Capítulo 2. Sistemas para detección de intrusos.

Este capítulo comienza con una introducción al concepto de los sistemas de detección de intrusos: se discuten los criterios primarios de medida de su efectividad, clasificación y enfoques aplicados en su implementación. Posteriormente se presenta una descripción de las características generales del protocolo HTTP, que representa el objeto de anomalías del presente trabajo. Finalmente se muestran distintas topologías de redes neuronales.

2.1. Introducción

Los primeros trabajos sobre detección de intrusos se remontan a 1970 por el pionero en seguridad informática James P. Anderson [AND72], donde presenta un reporte de dos volúmenes sobre los requerimientos de seguridad en sistemas informáticos de la fuerza aérea del gobierno estadounidense. En este reporte se plantea la seguridad necesaria en terminales de cómputo, principios de seguridad, técnicas de encriptado de archivos y modelos de seguridad en redes. Diez años más tarde Anderson introduce en [AND80] diversos conceptos que representan la base sobre la teoría de los sistemas de detección de intrusos, conceptos como tipos de amenazas, perfiles de usuarios clandestinos, sistemas de vigilancia y monitoreo de usuarios. La teoría de Anderson se desarrolla alrededor de la idea de obtener mediante sistemas de auditoría y vigilancia conjuntos de datos que permitan caracterizar el uso correcto y el uso intrusivo de los sistemas de cómputo, a partir de esta caracterización se podrían detectar actividades anómalas que representen amenazas o intrusiones. Otro trabajo representativo es desarrollado por Dorothy E. Denning [DEN87], donde describe un modelo de detección de intrusos en tiempo real basado en perfiles de conducta de usuarios capaz de detectar ataques exteriores e interiores, los perfiles son construidos a partir de registros de auditoría usando modelos estadísticos. Una de las principales ventajas es la generalización del IDS mediante la independencia de ambiente de aplicación, tipos de ataque y vulnerabilidades del sistema.

Como se reporta en [ORT04], en los siguientes años se desarrollaron IDS que utilizaron distintos enfoques en el análisis de sistemas de comunicación y detección de comportamientos intrusivos. En 1988 Harold S. Havitz y Alfonso Valdes desarrollaron IDES (Intrusion DetectionExpert System) que implementaba el modelo presentado por Dorothy E. Denning basado en un sistema experto. También en 1988, se desarrolla el primer IDS con estrategia de análisis de uso indebido, como parte del proyecto Haystack para la fuerza aérea de Estados Unidos de Norteamérica. Hacia 1990 Todd Heberlein

desarrolla en la Universidad de California el primer IDS basado en red llamado NSM (Network Security Monitor).

El término seguridad informática describe la tarea de proteger información en formato digital. La seguridad informática garantiza la implementación de medidas de protección contra ataques y evita el colapso total del sistema en caso de que un ataque ocurra [CIA08]. Existen tres características de la información que deben ser protegidas por la seguridad informática [WAT08]:

- Confidencialidad: asegura que solamente las entidades autorizadas puedan tener acceso a la información. Un sistema que posea características de confidencialidad podría implementar estrategias como corta fuegos, listas de control de acceso y encriptación de tráfico.
- Integridad: asegura que la información mantenga su estado de completitud y corrección, es decir, que ningún intruso o software malicioso haya alterado esos datos durante el tránsito entre los distintos nodos de la red. Algunos ejemplos de violaciones a la integridad incluyen la modificación de la apariencia de un sitio web, interceptar transacciones de comercio electrónico, modificar registros de bases de datos corporativas.
- Disponibilidad: es una medida de accesibilidad a los datos, asegura que los datos estén disponibles para los usuarios autorizados. Un ataque que represente vulnerabilidades en la disponibilidad puede efectuarse mediante la emisión sistemática de un gran número de peticiones a un sistema de red, ocasionando un consumo excesivo de sus recursos y que no se atiendan las peticiones legítimas emitidas por otros usuarios.

Las medidas de protección de seguridad informática son definidas a través de un modelo de seguridad, estos modelos definen las acciones que pueden ser ejecutadas por usuarios sobre distintos recursos (también llamados objetos). De esta manera el modelo de seguridad especifica que acciones son permitidas para cada usuario o grupos de usuarios [HEA90].

