INTRODUCCIÓN A LAS REDES NEURONALES EN LA DETECCIÓN DE INTRUSOS[1]

Angela Diez Diez, Francisco J. Rodríguez Sedano

Departamento de Ingeniería Eléctrica y Electrónica, Universidad de León
E.II.I.I., Campus de Vegazana, s/n. 24071. León (España)
E–mail: dieadd@unileon.es y diefrs@unileon.es

Contenido

1. INTRODUCCIÓN

En este trabajo se pretende analizar distintos sistemas de detección de intrusos basados en Redes Neuronales tanto para detección de anomalías como para detección de mal uso (*misuse*). La característica que se aprovecha es la de aprendizaje de estas redes, lo que permite predecir acciones de usuario. Por tanto, su consideración es lícita como una alternativa a los sistemas basados en reglas.

Para ello en este artículo se diferencian dos apartados: una primera, donde se definen ciertos términos de los Sistemas de Detección de Intrusos (SDI) y una segunda parte, donde se analizan diversos modelos de detección en los que se aplican las Redes Neuronales. Uno de los objetivos principales a tener en cuenta, serán la disminución de los falsos positivos y negativos y por tanto un aumento del porcentaje de aciertos en la detección tanto de mal uso como de anomalías.

[1] Este trabajo está soportado por el proyecto de investigación DPI 2001–0105 del MCT.

Las tareas que podemos destacar dentro de un SDI son: Recopilación de información (fuente de datos), Reducción de datos, Análisis de comportamiento, Información y respuesta.

La recopilación de información se puede clasificar por su localización y así definimos dos categorías iniciales host y red, pero de una forma más general podemos definir cuatro categorías: host, red, aplicación y objetivo. De esta forma podemos tener:

- monitores basados en máquina (recogen datos generados de un ordenador),
- monitores basados en red (recogen paquetes de la red)
- monitores basados en aplicación (registran la actividad de una aplicación)
- monitores basados en objetivos (generan sus propios registros, usan funciones para detectar alteraciones de sus objetivos, y contrastan los resultados con las políticas, se utiliza en elementos que no pueden ser monitorizados de otra forma)
- monitores híbridos si combinamos varias fuentes.

Después del proceso de recopilación de información, se lleva a cabo el proceso de análisis para la detección de ataques. En muchos casos la información se ordena cronológicamente, se clasifica, se evalúa de forma estadística en muchos casos o por otras técnicas, se reduce y se identifica mediante patrones o firmas de actividad relativos a distintos aspectos de seguridad.

Según el objetivo del motor de análisis podemos definir dos tipos principales de análisis:

- Detección de mal uso ("*misuse*"), es la técnica usada por la mayoría de sistemas comerciales, y analizan la actividad del sistema buscando eventos que coincidan con un patrón predefinido o firma que describe el ataque, por lo tanto se comparan firmas con la información recogida en busca de coincidencias. Se lleva a cabo a partir de modelos de ataque bien definidos, que utilizan fallos conocidos del sistema. Es una detección directa.
- Detección de anomalías o de comportamiento anómalo, en la que el análisis busca patrones anormales de actividad, es decir, se centra en identificar comportamientos inusuales, suelen construir perfiles que

representan el comportamiento normal, mediante la recogida de datos en periodos de operación normal. donde se manejan técnicas estadísticas que definen de forma aproximada lo que es el comportamiento usual o normal, es decir, cuantifican el comportamiento normal del sistema, las técnicas empleadas incluyen: detección de umbrales, medidas estadísticas o empleo de otras medidas que incluyen redes neuronales, algoritmos genéticos, y modelos de sistema inmune...

Las acciones que realiza el motor de análisis son: preproceso, clasificación y posproceso de la información.

En el apartado de respuesta, podemos definir dos tipos principales de respuesta: La respuestas pasivas (no se toman acciones que puedan cambiar el curso de un ataque, se limita a enviar o registrar la alarma correspondiente al responsable) y las respuestas activas (generan la alarma y reaccionan modificando el entorno).

Una de las limitaciones de los IDS tradicionales es la incapacidad de reconocer ataques ligeramente modificados respecto a los patrones con los que se realiza la comparación. Eso hace necesario el empleo de métodos alternativos como pueden ser las redes neuronales, que permiten acciones de aprendizaje lo que genera una mayor adaptabilidad.

