Universidad Nacional del Litoral
Proyecto Final de Carrera

Diseo de un sistema de deteccin de anomalas en redes de computadoras.

realizado por
Leandro Pineda

dirigido por Ing. Miguel Angel Robledo
codirigido por Ing. Gabriel Filippa

Santa Fe
Septiembre de 2016

Resumen

En la actualidad las redes informticas estn presentes en nuestro da a da, as como tambin en las operaciones diarias de prcticamente cualquier organizacin. Debido a la demanda masiva de servicios on-line, el tamao y el volumen de datos que manejan las entidades que proveen estos servicios incrementa ao tras ao. Por esto, las herramientas que facilitan la tarea de administracin y diagnstico de redes son esenciales para ofrecer un servicios de mejor calidad, y aprovechar al mximo la utilizacin de recursos existentes. An ms crticas son aquellas herramientas que permiten detectar ataques que atentan contra los datos que estas organizaciones manejan. Este documento describe un proyecto de desarrollo de un sistema de deteccin de anomalas basado en deteccin de comportamiento para la Secretara de Tecnologas para la Gestin de la Provincia de Santa Fe. En la primera seccin se describe el problema a solucionar y algunas herramientas disponibles en la actualidad. Luego se definen los objetivos y alcances del proyecto y la metodologa propuesta para su desarrollo. A continuacin se describe el plan de tareas y los informes de avance que se esperan obtener en cada etapa. Finalmente se identifican los riesgos y se estiman los costos del proyecto.

Palabras claves *intrusion detection system*, *redes informticas*, *data streaming*, *monitoreo*, *seguridad informtica*

Justificacin

La cantidad de servicios que organismos pblicos y privados ofrecen a travs de Internet aumenta constantemente, y en consecuencia tambin lo hace la complejidad de la infraestructura y los sistemas necesarios para proveerlos.

La integridad de estos sistemas puede verse comprometida principalmente por dos factores: las fallas de los componentes de la infraestructura de red[1], que afectan a una parte o a la totalidad de las funcionalidades del sistema, y los numerosos tipos de ataques[2][3] a los servicios ofrecidos al pblico. Hacer frente a estos ataques es de extrema importancia dado que su fin puede ser no solo afectar la calidad de los servicios ofrecidos, sino que su objetivo puede ser expropiar informacin sensible o crtica para el desarrollo de las actividades de la organizacin. Adems de ataques externos, los sistemas pueden ser vulnerados por usuarios malintensionados con acceso fsico a la red interna de la organizacin, o por actores inadvertidos[4][5][6][7].

La utilizacin de herramientas clsicas para implementar polticas de seguridad como firewall, lista de control de acceso, credenciales de usuario, proxy de red entre otros, resultaba adecuado 10 aos atrs, pero en la actualidad estas herramientas estn siendo cada vez menos efectivas en la tarea de bloquear ataques dirigidos y malware avanzados. Esto hace que cada vez sea ms comn la utilizacin de otros mtodos que monitorean continuamente eventos en busca de patrones que puedan indicar comportamientos anmalos en el trfico de red, provocados tanto por ataques como por fallas en los dispositivos de red. Los sistemas de deteccin/prevencin de intrusiones (IDS o IPS por las siglas en ingls *Intrusion Detection/Prevention System*) pueden realizar esta tarea utilizando dos enfoques: deteccin de uso indebido o *signature-based detection* y deteccin de anomalas o *anomaly-based detection*[8]. En el primer caso, el sistema determina la ocurrencia de ataques comparando la actividad de la red y de los subsistemas que la componen, contra un conjunto de patrones de ataques conocidos. Si los patrones coinciden, el ataque es informado y pueden tomarse medidas al respecto. En el segundo caso, se modela el comportamiento normal de la red y se usa como base para identificar comportamientos anmalos, los cuales pueden ser provocados por fallas o indicadores de potenciales ataques.

