IDS basado en mapas autoorganizados

IDS Based on Self-Organizing Maps

P. Cortada, G. Sanromà, P. García et al.

Resumen

Se presentan los resultados obtenidos con un nuevo IDS basado en el uso de redes neuronales. La aplicación desarrollada es capaz de clasificar en tiempo real el tráfico que llega a la interfaz de red del host que está protegiendo, detectando patrones de tráfico anómalos. Los únicos datos necesarios para el análisis del tráfico son las cabeceras de los paquetes.

Palabras clave: IDS, redes neuronales, SOM

Summary

The results obtained using a new IDS system based on the use of artificial neural networks are presented. The application developed is able to classify in real time the traffic that arrives at the network interface of the host that it is protecting, detecting anomalous traffic patterns. The only necessary data for the traffic analysis are the headers of the incoming packets.

Keywords: IDS, Neural Networks, SOM

Las redes neuronales son modelos matemáticos que intentan emular de una forma muy simplificada el funcionamiento de las neuronas en el cerebro humano

1.- Antecedentes y motivación

Cada vez se producen con mayor frecuencia ataques a los sistemas informáticos debido principalmente al auge experimentado por las redes de comunicaciones. Los efectos que pueden producir estos ataques son a menudo impredecibles y pueden abarcar aspectos tan diversos como la confidencialidad, la integridad, la disponibilidad o el control de un sistema informático. Las técnicas empleadas para realizar estos ataques son muy variadas y se caracterizan por una constante y rápida evolución.

Un elemento clave para minimizar sus efectos será intentar detectarlos inmediatamente. Con esta finalidad nace toda una serie de técnicas y aplicaciones denominadas genéricamente *Sistemas de Detección de Intrusiones* (IDS). Una de las líneas de investigación más activas hoy día en el marco de este tipo de sistemas es la que tiene por objetivo dotar de inteligencia y adaptabilidad a los IDS, con la finalidad de que sean capaces de adaptarse a la rápida dinámica que rige la evolución de las técnicas de ataque.

En este trabajo presentamos los resultados preliminares de un proyecto piloto cuyo objetivo es el uso de redes neuronales para proporcionar al IDS un sistema inteligente de análisis de datos.

2.- Clasificadores neuronales: Mapas autoorganizados

Las redes neuronales son modelos matemáticos que intentan emular de una forma muy simplificada el funcionamiento de las neuronas en el cerebro humano. Éste es un campo de investigación en constante evolución y que está dando origen a numerosas aplicaciones. En general, podemos clasificar las redes neuronales en dos grandes grupos en función de la estrategia empleada en el aprendizaje: sistemas supervisados y sistemas no supervisados. Los primeros necesitan de un "maestro" que les indique la relación causa-efecto a aprender. En cambio los segundos actúan como clasificadores, agrupando los datos de entrada (denominados patrones) que presentan características similares. Entre los clasificadores no supervisados uno de modelos más extendidos son los mapas autoorganizados (SOM) [Ko90].

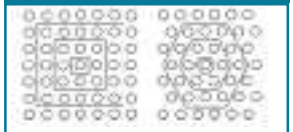
El objetivo que se ha fijado para esta fase preliminar de estudio es explorar las posibilidades que ofrece el uso de un mapa autoorganizado como núcleo del sistema de análisis de un IDS

El algoritmo del SOM realiza la transformación de un espacio de entrada de dimensión alta a otro de dimensión más baja (usualmente bidimensional o tridimensional) formado por una red regular de nodos. El proceso se realiza preservando la topología inicial del espacio de entrada.

Cada nodo del mapa tiene asociado un vector de pesos $m1=[m_{11},...,m_{in}]^t$, donde n representa la dimensión del espacio de entrada. Sobre cada una de estas unidades se define una vecindad N_i que estará determinada por la forma de la red (regular, hexagonal, toroidal).

El número de unidades que forman la red así como su topología se fijan antes de empezar el proceso de entrenamiento. Durante el proceso de aprendizaje el SOM se comporta como una red elástica, deformándose para adaptarse a la estructura de los datos que se le presentan. De esta manera la red aproxima la función de densidad de probabilidad de los datos iniciales, situando un mayor número de nodos donde el espacio de entrada es más denso y menos nodos donde los datos están más dispersos.

