Detección de Intrusiones mediante el uso de Redes Neuronales

Britos J. Daniel[12], Arias. Silvia[13], Vargas Laura[14]

1. Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de Córdoba. Laboratorio de Redes y Comunicaciones (LARyC).
2. Departamento Universitario de Informática. Universidad Nacional de Córdoba.
3. Departamento Tecnologías de Información. Instituto Universitario Aeronáutico.
4. Facultad Regional Córdoba. Universidad Tecnológica Nacional.

dbritos@gmail.com, edith.edit@gmail.com, lauramonicavargas@gmail.com

Resumen:
Con el crecimiento explosivo de Internet y particularmente de las aplicaciones de comercio electrónico, los ataques a las redes se han vuelto más comunes y sofisticados. Las redes demandan medidas de protección más elaboradas para garantizar que éstas operen seguras y dar continuidad a los servicios críticos, estas medidas incluyen métodos de detección y repuesta en tiempo real a los intentos de intrusión. Este trabajo usa modelos estadísticos y clasificadores multivariables para detectar perfiles de tráfico anómalos, utilizando redes neuronales. El análisis estadístico basa su cálculo en el álgebra de las funciones de densidad de probabilidad (PDF). La red neuronal integra esta información en una sola salida reportando el estado de la red, la que alimenta un software que configura al firewall, para producir un rechazo de la amenaza en tiempo real.

Palabras claves: Detección de Intrusiones, Seguridad Integral, Alerta Temprana, Inundación de Paquetes, Denegación de Servicios, Redes de Datos.

Contexto

El presente trabajo de investigación se ocupa de los sistemas IDS del tipo NID, está avalada y subsidiado por la SECyT – UNC 05/M067.

Estado del Arte

Los intentos de vulnerar los sistemas de computadoras crecen día a día y el uso de métodos avanzados de encriptación no es suficiente para proteger los sistemas informáticos. Esto hace que sea necesario proveer a las redes con sistemas de protección bien planeados y políticas integrales de defensa contra los ataques.
Para ese rol juegan un papel importante los sistemas de alerta temprana o sistemas de detección de intrusiones (IDS Intrusion Detection Systems), como amplia contramedida de defensa. Un antecedente en este tema es el trabajo de Papavassiliou [1] que propone el método de detección de intrusiones estadística, usando como herramienta la estadística de Kolmogorov-Smirnov y redes neuronales para modelar y detectar ataques.

Introducción

Ataques a Computadoras
Existen básicamente dos tipos de ellos:

- Ataques Pasivos
- Ataques Activos

Ataques Pasivos

Estos son los de escucha sin autorización o de monitoreo de tráfico. Los objetivos de estos ataques consisten en obtener la mayor cantidad de información del mensaje transmitido y del oponente. Las distintas modalidades de ataques pasivos son las siguientes:

Descarga de contenidos del mensaje: Están incluidos dentro de este tipo de ataque la escucha de una conversación telefónica, la lectura de un mensaje de correo electrónico o la información confidencial capturada por un oponente.

Análisis de tráfico: Este es un ataque muy sutil. Se supone que hay medios de envíos de mensajes confidenciales, que no permiten al atacante poder acceder al contenido del mensaje. El atacante tiene sólo la posibilidad de observar la transmisión de los mensajes y obtener de éstos, por ejemplo datos tales como: la frecuencia de emisión de los mensajes y la longitud del mensaje. Esta información puede ser de mucha ayuda para inferir la naturaleza de la comunicación.

Los ataques pasivos son muy difíciles de detectar y reconocer, porque ellos son un medio de reconocimiento previo a la realización de ataques activos.

Ataques Activos

Los ataques activos involucran y comprometen los pilares básicos de las prácticas de seguridad: la confidencialidad, la integridad y la disponibilidad (CIA Confidentiality, Integrity and Availability). Los ataques activos son:

Denegación de Servicio: El efecto de este ataque es impedir la posibilidad de acceso a toda persona a un determinado servidor.

Enmascarado (“Masquerade"): En este caso el atacante se representa él mismo como un legítimo usuario con el objeto de robar, alterar o destruir recursos informáticos.

Reinterpretar (“Replay"): Este ataque es llevado a cabo mediante una captura pasiva de datos, para que luego sean retransmitidos y con ello producir efectos no autorizados.