Como alternativa a las soluciones de hardware dedicado o sistemas privativos que ofrecen algunas empresas, la comunidad de software libre desarrolla y mantiene la aplicacin Snort[1], un IDS basado en deteccin de uso indebido. Si bien esta herramienta es efectiva detectando ataques conocidos, es dependiente de la definicin de un conjunto efectivo de reglas por parte del administrador (las cuales son difciles de mantener) y de una base de datos actualizada con los patrones de ataques o *malware*. Otra desventaja de esta tcnica de deteccin es la imposibilidad de detectar ataques *zero-day*[2][8]. Otra herramienta que suele utilizarse en conjunto con Snort es Bro[3], una plataforma para anlisis de trfico de red. Sin embargo, Bro tambin se basa en la definicin de reglas de monitoreo y utiliza deteccin de uso indebido. Por otro lado, las herramientas de deteccin de anomalas cmo Ourmon[4] utilizando mtodos de deteccin estadsticos y no estn diseados para procesar grandes volmenes de datos. Es importante destacar que el desarrollo de este proyecto genera adems un aporte a la comunidad de software libre.

El diseo de una herramienta para la identificacin de comportamientos sospechosos en redes de informacin, mediante el anlisis del trfico de red con mtodos de *streaming* para modelar de forma automtica el comportamiento normal de la red utilizando tcnicas de conteo, permite identificar y alertar sobre comportamiento anmalo en una infraestructura de red de datos. Esto complementa las funcionalidades de las herramientas basadas en deteccin de uso indebido dado que el modelado

[1] https://www.snort.org/

[2] ataques que explotan vulnerabilidades que an no han sido conocidas o que no estn contempladas en al conjunto definido de reglas

[3] https://www.bro.org/

[4] http://ourmon.sourceforge.net/

automtico de comportamiento hace que no sea necesario utilizar un conjunto de reglas definidas por administradores. El modelo de comportamiento normal del trfico de la red ser calculado por el software, brindando la posibilidad de detectar efectivamente cualquier evento poco usual en los sistemas. Para esto, se har uso de la informacin generada por los dispositivos de infraestructura de red como routers y los logs de servidores y subsistemas que interactuan dentro de la red.

Adems, al calcular automticamente el modelo de comportamiento, es fcil de implementar y al utilizar informacin de la red su implementacin requiere pocas modificaciones en los sistemas que se encuentran funcionando. Se utilizar programacin funcional para el desarrollo de los algoritmos, lo que aporta un complemento a la formacin de grado y a la comunidad acadmica debido a la importancia de este paradigma en la resolucin de problemas de procesamiento *on-line*.

Las pruebas del proyecto se harn en la Secretara de Tecnologas para la Gestin de la Provincia de Santa Fe, lo cual la convierte en potencial beneficiaria de los resultados obtenidos del desarrollo de este proyecto.

Objetivos

Generales

Disear y codificar un sistema de deteccin de anomalas en redes de datos TCP/IP utilizando tcnicas de modelado automtico del comportamiento del trfico de la red.

Especficos

- Identificar un conjunto de tecnologas y herramientas adecuadas para el diseo del sistema.
- Elaborar un documento que describa la arquitectura del sistema.
- Determinar las tcnicas de modelado de trfico de red necesarias.
- Desarrollar un software que modele el trfico normal de red de forma automtica, utilizando tcnicas no supervisadas.
- Implementar un software que genere alertas ante la ocurrencia de eventos anmalos.
- Disear un sistema que minimice la clasificacin de eventos como falsos negativos. Este objetivo es de particular importancia dado que un evento anmalo clasificado como falso negativo es un evento potencialmente daino considerado como normal.

Alcances

Funcionales

- El sistema utilizar informacin de capa de red[9] y capa de transporte[10][11]. Quedan fuera del alcance del proyecto el procesamiento de paquetes IPv6.
- Los datos sern procesados utilizando tcnicas de *data streaming*[12].
- El sistema identificar potenciales comportamientos anmalos y generar las alarmas correspondientes.
- Con cada alarma el sistema mostrar con que probabilidad se clasific el evento como tal.

- El sistema proveer al usuario una interfaz web para la visualizacin de los eventos y las alarmas.
- El sistema almacenar informacin de eventos ocurridos en archivos de texto plano y en formato semi-estructurado.

No Funcionales

- El sistema procesar flujos de datos en tiempo real.
- Se proveer un manual de instalacin y configuracin.

Exclusiones

El proyecto no contempla la instalacin del sistema en un mbito de produccin.

Supuestos

Los datos necesarios para realizar pruebas pueden ser generados automticamente. Adems, se utilizarn datos provistos por el Centro de Cmputos de la Provincia de Santa Fe.