TOPOLOGÍA DE MAPA AUTOORGANIZADO



a) retículo regular b) retículo hexagonal

El algoritmo de aprendizaje itera sobre el conjunto de datos usado para entrenar la red hasta que se alcanza un valor prefijado en el error de clasificación. En cada iteración se escoge aleatoriamente un patrón x del conjunto de entrenamiento. Se mide la similitud entre este patrón y todos los vectores de pesos asociados a las unidades que forman la red usando la distancia euclídea. Como resultado se obtiene la unidad más similar al patrón de entrada, denominada *best matching unit* (BMU). Una vez detectada la BMU, los pesos de ésta y de sus unidades vecinas se adaptan, acercándolos a la posición del patrón de entrada. Este proceso de adaptación puede formalizarse como:

$$m_i(t+1) = \begin{cases} m_i(t) + \alpha(t)\left[x(t) - m_i(t)\right]; & i \in N_{bmu}(t) \\ m_i(t); & i \notin N_{bmu}(t) \end{cases}$$

donde $\partial(t)$ es un parámetro que controla la velocidad del proceso de aprendizaje y N_{bmu} representa la vecindad topológica de la BMU. Los nodos que quedan fuera de la vecindad topológica de la BMU quedan inalterados.

3.- Arquitectura del IDS neuronal

El objetivo que se ha fijado para esta fase preliminar de estudio es explorar las posibilidades que ofrece el uso de un *mapa autoorganizado* como núcleo del sistema de análisis de un IDS. Para ello diseñaremos e implementaremos un sistema IDS basado en host que clasificará en tiempo real cada uno de los paquetes que llegan al interfaz de red de la máquina a proteger. Las hipótesis en las que se ha basado el diseño del IDS son:

- El IDS actúa como un **clasificador del tráfico entrante** en la máquina que está protegiendo.
- **Patrones de tráfico similares se asignarán a un mismo cluster**, dado que el mapa preserva las relaciones topológicas del espació de entrada.
- Al entrenar el mapa con tráfico "normal", **el tráfico anómalo quedará fuera de los clusters** que se hayan formado, siendo por tanto fácilmente detectable.

El propósito del IDS neuronal es la detección, sobre la base del análisis de los datos de la cabecera del paquete, de ataques cuya finalidad sea el escaneo de puertos o la denegación de servicios. No se analizará el contenido de los paquetes, y por tanto el IDS neuronal no detectará los ataques que puedan realizarse a nivel de aplicación.

El propósito del IDS neuronal es la detección de ataques cuya finalidad sea el escaneo de puertos o la denegación de servicios

Para la captura del tráfico se ha desarrollado una aplicación basada en la librería libpcap (http://www.tcpdump.org). Los datos de tráfico capturados se separan en diversos ficheros en función del protocolo. A los datos que nos proporciona la cabecera IP añadiremos los campos adicionales que nos proporcionan las cabeceras de los protocolos de transporte. El hecho de separar tráficos correspondientes a protocolos distintos facilitará la construcción de mapas específicos para cada uno de ellos, obteniéndose mejores resultados en la clasificación. Las características escogidas para el análisis del tráfico son las que aparecen en la tabla.

PROTOCOLO	DESCRIPTORES SELECCIONADOS	CARACTERÍSTICAS COMUNES
IP	versión long. cabecera ToS, long paquete id. datagrama NF, MF ttl dir. origen dir. destino	promedio bytes/sec. durante el último segundo promedio de pkts/sec. durante el último segundo promedio de bytes/sec. durante los últimos N segundos promedio de pkts/sec. durante los últimos N segundos
TCP	puerto origen puerto destino nº seq nº. ACK flags wsize nº paq. SYN último segundo nº paq. SYN últimos N segundos	
UDP	puerto origen puerto destino long. paquete	
ICMP	tipo código	

Tabla: Características para el análisis del tráfico en función del protocolo

La parte fija de datos para cada paquete analizado estará formada por la cabecera IP y una serie de variables estadísticas que tienen en cuenta la evolución del tráfico en una cierta ventana de tiempo. A esta información fija se añadirán los descriptores propios de cada protocolo.