Modificación de contenidos del mensaje: La información original es alterada de tal forma que permita obtener un resultado no autorizado.

Sistema de detección de intrusiones

La detección de intrusión en redes IDS es un componente vital en la defensa contra ataques a redes y es abordado desde diferentes perspectivas como puede verse en el trabajo de Bai[2]. En él se implementan dos métodos principales de detección: La detección de intrusiones de redes NID (Network Intrusion Detection), y la detección de intrusiones de servidores HID (Host Intrusion Detection).

Detección de intrusiones de redes

Los sistemas NID están relacionados con el tráfico de información entre servidores y clientes. Típicamente referidos como espías de paquetes (packet-sniffers), estos dispositivos interceptan paquetes que viajan por los medios de comunicación y transportan datos encapsulados en diferentes protocolos tales como, “Frame Relay" o ATM (Modo de Transferencia Asincrónica, Asynchronous Transfer Mode). Algunos dispositivos NID comparan el paquete con una base de datos de firmas de ataques conocidos y huellas digitales de paquetes maliciosos, mientras que otros analizan la actividad de paquetes buscando un comportamiento anómalo que pueda ser malicioso. En cualquiera de ambos casos un dispositivo IDS debe ser visto principalmente como una defensa perimetral de la red.

Los dispositivos NID en el pasado y por la complejidad de la tarea que realizan, han sido incapaces de operar en los siguiente ambientes:

- Redes conmutadas
- Redes encriptadas
- Redes de alta velocidad

Recientemente esta limitación ha sido superada por la potencia de procesamiento y cálculo de los microprocesadores modernos. Los conmutadores de redes ya vienen equipados con dispositivos IDS, capaces inclusive de realizar “packetsniffing" (espías de paquetes) a velocidades de gigabit. Las técnicas de NID pueden desagregarse en dos tendencias principales y complementarias entre sí: Detección de mal uso y Detección de anomalías.
La detección de mal uso como lo explica Vigna [3] modela ataques conocidos y realiza búsquedas de la ocurrencia de esos patrones.
Los sistemas de detección de anomalías, como lo señala Valdes [4] alertan de intrusiones mediante la observación de las desviaciones del comportamiento típico del tráfico de la red.

Detección de intrusiones de servidores
La detección de intrusiones basada en servidores, está diseñada para responder a ataques sobre un determinado servidor. Se basan en la supervisión de las acciones de los usuarios y de los archivos del servidor. En auditoria de los registros de actividad de los servidores y estado de los archivos del sistema, existen técnicas robustas que ofrecen administración de políticas de auditoria, análisis estadístico y soporte de evidencias, los que proveen medidas de control de la actividad de los servidores. La detección de intrusión en servidores sirve tanto para detectar ataques externos como internos.

Materiales y Métodos

En el análisis de la información se utilizan programas específicos y herramientas GNU. Los programas se desarrollaron en lenguaje C, en un ambiente Linux, bajo normas(CMM/ ISO/IEC 12207), para asegurar calidad y capacitar en las prácticas adecuadas de ingeniería de software. Se utilizan simuladores de redes como el ns2 de la universidad de Berkeley. Para la simulación de las redes neuronales se utilizarán programas específicos y herramientas tales como Stuttgart Neural Network Simulator. El sistema de detección de intrusiones propuesto en el siguiente trabajo de investigación utiliza modelos estadísticos y una red neuronal clasificadora tipo, para detectar las condiciones anómalas de la red. El análisis estadístico basa su cálculo en el álgebra de las funciones de densidad de probabilidad (pdf), en lugar de usar muestras aisladas o sus promedios solamente. Éstas funciones pdf serán procesadas estadísticamente antes de ser ingresados a la red neuronal. la red neuronal integrará esta información en una sola salida reportando el estado de la red. En el presente trabajo se propone la implementación y las pruebas de funcionamiento en un ambiente real, utilizando el firewall (Pared cortafuegos) de la Universidad Nacional de Córdoba.