Criterios de Aceptacin

Se considera que el proyecto est aceptado cuando cumple en un 90 % los requisitos funcionales, con un nivel mnimo de aceptacin del 75 %. Los prototipos deben tener implementadas todas las funcionalidades planificadas.

Metodologa

La metodologa de desarrollo del proyecto ser incremental e iterativa. Incremental porque varios componentes y funcionalidades del sistema se desarrollarn en momentos diferentes y sern integradas cuando sean completadas. Iterativa pues se invertirn esfuerzos en revisar constantemente partes del sistema, tanto para mejorar la calidad externa como interna del software[13].

Dado que los requerimientos de los interesados pueden cambiar en el transcurso de la ejecucin del proyecto, se propone utilizar un enfoque de desarrollo gil. El mismo contempla la posibilidad de priorizacin y seleccin en los alcances y en las prioridades de las diferentes funcionalidades, y fundamentalmente promueve la entrega continua de software y la inclusin de los interesados en el proceso de desarrollo.

Plan de Tareas

El proyecto se divide en 5 incrementos. A continuacin se da una breve descripcin de los mismos:

Incremento 1: Investigacin preliminar La primer parte del proyecto consiste en determinar que conjunto de tecnologas sern utilizadas, y elaborar una descripcin a alto nivel de las diferentes componentes del sistema. Adems se realizar una investigacin sobre las tcnicas de modelado de comportamiento existentes con el fin de determinar cuales sern implementadas.

Incremento 2: Modelado de comportamiento En esta etapa se implementarn las tcnicas de modelado de comportamiento seleccionadas en el incremento anterior. Se utilizarn solo los datos provenientes de la capa de transporte y se desarrollar un software con una interfaz web sencilla donde se muestren los principales parmetros del modelado.

Incremento 3: Deteccin de anomalas Se implementarn las funcionalidades de deteccin de anomalas y se agregar al software mecanismos de generacin de alarmas. Adems se mejorar la interfaz de usuario.

Incremento 4: Modelado de comportamiento de subsistemas Con el fin de mejorar la identificacin de anomalas se incorporar al modelo de comportamiento informacin de los diferentes subsistemas que componen la infraestructura de red.

Incremento 5: Pruebas y documentacin Se completar el desarrollo del software y se redactar la documentacin necesaria. Una vez concluidas las pruebas, se elaborar un informe con los resultados obtenidos

Al finalizar la primera iteracin de cada incremento se obtiene una herramienta de software con las funcionalidades descriptas y calidad de producto final, con excepcin de la etapa de investigacin preliminar donde se obtendr un informe en soporte escrito o digital.

Entregable	Fecha de entrega
Informe de avance 1	04/10/2016
Informe de avance 2	08/11/2016
Informe de avance 3	16/12/2016
Informe de avance 4	14/02/2017

Cuadro 1: Fechas de entrega de informes de avance.

Plan de Tareas

La duracin total del proyecto es de 466 horas, con una dedicacin de 20 horas semanales. A continuacin se detalla el plan de tareas.

1. **Investigacin preliminar** (66hs)
   1.1. Estudio comparativo de las tecnologas y mtodos de modelado. (24hs)
   1.2. Diseo conceptual del sistema. (30hs)
   1.3. Documentacin. (12hs)
2. **Modelado de comportamiento** (84hs)
   2.1. Instalacin y configuracin de la plataforma de desarrollo. (12hs)
   2.2. Implementacin de funcionalidad de captura de trfico de red. (24hs)
   2.3. Implementacin de modelado de trfico. (24hs)
   2.4. Implementacin de interfaz web. (24hs)
3. **Deteccin de anomalas** (98hs)

3.1. Implementacin de funcionalidad de deteccin de anomalas. (30hs)

3.2. Pruebas de implementacin. (24hs)

3.3. Implementacin de mdulo de alarmas. (24hs)

3.4. Mejora de interfaz web. (20hs)

4. **Modelado de comportamiento de subsistemas** (146hs)
   4.1. Implementacin de funcionalidad de captura de logs de subsistemas. (72hs)
      - Servidores web (24hs).
      - Gestores de base de datos (24hs).
      - Firewalls (24hs).
   4.2. Implementacin de modelado de comportamiento de subsistemas. (30hs)
   4.3. Pruebas de implementacin. (24hs)
   4.4. Implementacin de funcionalidad de deteccin de anomalas y alarmas. (20hs)
5. **Pruebas y documentacin** (72hs)
   5.1. Pruebas de deteccin. (20hs)
   5.2. Elaboracin de informe de desempeo. (12hs)
   5.3. Elaboracin de informe final. (40hs)