Como se señala en [HEA90] existen al menos tres maneras en las que el modelo de seguridad puede ser comprometido: una implementación incorrecta del modelo, autentificación errónea de usuarios y un atacante interno. Dadas estas vulnerabilidades, en [HEA90] se propuso que el modelo de seguridad pueda ser aumentando mediante un sistema de detección de intrusos, el cual tenga por objetivo el monitoreo de actividades específicas y alertar sobre anomalías en las actividades observadas. El monitoreo que lleva a cabo un sistema de detección de intrusos puede considerar las acciones desarrolladas por un solo usuario y desde esta perspectiva, autentificar su identidad mientras sus acciones se mantengan dentro de los parámetros definidos en el perfil de conducta normal. Por lo tanto,

si las actividades desarrolladas por el usuario difieren en gran medida de su perfil normal, existen razones para emitir un tipo de alerta.

2.2. Desempeño de los Sistemas de detección de Intrusos

Para determinar el desempeño de un sistema de detección de intrusos necesitamos conocer los índices de detección de falsos positivos y falsos negativos. Un falso positivo es un evento anómalo que resulta inofensivo, es decir una falsa alarma. Un falso negativo es una conducta intrusiva no detectada. Los casos posibles en la detección de anomalías se presentan en la Figura 2.1 [ORT04].

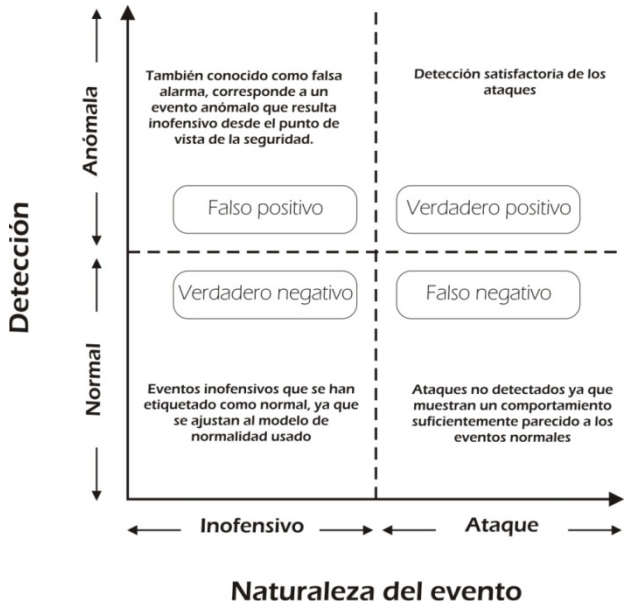

Figura 2.1: Detección de anomalías

2.3. Clasificación de los sistemas de detección de intrusos

La clasificación de los sistemas de detección de intrusos se realiza de acuerdo a dos características [ORT04], Figura 2.2:

- Fuentes de información. Se refiere al origen de los datos que se analizan para determinar si un ataque se ha llevado a cabo.

- Técnica de análisis. Es el método por el cual se va a analizar los datos obtenidos de la fuente de información.

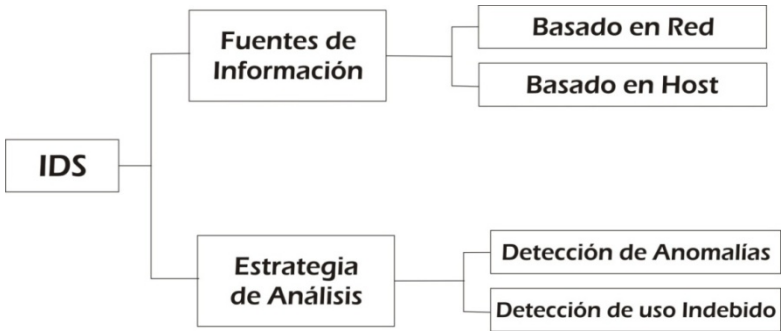

Figura 2.2: Clasificación de los sistemas

2.3.1. Fuentes de información

Los IDS pueden obtener la información para el análisis de distintas fuentes, cuando se consideran solamente los datos obtenidos de terminales se dice que está basado en host, en cambio cuando se evalúa el flujo de información que viaja a través de la red se clasifica al IDS como basado en red.

Dentro de los IDS basados en host encontramos herramientas como informes del sistema, registros de auditoría, llamadas del sistema de procesos en ejecución, OSSEC HIDS, o algunas comerciales como PGP Endpoint Application Control, Kane Security Monitor y TripWire. En los IDS basados en red se pueden utilizar herramientas como Snort, Bro, Cisco IDS, para capturar y analizar paquetes de red [ATH03].