El IDS implementado se ha denominado *Sistema Automático para la Detección de Actividades Maliciosas (SADAM)*, y funcionalmente podemos representarlo como aparece en la figura.

El proceso de operación del IDS consta de dos fases:

Al usar un modelo de aprendizaje no supervisado carecemos de información *a priori* sobre los valores que se están clasificando

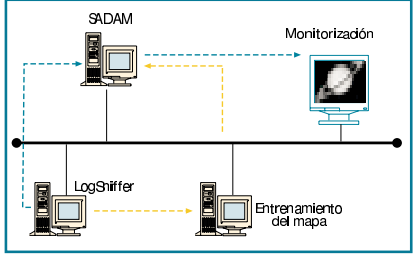
1.- Entrenamiento del sistema (off-line)

- Captura de datos para entrenamiento
- Preprocesado y normalización
- Entrenamiento y adaptación del mapa

2.- Análisis del tráfico (on-line)

- Captura y preprocesado de tráfico en tiempo real
- Visualización del tráfico.

FIGURA 1: SADAM



Al usar un modelo de aprendizaje no supervisado carecemos de información *a priori* sobre los valores que se están clasificando. Por tanto, para la correcta interpretación de los mapas será imprescindible un proceso de etiquetado con la finalidad de detectar el tipo de tráfico asociado a cada cluster.

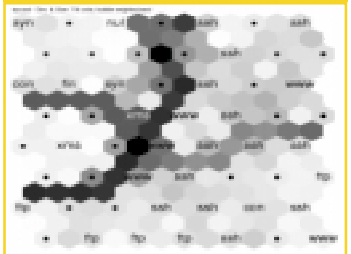
4.- Resultados obtenidos

Uno de los problemas fundamentales que implica el uso de mapas autoorganizados es la interpretación del mapa una vez entrenado. Habitualmente empleamos una representación gráfica que nos proporcione una idea precisa de su estado. Existen diversas técnicas para la visualización del estado del mapa. Las más usadas son: los planos de componentes (c-planes) y las matrices de distancias (u-matrix) [Ults90].

Con los primeros, representamos sobre el espacio del mapa los valores correspondientes a cada una de las variables de entrada mediante un código de color basado en su valor numérico. Esta visualización nos permite saber en todo momento en qué estado se encuentra el sistema que estamos monitorizando, relacionándolo con cada una de las variables que forman el patrón de entrada.

Las matrices de distancia las obtenemos al calcular las distancias entre cada nodo del mapa y los nodos de su vecindad topológica. Como resultado de este proceso obtenemos una visualización como la mostrada en la figura 2, que nos permite detectar los clusters que se forman sobre el mapa ya que sus fronteras quedan delimitadas por valores de distancia altos (líneas oscuras).

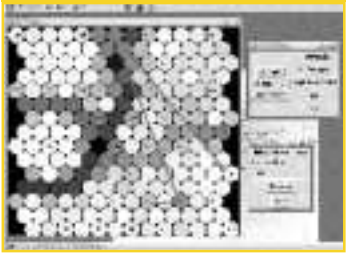
FIGURA 2: MATRIZ DE DISTANCIAS



La figura 3 muestra la herramienta de monitorización desarrollada para SADAM. Esta aplicación permite superponer las trayectorias que sigue el estado del sistema sobre una matriz de distancias etiquetada. De esta forma vemos en tiempo real cómo se clasifica cada uno de los paquetes que llegan a la interfaz de red y tenemos una idea de cuál ha sido su evolución temporal.

En el ejemplo de la figura 3 se está realizando una conexión FTP y previamente se había establecido una sesión SSH.