Informes de avance

Se presentarn 4 informes de avance en las fechas de finalizacin de cada etapa, detalladas en el Cuadro 2. A continuacin se detalla que informacin ser incluida en cada informe:

Informe de avance 1 Contendr los resultados obtenidos en los estudios comparativos de las tecnologas y las tcnicas de modelado y deteccin y justificar la eleccin de las mismas. Adems se incluir una descripcin general de la arquitectura del sistema.

Informe de avance 2 Contendr informacin sobre los criterios de seleccin de caractersticas para el modelado de trfico de red. Se proveer la gua de instalacin y configuracin de la plataforma. Adems tendr informacin sobre cambios realizados en los entregables anteriores y el desempeo del modelo implementado.

Informe de avance 3 Contendr informacin sobre los criterios de seleccin de caractersticas para el modelado del comportamiento de los subsistemas, y el desempeo del modelo implementado. Adems se detallarn los cambios realizados en los entregables anteriores.

Informe de avance 4 Se describirn las pruebas de integracin del sistema. Adems tendr informacin sobre cambios realizados en los entregables anteriores y el desempeo general del sistema. Tambin se incluirn los resultados de las pruebas de deteccin de la ltima iteracin.

Etapa	Inicio	Finalizacin	Duracin
Investigacin preliminar	01/09/2016	30/09/2016	4 semanas
Modelado de comportamiento	03/10/2016	04/11/2016	4 semanas
Deteccin de anomalas	07/11/2016	16/12/2016	5 semanas
Modelado de comportamiento de subsistemas	19/12/2016	18/02/2017	8 semanas
Pruebas y documentacin	20/02/2017	25/03/2017	5 semanas

Cuadro 2: Fechas estimativas de inicio y fin de actividades.