Resultados

Síntesis de Diseño e Implementación:

Módulo de captura: Este módulo recolecta datos del comportamiento de la red, midiendo seis parámetros: Tráfico de Protocolo Internet (IP Internet Protocol) en bps (bits por segundo), Tráfico IP en pps (paquetes por segundo), Tráfico de Protocolo de Datagrama de

Usuario (UDP, User Datagram Protocol) en bps, Tráfico UDP en pps, longitud de paquetes IP en bytes y longitud de paquetes UDP en bytes.

Preprocesador de eventos: Recibe reportes del módulo de captura y calcula la función de distribución de probabilidad (PDF) para los seis parámetros en estudio.

Procesador estadístico: Mantiene un modelo de referencia de la actividad normal de red, lo compara con los reportes del preprocesador de eventos para cada uno de los seis parámetros y forma un vector estímulo que combina todos ellos para alimentar la red neuronal.

Red neuronal clasificadora de ataques: Procesa el vector estímulo del procesador estadístico para decidir si el estado de la red es normal o se encuentra bajo ataque.

Visualizador de estado: Obtiene la información de salida de la red neuronal y la representa gráficamente en una curva en función del tiempo. Esta curva podrá estar disponible a través de una página web.

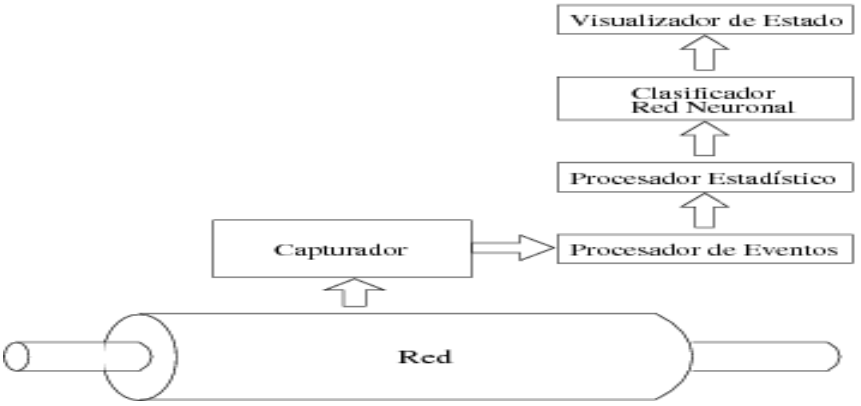

Diagrama de Bloques del IDS

Todo el sistema completo genera una salida indicando el nivel de ataque con una variable real entre los valores 0 y 1, donde el valor 1 representa una intrusión con absoluta certeza y una salida con valor 0 indica normalidad. Esta salida es actualizada con una frecuencia determinada.

Obtención de los parámetros de entrada y salida
Los patrones utilizados en el entrenamiento es un conjunto de entradas salidas conocidas. Este conjunto esta integrado tanto de patrones de tráfico normal de la red como de patrones de tráfico bajo condición de ataque. Para obtener los patrones de entrada de la red neuronal (vector k-dimensional cuyos componentes son los valores de similitud) se utiliza un programa que incluye solamente el bloque de captura. Con las muestras de entrada, se procedió al entrenamiento de la red propiamente dicha. Para ello se utilizó un programa que lee los datos del archivo de las muestras de la red y las procesa estadísticamente, generando los patrones de entrada para la red neuronal. Los patrones de salida, son conocidos, mediante el registro de los períodos de tiempo en los cuales se realizaron los ataques. Se asignó 1 para condición de tráfico normal y -1 para la condición de ataque. Una vez obtenidos los patrones de entrada y salida, se realizó el entrenamiento de la red. Los pesos

calculados de la red se almacenan en un archivo, que se utilizará en el próximo paso. El error de aprendizaje resultante fue de 0,56 %. Este error es el porcentaje de salidas de la red neuronal con falsos positivos y falsos negativos en relación al total de salidas. Se observa la convergencia del error cuadrático medio de la red neuronal, a medida que transcurren las iteraciones. Es importante notar que el error cuadrático medio disminuye muy rápidamente luego de sólo algunas iteraciones, alcanzando niveles de convergencia satisfactorios dentro de las 15 iteraciones. Estas características son especialmente deseables para sistemas de detección de intrusiones en redes (NID), los cuales necesitan supervisión en tiempo real y entrenamiento en línea. Una vez realizado el entrenamiento, utilizamos el segundo conjunto de muestras de tráfico de red y un programa para verificar el correcto aprendizaje de la red neuronal y analizar su generalización. El error de generalización obtenido fue del 1, 97 %.