2. LAS REDES NEURONALES

La idea base de este modelo es el de imitar el sistema más complejo que se conoce hasta ahora, el cerebro. Éste esta formado por millones de células llamadas neuronas. Estas neuronas son unos procesadores de información sencillos con un canal de entrada de información, un órgano de cómputo y un canal de salida de información.

Las cualidades que presenta el cerebro son: procesamiento paralelo (pequeñas unidades elementales neuronas con un procesamiento simple), procesamiento distribuido (la información no se almacena localmente en determinadas zonas, sino que están presentes en toda ella) y adaptabilidad (capacidad de aprendizaje y generalización, por lo que puede responder a casos desconocidos), acción que es de mucho interés para los IDS.

Elementos de la neurona:

- Las entradas que reciben los datos de otras neuronas.

- Los pesos número, que se modifica durante el entrenamiento de la red neuronal, y es aquí por tanto donde se almacena la información que hará que la red sirva para un propósito u otro.
- Una regla de propagación. Con las entradas y los pesos se suele hacer algún tipo de operación para obtener el valor del potencial
- Una función de activación. El valor obtenido con la regla de propagación, se filtra a través de una función conocida como función de activación y es la que nos da la salida de la neurona.

Matemáticamente, una red neuronal la podemos ver como un grafo dirigido y ponderado donde cada uno de los nodos son neuronas y los arcos que unen los nodos son las conexiones. Los arcos serán unidireccionales la información se propaga en un único sentido; es ponderado, las conexiones tienen asociado un peso. Además las neuronas se agrupen en capas. Las básicas son: capa de entrada, capas ocultas y capa de salida.

Los sistemas de Redes Neuronales los podemos clasificar por la arquitectura y el tipo de aprendizaje.

Según la arquitectura la red se denomina unidireccional (feedforward) y recurrentes o realimentados (recurrent). Según el aprendizaje tenemos 4 tipos: con aprendizaje supervisado, con aprendizaje no supervisado o autoorganizado, con aprendizaje híbrido (mezcla de los anteriores) y con aprendizaje reforzado (reinforcement learning).

2.1. Tipos de redes

El perceptron multicapa (MLP)

Este es uno de los tipos de redes más comunes, basado en la red más simple llamada perceptrón (salvo que el número de capas ocultas puede ser mayor o igual que una). Es una red unidireccional (feedforward). Las neuronas de la capa oculta usan como regla de propagación la suma ponderada de las entradas con los pesos y sobre dicha suma ponderada se aplica una función de transferencia de tipo sigmoide, que es acotada en respuesta. El aprendizaje que se suele usar en este tipo de redes recibe el nombre de retropropagacion del error (backpropagation). Como función de coste global, se usa el error cuadrático medio. Sobre esta función de coste global se aplica algún procedimiento de minimización

Redes Autoorganizadas. Redes SOFM

En este tipo de redes el entrenamiento o aprendizaje es diferente al de las redes con entrenamiento supervisado. A la red no se le suministra junto a los patrones de entrenamiento, una salida deseada. Lo que hará la red es encontrar regularidades o clases en los datos de entrada, y modificar sus pesos para ser capaz de reconocer estas regularidades o clases.

Uno de los tipos de redes que pertenece a esta familia y que se ha usado bastante son los mapas autoorganizados, SOM (Self–Organizing Maps).

Es una red de tipo unidireccional, y se organiza en dos capas: la primera capa esta formada por las neuronas de entrada, y la segunda consiste en un array de neuronas de dos dimensiones. En este caso se necesitan dos índices para etiquetar cada neurona, los pesos asociados a cada neurona tendrán tres índices donde dos de ellos indican la posición de la neurona en la capa y el tercero la conexión con cierta neurona de entrada.

En cuanto al entrenamiento, utiliza un aprendizaje de tipo no supervisado, y cada neurona utiliza como regla de propagación una distancia de su vector de pesos al patrón de entrada. En este tipo se emplean dos conceptos que son los de neurona ganadora y vecindad de la misma. Uno de los algoritmos de aprendizaje usado es el algoritmo de Kohonen.

Redes de función de base radial (RBF)

Este tipo de redes se caracteriza por tener un aprendizaje o entrenamiento híbrido. La arquitectura de estas redes se caracteriza por la presencia de tres capas: una de entrada, una única capa oculta y una capa de salida. Se diferencia de la red MLP en que las neuronas de la capa oculta calculan la distancia euclídea entre el vector de pesos y la entrada y sobre esa distancia se aplica una función de tipo radial con forma gaussiana.