Diagrama de Gantt

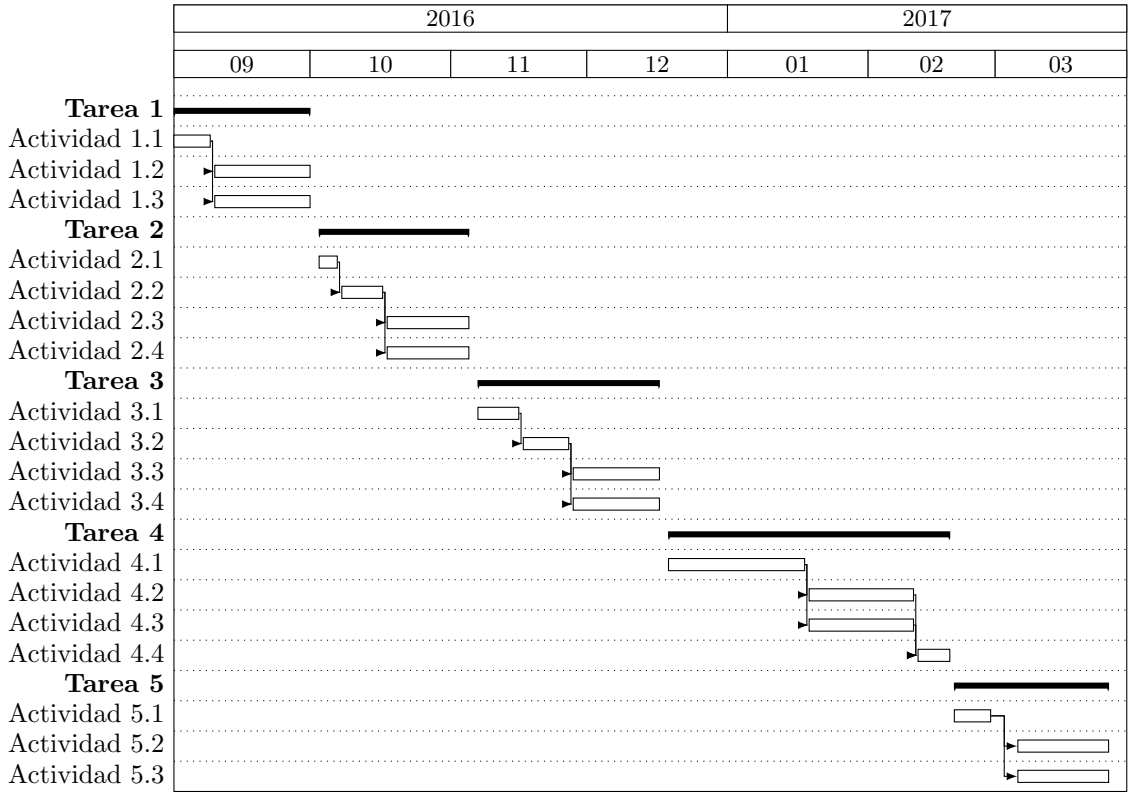

Figura 1: Diagrama de Gantt del proyecto

Riesgos

En esta seccin se enumeran los riesgos identificados, indicadores y estrategia a adoptar segn corresponda. Para realizar el anlisis cualitativo de los riesgos se asigna una probabilidad de ocurrencia y un impacto a cada riesgo. Luego se priorizan segn su *severidad*. Los riesgos identificados son:

1. **No se dispone del hardware necesario:** el servidor de aplicaciones necesario para el desarrollo de la aplicacin no se encuentra disponible

   Indicador: no se puede comenzar con el segundo incremento.
   Probabilidad de ocurrencia: Baja.
   Impacto: Alto.
   Estrategia a adoptar: Mitigar (impacto). Se utilizarn servidores virtualizados.

	1	2	3	4	5
1					
2					
3					
4					
5					

Cuadro 4: Probabilidad/impacto.

Severidad	Estrategia
menor que 4	Aceptar
5 a 15	Mitigar
16 a 25	Evitar

Cuadro 5: Estrategia segn severidad.

2. **Prdida de inters en el proyecto:** dado que el proyecto ser desarrollado para un organismo pblico y las decisiones administrativas estn sujetas a numerosos factores, la entidad solicitante puede dejar de lado el desarrollo del sistema.

   Indicador: el progreso del proyecto se ve interrumpido por falta de comunicacin con los interesados.
   Probabilidad de ocurrencia: Baja.
   Impacto: Alto.
   Estrategia a adoptar: Mitigar (impacto).
   El proyecto ser desarrollado utilizando datos de prueba generados o base de datos existentes.

3. **Las tcnicas de deteccin no pueden ser implementadas:** dada su complejidad, algunas tcnicas de deteccin pueden ser complejas en su implementacin.

   Probabilidad de ocurrencia: Media.
   Impacto: Alta.
   Estrategia a adoptar: Mitigar (probabilidad).
   Se analizarn trabajos previos relacionados y se recurrir a asesoramiento de expertos.

Anlisis Riesgos

En esta seccin se muestra el anlisis realizado de los riesgos del proyecto, ordenados por importancia segn su severidad.

Riesgo	Imp.	% ocur.	Sev.
Las tcnicas de deteccin no pueden ser implementadas.	4	3	12
No se dispone del hardware necesario.	4	2	8
Prdida de inters en el proyecto.	4	2	8

Cuadro 3: Lista de riesgos ordenados por severidad.

Presupuesto

A continuacin se detalla el presupuesto necesario para el desarrollo del proyecto. El servidor de desarrollo cuenta con hardware de red especfico y ser utilizado para ejecutar las pruebas del sistema.

Bienes			
Estacin de trabajo (costo de amortizacin) *Procesador AMD FX6* *Memoria 8GB DDR3* *Monitores (configuracin dual)*			$2691.
Servidor de desarrollo *Procesador Intel i5 o similar* *Memoria 16GB DDR3* *Interfaz de red para servidor (2 puertos)*			$18000.
Impresora (costo de amortizacin)			$318
Servicios			
Servicio de Internet			$4000
Insumos			
Resma de Hojas			$80
Toner			$200
Artculos de librera			$150
Recursos humanos	**Costo por hora**	**Horas**	
Director de proyecto	$150	40	$6000
Codirector de proyecto	$150	40	$6000
Diseador	$110	84	$9420
Desarrollador	$100	364	$36400
Tester	$85	40	$3400
Costo total			$**86659**.