Conclusiones

En primer lugar, se vio que la red neuronal puede entrenarse para distinguir entre condiciones normales y en condiciones de ataque de la red. Se obtuvo un error de aprendizaje de 0,56 % y un error de generalización de 1,97 %, utilizando una configuración sencilla de la red neuronal (configuración back-propagation con sólo dos neuronas en la capa oculta). Este resultado se ha logrado por el uso del módulo procesador estadístico que colabora en forma eficiente en la detección de intrusiones.

Como se puede advertir, para la detección de inundaciones UDP, el aporte realizado por los parámetros relacionados con el protocolo IP es pequeño en relación al aporte de los parámetros del protocolo UDP al momento de llevar a cabo la detección. Por esta razón, se podría omitir la utilización de los parámetros del protocolo IP.

Trabajos Futuros

El sistema se podría adaptar para detectar otros tipos de intrusiones, tales como inundaciones TCP-SYN e ICMP (Internet Control Message Protocol, Protocolo Internet de Mensajes de Control). En el caso de inundaciones TCP-SYN, se debería incluir el uso de parámetros relacionados al protocolo TCP mientras que las inundaciones de paquetes ICMP se detectarían a través de los parámetros del protocolo IP. Todo ello sería factible siempre y cuando la red neuronal pueda aprender los distintos tipos de ataque.

Se debe tener en cuenta que en este caso, no tenemos control sobre el tráfico de fondo de la red. Por el contrario sí se tuviese, a través de programas de simulación y de modelado de tráfico de red, se podría analizar el rendimiento del sistema para distintos tráficos de fondo y distintos niveles de ataque para mejorar su desempeño.

Con el sistema IDS implementado, se detectaron ataques en la red, se dejó para una etapa posterior la elaboración de ataques más complejos y elaborar acciones defensivas a través de la construcción automática de reglas al firewall.

Publicaciones

"Statistical Intrusion Detection in Data Networks". IEE Latin America Transactions. Page(s): 373-380. Digital Object Identifier:10.1109/TLA.2007.4378531

Formación de Recursos Humanos

Obtención del grado de Maestría en Ciencias de la Ingeniería Mención Telecomunicaciones. En la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de Córdoba., por parte de uno de los integrantes del equipo de investigación.

Agradecimientos

A la Secretaría de Ciencia y Técnica de la Universidad Nacional de Córdoba que avala el proyecto.

Bibliografía

[1] C. Manikopoulos and S. Papavassiliou, “Network intrusion and fault detection: a statistical anomaly approach", Communications Magazine, vol. 40, pp. 76-82, Oct. 2002.

[2] Y. Bai and H. Kobayashi, “Intrusion detection systems: technology and development", Advanced Information Networking and Applications, 2003. AINA 2003.17th International Conference on, vol. Issue, 27-29, pp. 710 - 715, March 2003.

[3] G. Vigna and A. Kemmerer, “Netstat: a network-based intrusion detection approach", in Computer Security Applications Conference, vol. 1, Dec. 1998, pp. 25-34.

[4] A. Valdes and D. Anderson, “Statistical methods for computer usage anomaly detection using nides”, Tech. rep.: SRI International, Jan 1995.

[5] Y.Ñong et al., “Statistical process control for computer intrusion detection", DARPA Information Survivability Conference Exposition II, 2001. DISCEX '01. Proceedings, vol. Volume 1, pp. 3 - 14, June 2001.

[6] C. M. Bishop, Neural Networks for Pattern Recognition. New York, NY: Oxford Univ. Press, 1995.

[7] P. Ramasubramanian and A. Kannan, “Intelligent multi-agent based backpropagation neural network forecasting model for statistical database anomaly prevention system”, Intelligent Sensing and Information Processing, 2004. Proceedings of International Conference on, pp. 108 - 113, 2004.

[8] S. C. Lee and D. Heinbuch, “Training a neural-network based intrusion detector to recognize novel attacks", Systems, Man and Cybernetics, Part A, IEEE Transactions on, vol. Volume 31, pp. 294 - 299, July 2001.