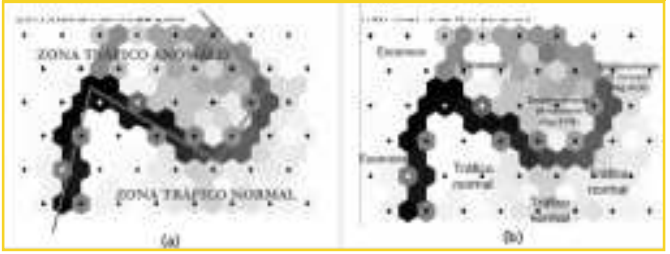
FIGURA 3: HERRAMIENTA DE MONITORIZACIÓN DE SADAM



Los mapas autoorganizados son capaces de aprender las características del tráfico de red

Los experimentos realizados con tráfico real han permitido caracterizar automáticamente el tráfico entrante en el sistema. Entre los experimentos realizados es importante destacar los resultados obtenidos en la caracterización de patrones de tráfico considerados anómalos. Mediante el entrenamiento y posterior etiquetado del mapa con tráfico normal y tráfico resultante de diversos ataques de tipo escaneo de puertos y DoS, obtenemos una representación sobre la matriz de distancias del mapa que muestra claramente la frontera que separa ambos comportamientos.

FIGURA 4: (a) Separación entre patrones de tráfico anómalos y tráfico considerado normal. (b) Etiquetaje del mapa mostrando las características principales de cada tipo de tráfico



En la figura 4a vemos en la parte derecha el cluster correspondiente al tráfico normal. En cambio en la parte izquierda se sitúa el tráfico anómalo correspondiente a ataques realizados con herramientas de escaneo (como NMAP). En la figura 4b se aprecia que en la zona cercana a la frontera se sitúan los paquetes correspondientes a establecimiento de conexión y que inicialmente pueden ser atribuidos tanto a tráfico normal como anómalo.

5.- Conclusiones y vías de futuro

Las redes neuronales, y en especial los mapas autoorganizados constituyen una alternativa válida y eficaz para su uso como sistema de análisis en un IDS. Los mapas autoorganizados son capaces de aprender las características del tráfico de red y por tanto son capaces de diferenciar entre las condiciones de tráfico normal y las de tráfico considerado como anómalo. Una vez entrenados son capaces de clasificar el tráfico en tiempo real e incluso pueden ser implementados en hardware para operar en situaciones que requieran una gran velocidad de análisis (por ejemplo, redes Gigabit Ethernet)

Respecto a las vías de futuro que se plantean a partir del presente trabajo, podemos destacar:

- Aplicación de las mismas técnicas a la caracterización del régimen de operación de un sistema operativo. De esta manera obtendríamos un IDS capaz de detectar estados anómalos en la operación de un ordenador y que podrían considerarse como indicios de una intrusión.
- Adaptar la arquitectura del IDS con el objetivo de obtener un sistema totalmente distribuido en el que agentes IDS especializados intercambien información y realicen diagnósticos conjuntos, tanto en la red como en los hosts.
- Generación de respuestas activas, que permitan en función de las características calculadas para el tráfico anómalo, generar reglas de filtraje en un firewall.

Todo ello pasa previamente por un análisis y una validación a fondo del sistema actual. Actualmente se está realizando esta validación a partir de la base de datos del MIT Lincoln Laboratory [Hai99] que contiene alrededor de 4Gb de información correspondiente a ataques simulados en una red militar.

6.- Referencias

[Ko90] Kohonen, T. "The Self-Organizing Map" Proceeding of IEEE, 78(9), 1464 (1990).

[Ults90] Ultsch, A., Siemon, H. "Kohonen's Self Organizing feature maps for exploratory data analysis" Proceedings of INNC'90. International Neural Networks Conference, 305 (1990).

[Hai99] Haines, J.W., Lippmann, R.P., Fried, D., Tran, E., Boswell, S. and Zissman, M. "1999 DARPA Intrusion Detection System Evaluation: Design and Procedures", MIT Lincoln Laboratory Technical Report, (1999). (http://www.ll.mit.edu/IST/ideval/index.html)

Pere Cortada Bonjoch
(pcb.si@estudiants.urv.es)
Gerard Sanromà Güell
(gsg.is@estudiants.urv.es)
Pedro García López
(pgarcia@etse.urv.es)
Àlex Arenas Moreno
(aarenas@etse.urv.es)
Robert Rallo Moya
(Robert.Rallo@etse.urv.es)
Departament d'Enginyeria Informàtica I Matemàtiques
Escola Tècnica Superior d'Enginyeria - URV