Bibliografía

[1] P. Gill, N. Jain y N. Nagappan, "Understanding network failures in data centers: measurement, analysis, and implications", *SIGCOMM Comput. Commun. Rev.*, vol. 41, n.º 4, págs. 350-361, ago. de 2011, ISSN: 0146-4833. DOI: 10.1145/2043164.2018477. dirección: http://doi.acm.org/10.1145/2043164.2018477.

[2] S. Karumanchi y A. C. Squicciarini, "In the wild: a large scale study of web services vulnerabilities", en *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, ép. SAC '14, Gyeongju, Republic of Korea: ACM, 2014, págs. 1239-1246, ISBN: 978-1-4503-2469-4. DOI: 10.1145/2554850.2555010. dirección: http://doi.acm.org/10.1145/2554850.2555010.

[3] P. Mutchler, A. Doupé, J. Mitchell, C. Kruegel y G. Vigna, "A Large-Scale Study of Mobile Web App Security", en *Proceedings of the Mobile Security Technologies Workshop (MoST)*, mayo de 2015.

[4] "Human errors and violations in computer and information security: the viewpoint of network administrators and security specialists", *Applied Ergonomic*,

[5] S. Kraemer, P. Carayon y J. Clem, "Human and organizational factors in computer and information security: pathways to vulnerabilities", *Computers & Security*, vol. 28, n.º 7, págs. 509-520, 2009, ISSN: 0167-4048. DOI: http://dx.doi.org/10.1016/j.cose.2009.04.006". dirección: http://www.sciencedirect.com/science/article/pii/S0167404809000467.

[6] D. Liginlal, I. Sim y L. Khansa, "How significant is human error as a cause of privacy breaches? an empirical study and a framework for error management", *Computers & Security*, vol. 28, n.º 34, págs. 215 -228, 2009, ISSN: 0167-4048. DOI: http://dx.doi.org/10.1016/j.cose.2008.11.003. dirección: http://www.sciencedirect.com/science/article/pii/S0167404808001181.

[7] M. Ahmed, L. Sharif, M. Kabir y M. Al-maimani, *Human errors in information security*, 2012.

[8] A. Milenkoski, M. Vieira, S. Kounev, A. Avritzer y B. D. Payne, "Evaluating computer intrusion detection systems: a survey of common practices", *ACM Comput. Surv.*, vol. 48, n.º 1, 12:1-12:41, sep. de 2015, ISSN: 0360-0300. DOI: 10.1145/2808691. dirección: http://doi.acm.org/10.1145/2808691.

[9] J. Postel, *Internet Protocol*, RFC 791 (INTERNET STANDARD), Updated by RFCs 1349, 2474, 6864, Internet Engineering Task Force, sep. de 1981. dirección: http://www.ietf.org/rfc/rfc791.txt.

[10] ——, *Transmission Control Protocol*, RFC 793 (INTERNET STANDARD), Updated by RFCs 1122, 3168, 6093, 6528, Internet Engineering Task Force, sep. de 1981. dirección: http://www.ietf.org/rfc/rfc793.txt.

[11] T. Socolofsky y C. Kale, *TCP/IP tutorial*, RFC 1180 (Informational), Internet Engineering Task Force, ene. de 1991. dirección: `http://www.ietf.org/rfc/rfc1180.txt`.

[12] F. Fischer, F. Mansmann y D. A. Keim, "Real-time visual analytics for event data streams", en *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, ép. SAC '12, Trento, Italy: ACM, 2012, págs. 801-806, ISBN: 978-1-4503-0857-1. DOI: `10.1145/2245276.2245432`. dirección: `http://doi.acm.org/10.1145/2245276.2245432`.

[13] ISO/IEC, *ISO/IEC 9126. Software engineering – Product quality*. ISO/IEC, 2001.

[14] M. Xie, S. Han, B. Tian y S. Parvin, "Anomaly detection in wireless sensor networks: a survey", *J. Netw. Comput. Appl.*, vol. 34, n.º 4, págs. 1302-1325, jul. de 2011, ISSN: 1084-8045. DOI: `10.1016/j.jnca.2011.03.004`. dirección: `http://dx.doi.org/10.1016/j.jnca.2011.03.004`.

[15] A. Tanenbaum, *Computer Networks*, 4th. Prentice Hall Professional Technical Reference, 2002, ISBN: 0130661023.

[16] W. Stallings, *Data and Computer Communications (5th Ed.)* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997, ISBN: 0-02-415425-3.