Para el aprendizaje de la capa oculta, hay varios métodos, siendo uno de los más conocidos el algoritmo denominado k–medias (*k–means*) que: un algoritmo no supervisado de clustering. Se fijan los valores de los centros, se ajusta las anchuras (parámetros función gaussiana) de cada neurona. Finalmente, se entrena la capa de salida, se suele emplear algoritmos semejantes a los empleados en capa de salida del MLP.

De todos estos esquemas de redes neuronales los que más se emplean son los dos primeros.

3. PROTOTIPOS A DESTACAR DESARROLLADOS

Destacaremos cuatro estudios recientes en el manejo de Redes Neuronales para la detección de la intrusión, para el modelo de mal uso y para el modelo de anomalía.

- Instituto De Investigación Técnica de Georgia (GTRI) [13]
- Instituto Tecnológico De Massachusetts (MIT) [14, 15]
- UBILAB Laboratory [10, 11]
- Reliable Software Technologies Corp (RST) [6]

Prot.	Tipo	Modelo	Carácter	Ataques
GTRI	Mal uso	MLP	4 niveles 9 nodos de entrada y 2 nodos de salida	escaneo ISS, escaneo SATÁN y SYNFlood
		MLP/ SOM	Red unidireccional (feedforward) Aprendizaje retropropagación (backpropagation)	ataques dispersos y posibles ataques cooperativos Fallos FTP
MIT (1999)	Mal uso Red	MLP	2 niveles k nodos de entrada 2k nodos ocultos 2 salidas aprendizaje por retropropagación	Palabras Clave Ataques a Host Unix, cuando el al obtener los privilegios de root sobre el servidor
UBILAB	Análisis de red	SOM	Monitoriza actividades de la red	IP spoofing, adivinación de contraseña de FTP, red escaseando hopping de red saltando
RST	Detección anomalías	MLP	Utilizaron el algoritmo Leacky Bucket que almacena temporalmente en memoria los eventos recientes de anomalías.	Analiza comportamiento de programas, captando llamadas al sistema.

Se detectan en los mismos un nivel bueno de rendimiento. Dicho rendimiento aumenta con la Red de Elman.

4. ANÁLISIS DE SISTEMAS DE REDES NEURONALES APLICADOS A LA DETECCIÓN DE INTRUSOS

Los sistemas a analizar son:

1. Red Neuronal detector de Intrusión [1]
2. IDS de anomalías de red jerárquico usando clasificación de red neuronal [2]
3. Detección de intrusos basado en Host con SOM [3] [10]
4. Detección de anomalías [4]
5. Detección de anomalías de usuario basado en Host Unís usando SOM [5]
6. Detección de anomalías y mal uso (misuse) mediante Redes Neuronales [6]
7. Aplicación de Redes Neuronales a la Seguridad Unís [7]
8.. Red Neuronal para detección de mal uso (misuse) [8]
9. Detección de anomalía por redes Elman [9]

Sist.	Año	Tipo RN empleada	Tipo detección	Carácter	% detección	% Falso Positivo	% Falso Negativo
1	1998	MLP	Detección Anomalías	Análisis de comportamiento de usuarios		7%	4%
2	2001	MLP y Análisis estadístico	Detección Anomalías	Recoge tráfico de la red			
3	2002 2003	SOM	Detección de anomalías	Aprendizaje no supervisado	89%	4,6%	
4	1998	MLP	Detección de abuso	Comportamiento de usuario a través de Aplicaciones a nivel de proceso		0	20%
5	2000	SOM	Detección de anomalías	Aprendizaje no supervisado		Alto o bajo según variable	Bajo o alto según variable
6	1999	MLP	Detección de abuso	Detección ataques y clasificación de ataques (datos normales o no)	77%	2,2%	
7	1995	MLP – TR	Detección Anomalías	Análisis de usuario			
8	1998	MLP	Detección de abuso (misuse)	IDS de red		Reduce las falsas alarmas respectos a los Sist. Expertos	
9	1999	ELMAN	Detección de anomalías		77% 100%	0	9%

(Nota: únicamente se recogen los datos incluidos en los artículos.). Se comprueba que las Redes Neuronales en las arquitecturas de detección de niveles se emplean para verificar alertas y minimizar los falsos positivos (caso 4).

