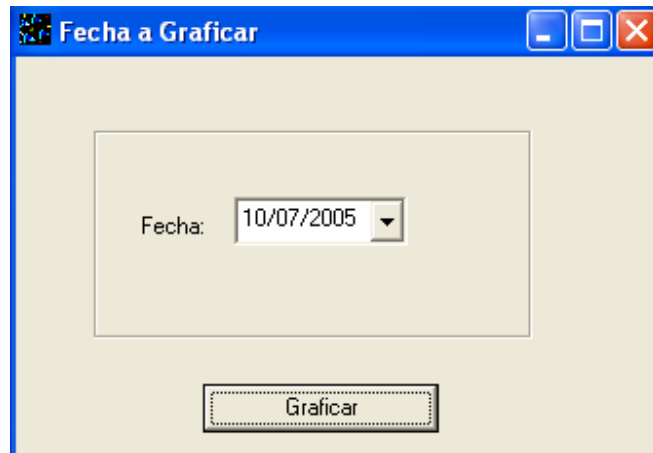


Figura 39. Pantalla para Escoger la Fecha para graficar.

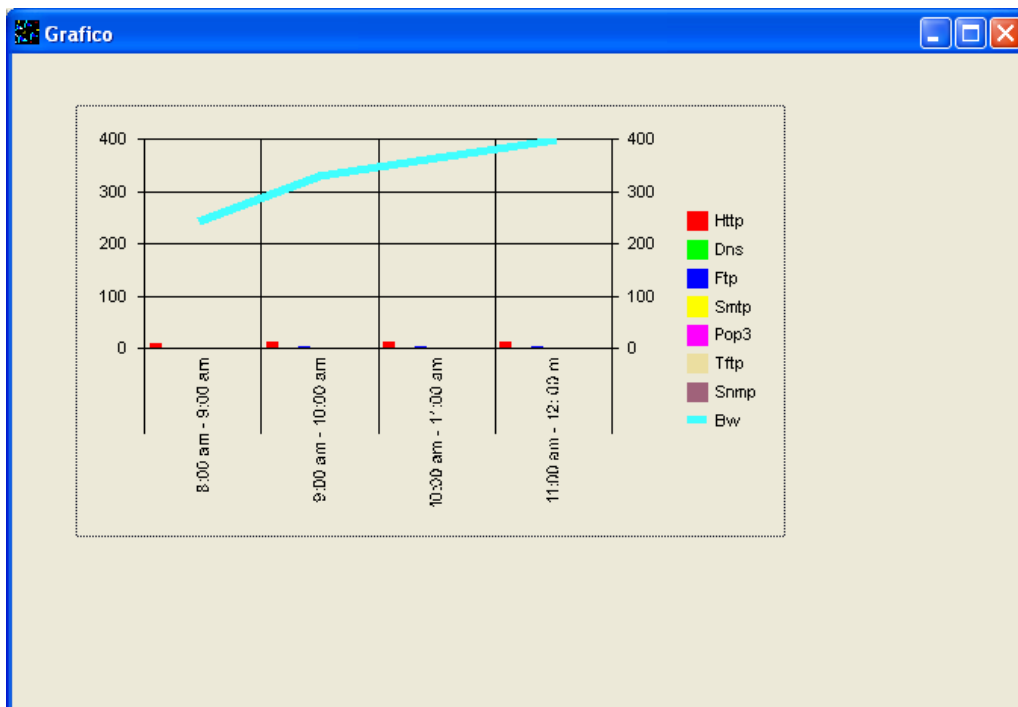


Figura 40. Pantalla de Gráfica, clasificado por hora y fecha.

REFERENCIAS BIBLIOGRAFICAS

- Alvarez, C., Azcorra, A., Berrocal, J., Domingo, J., Larrabeiti, D., Martinez, X., Moreno, J., Pérez, J y Solé J. 1998. CASTBA: Medidas de tráfico sobre la Red Académica Española de Banda Ancha. Recuperado el 30 de Septiembre de 2005 de URL: <http://greco.dit.upm.es/~enrique/pub/castba-telecom98.pdf>.
- Arsham, H. 2002. Razonamiento estadísticos para Decisiones Gerenciales. Recuperado el 25 de Noviembre de 2005 de URL: <http://home.ubalt.edu/ntsbarsh/Business-stat/opre504s.htm>.
- Barakat, C., Thiran, P., Iannaccone, G. y Diot, C. 2005. Modelando Trafico en el Enlace Principal de Internet. Recuperado el 14 de junio de 2005 de URL: http://ipmon.sprint.com/pubs_trs/pubs/gianluca/flowmodel.2pages.pdf.
- Barón, L. y Téllez, M. 2000. Regresión Múltiple. Recuperado el 25 de Noviembre de 2005 de URL: http://campusvirtual.uma.es/est_fisio/apuntes/ficheros/cap06.pdf.
- Borray, G., Sánchez, J. y Wilches J. 2001. Modelado matemático del tráfico de redes de telecomunicaciones de acceso fijo. Recuperado el 28 de Septiembre de 2005 de URL: http://www.crt.gov.co/documentos/biblioteca/Modelo_matematico_de_redes.PDF.
- Cheng, R. 2005. Análisis de Protocolo. Recuperado el 28 de Septiembre de 2005 de URL: <http://ruben.cheng-ca.com/es/knowledge/network/trafficanalysis.htm>.
- Cheng, R. 2005. Monitoreo de tráfico. Recuperado el 30 de Septiembre de 2005 de URL: <http://ruben.cheng-ca.com/es/knowledge/network/trafficmon.htm>.
- Cole, J. 2004. Nociones de Regresión Lineal. Recuperado el 20 de enero de 2005 de URL: <http://eumed.net/cursecon/medir/index.htm>.
- Conde, A. 2002. Matriz Inversa. Recuperado el 04 de Marzo de 2006 de URL: <http://thales.cica.es/rd/Recursos/rd99/ed99-0289-02/index.html>
- Cox, G. 2004. Modelo de Regresión Lineal Múltiple. Recuperado el 29 de Noviembre de 2005 de URL: <http://www.udc.es/dep/mate/estadistica2/cap8.html>.
- Galan, F., Romeral, R., Garcia, A., Garcia, Ana., Alvarez, M. y Azcorra, M. 2003. MIRA: Plataforma de Monitorización y Análisis de Tráfico para Redes IP. Recuperado el 30 de septiembre de 2005 de URL: <http://greco.dit.upm.es/~abgarcia/publications/2003TELECOM.pdf>.