5. CONCLUSIÓN

Los IDS aparecen como un elemento más dentro de la seguridad de nuestros sistemas, y que están influenciados por la Política de Seguridad definida. Los SDI basados en Redes Neuronales aparecen con un sistema muy prometedor dentro de los entornos de IDS, ya que permiten una mayor adaptabilidad al entorno permitiendo acciones de aprendizaje, lo que permite predecir las acciones de los usuarios dada una serie de acciones definidas. Otra de las ventajas, es que funcionan bien en entornos con ruido, y son capaces de detectar nuevas formas de ataque no conocidas sin necesidad de introducir reglas de forma manual. Como inconveniente podemos destacar la gran cantidad de datos a utilizar durante el entrenamiento.

6. REFERENCIAS

[1] Ryan, J.,Lin, M., Miikkulainen, R., "*Intrusion Detection with Neural Networks*" 1998. http://citeseer.nj.nec.com/ryan98intrusion.html.

[2] Z. Zhang, J. Li, C. Manikopoulos, and J. Ucles, "*A hierarchical anomalynetwork intrusion detection system using neural network classification*". Proceedings of 2001 WSES International Conference on: Neural Networks andApplications (NAA'01), Feb 2001.

[3] Lichodzijewski P., Zincir–Heywood A.N., Heywood M.I., "*Host–Basedintrusion detection using Self–Organizing Maps*," IEEE InternationalJoint Conference on Neural Networks, pp. 1714–1719, May 12–17, 2002.

[4] Ghosh, A., Wanken, J., and Charron, F. 1998. "*Detecting anomalous and unknown intrusions against programs".* In Proceedings of the 1998 Annual Computer Security Applications Conference (ACSAC'98).

[5] Höglund A.J. and Hätönen K and Sorvari A.S. "*A Computer Host–based User Anomaly Detection System using Self–Organizing Map*". In Proceedings of the IEEE–INNS–ENNS International Joint Conference on Neural Networks, volume 5, pages 411–416, 2000.

[6] Ghosh, A., Schwartzbard, A., "*A study in using neural networks for anomaly and misuse detection*", In Proceedings of the Eighth USENIX Security Symposium, 1999.

[7] Tan, K., "*The Application of Neural Networks to UNIX Computer Security*". In Proceedings of the IEEE International Conference on Neural Networks, Vol.1 pp. 476–481, 1995.

[8] Cannady, J., "*Artificial Neural Networks for Misuse Detection*". Proceedings, National Information Systems Security Conference (NISSC'98), October, Arlington, VA, pp. 443–456, 1998.

[9] Anup K. Ghosh, Aaron Schwartzbard, Michael Schatz: "*Learning Program Behavior Profiles for Intrusion Detection*". Workshop on Intrusion Detection and Network Monitoring 1999: 51–62.

[10] Kayacik, H.G., Zincir–Heywood, A.N., Heywood M.I., "*On the Capability of an SOM based Intrusion Detection System*", IEEE International Joint Conference on Neural Networks, July 20th–24th 2003

[11] Girardin, L., "*An eye on network intruder–administrator shootouts – UBS UBILAB*", In Proceedings of the 1st Workshop on Intrusion Detection and Network Monitoring (ID '99), Santa Clara, CA, (http://www.ubilab.org/publications/index.html)

[12] **Ghosh A., Schwartzbard, A., "*A study using Neural Networks for anomaly detection and misuse detection*", Reliable Software Technologies (http://www.docshow.net/ids /usenix_sec99.zip)

[13] Ghosh, A., Schwartzbard, A., Schatz, A., "*Learning program behavior profiles for Intrusion Detection*", Proceedings of the workshop on Intrusion Detection and Network Monitoring – Santa Clara, April 9–12 1999

[14] Cannady, J., & Mahaffey, J., "*The application of Artificial Neural Networks to Misuse detection: initial results*". Georgia Tech Research Institute. (http://www.raid–symposium.org/raid98/Prog_RAID98 /Talks.html #Cannady_34)

[15] Cunningham, R., Lippmann, R., "*Improving Intrusion Detection performance using Keyword selection and Neural Networks*", MIT Lincoln University (http://www.ll.mit.edu/IST/pubs.html)

[16] Cunningham, R., Lippmann R., "*Detecting Computer Attackers: recognizing patterns of malicious, stealthy behavior* ", MIT Lincoln Laboratory –Presentation to CERIAS 11/29/2000 (http://www.cerias.purdue.edu /secsem/abstracts0001.php).