- García, R., Mateo, M., Pérez, M. 2002. Analizador de protocolos TCP/IP para ayuda en la docencia de redes de computadores. Recuperado el 28 de Septiembre de 2005 de URL: <http://www.grc.upv.es/mperez/teaching/pubs/Jenui2k1.pdf>.
- Gibergasn, J. 2002. Regresión Lineal Múltiple. Recuperado el 22 de Febrero de 2006 de URL: <http://cv.uoc.es/cdocent/IL82WF018QCT0FBUCJ00.pdf>.
- Gordo, R. 1998. La transmisión de información en Internet. Recuperado el 30 de Septiembre de 2005 de URL: <http://www.geocities.com/SiliconValley/Bay/8259/contenido.html>.
- Grajales, T. 2000. Tipos de Investigación. Recuperado el 17 de Noviembre de 2005 de URL : <http://tgrajales.net/investigtipos.pdf>.
- Guedez, J. 2005. Propuesta de un modelo de predicción de Tráfico para el acceso a redes de banda ancha en la universidad Fermín Toro. Trabajo de Grado. Universidad Centrocidental “Lisandro Alvarado”. Barquisimeto. 66 p.
- MacLaurin. 1978. Cálculos de determinantes. Recuperado 04 de Marzo de 2006 de URL: <http://www.terra.es/personal2/jpb00000/tdeterminantes.htm>
- Magaña, E. y Izukue, E. y Villadangos, J. 2001. Análisis de Prestaciones de un Planificador de Tráfico sobre Plataformas de Propósito General. Recuperado el 20 de agosto de 2003 de URL: <http://www.tlm.unavarra.es/~eduardo/publicaciones/20013-alcom-espanol.pdf>.
- Maraboli, M. 1997. Monitor de Tráfico Ethernet Netgraph. Recuperado el 30 de Septiembre de 2005 de URL: <http://elqui.dscs.utfsm.cl/netgraph/netgraph/>.
- Nicaró, O. 2002. Fundamentos esenciales sobre redes. Recuperado el 30 de Octubre de 2005 de URL: <http://www.geocities.com/nicaraocalli/Redes/Redes.htm>.
- Pértega Díaz, S. 2001. Técnicas de Regresión Lineal Múltiple. Recuperado el 15 de Noviembre de 2005 de URL: http://www.fisterra.com/mbe/investiga/regre_lineal_multi/regre_lineal_multi.htm
- Rodríguez, M y Escanciano, J. 2001. Métodos de Regresión. Recuperado el 25 de Noviembre de 2005 de URL: http://www.uam.es/personal_pdi/ciencias/mayter/docencia/regresion/regresion.ht

ml.

Roca, F. 2004. Tabla de la Distribución T de Student. Recuperado el 04 de Marzo de 2006 de URL: http://usuarios.lycos.es/guillemat/t_student.htm.

Rincón, M. 2005. Regresión Lineal. Recuperado el 10 de Diciembre de 2005 de URL: <http://www.monografias.com/trabajos27/regresion-simple/regresion-simple.shtml>.

Sabino, C. 2000. El proceso de investigación. Editorial Panapo. Caracas.

Enrique Sánchez, I. 2001. Cuando el problema es la colinealidad. Recuperado el 04 de Noviembre de 2005 de URL: <http://congresoestadistica.umh.es/web/resumenes/resumen%20II/E.S%C3%A1nchez.htm>

Universidad Centroccidental “Lisandro Alvarado” (UCLA). 2002. Manual para la Elaboración del Trabajo Conducente al Grado Académico de: Especialización, Maestría y Doctorado.

Universidad Pedagógica Experimental Libertador. 2002. Manual de Trabajos de Grado de Especialización, Maestría y Tesis Doctorales. Caracas – Venezuela.

Villadangos, J. y Magaña, E. 2001. Garantía de calidad de servicio basada en la predicción del ancho de banda. Recuperado el 25 de septiembre de 2003 de URL: <http://www.tlm.unavarra.es/~eduardo/publicaciones/20010901-jitel.pdf>.

Walpole, R. y Myers, R. 1992. Probabilidad y Estadística. Editorial McGraw Hill. México.

Wolti, L. 2002. Introducción al Análisis de Regresión Lineal. Recuperado el 01 de Septiembre de 2005 de URL: <http://web.media.mit.edu/~lwolti/documents/statistics/Analisis%20de%20Regresion.pdf>.

Wikipedia. 2006. Matriz Invertible. Recuperado el 04 de Marzo de 2006 de URL: http://es.wikipedia.org/wiki/Matriz_inversa

ANEXO A

α Γ	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,378	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,766	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,696	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290