2.3.2. Técnica de análisis

Cuando se consideran las técnicas de análisis los sistemas de detección de intrusos se clasifican en la detección de uso indebido y la detección de anomalías. La detección de uso indebido utiliza una base de datos que contiene firmas de ataques, cada firma corresponde a un ataque en específico y describe sus características, en esta técnica es importante mantener actualizada la base de datos que contiene las firmas de ataques pues solamente se pueden reconocer ataques cuyo perfil coincida con alguna de ellas. La principal desventaja que tiene la técnica de análisis de uso indebido es que solamente se pueden detectar ataques ya conocidos cuyo patrón ha sido descrito anteriormente. Una de las ventajas de los IDS basados en uso indebido es que se obtiene un índice bajo de falsos positivos.

La detección de anomalías analiza el estado y conducta del sistema para categorizar conductas normales y anómalas. La principal ventaja de la técnica basada en anomalías es

que no necesitamos establecer una base de firmas de características de cada ataque sino que se pueden detectar ataques nunca antes vistos. Sin embargo la desventaja de éste enfoque es la volatilidad del estado normal de operación de la red, que puede llevar a tener una tasa considerable de falsos positivos [FAN09].

La técnica de análisis de los sistemas de detección de intrusos usada actualmente por la mayoría de los sistemas de detección de intrusos comerciales se basa en el conocimiento de ataques previos por medio de la actualización de firmas dejando que nuevos ataques o variaciones de ataques pasados no puedan ser reconocidos [BAI03].

2.4. Investigación en los sistemas de detección de intrusos

Debido al crecimiento de ataques en las redes de comunicaciones, los sistemas de detección de intrusos se han convertido en un componente necesario que combinado con otras técnicas buscan brindar protección en contra de actividades ilegales o no autorizadas. Actualmente existe una tendencia a incorporar distintas disciplinas como inteligencia artificial y estadística clásica a los sistemas de detección de intrusos [ORT04], algunas de técnicas aplicadas en el desarrollo de los IDS son las siguientes [SHU08], [BAI03]:

- Sistemas expertos. Los datos recolectados son comparados con una serie predefinida de reglas que describen perfiles de ataques, inicialmente los sistemas expertos fueron los más utilizados en la implementación de IDS.
- Análisis de firmas. Es similar a los sistemas expertos, tiene descripciones semánticas de los ataques, y es el método más comúnmente usado por los sistemas comerciales.
- Redes de Petri. Utilizando bases de conocimiento expertas, se genera una representación gráfica de los ataques.
- Análisis Estadístico. La conducta de los datos es comparada contra un número de variables a través del tiempo. Algunas de las variables pueden ser inicios de sesión, uso de disco duro, memoria, procesador, etc. La principal ventaja de los enfoques estadísticos subyace en su habilidad de adaptación en el aprendizaje de la conducta de los usuarios.
- Minería de datos. Trata de recuperar relaciones entre largas cantidades de datos, se trata de un proceso en el que se buscan patrones de comportamiento a partir de los datos analizados y de esta forma desarrollar un modelo predictivo.
- Redes neuronales. Vectores de entrada representativos a objetos de anomalías son presentados a la arquitectura para que al finalizar una etapa de aprendizaje detecte correctamente patrones que describan ataques.

2.5. Protocolo HTTP

Dentro de la definición del sistema de detección de intrusos es indispensable definir el objeto de anomalías en el que se va a concentrar. Existen muchas posibles opciones sobre la variable que se va a utilizar en la detección de conductas anómalas, algunos sistemas de detección de intrusos utilizan comandos del usuario, funciones sobre el porcentaje de uso de interfaces de red o determinados protocolos.

Debido a que el protocolo de nivel de aplicación HTTP (protocolo de transferencia de híper texto) es usado en la mayoría de las transacciones web, se vuelve más propicio a sufrir ataques por medio de este tipo de mensajes, razón por la cual este trabajo se concentra en la detección de conductas intrusivas sobre este protocolo.

El protocolo de transferencia usado en la comunicación entre navegadores web y servidores es HTTP (protocolo de transferencia de híper texto), este protocolo especifica el tipo de mensajes que los clientes pueden enviar a los servidores y las respuestas emitidas de vuelta al cliente. La manera usual en que un navegador web se comunica con un servidor es mediante el establecimiento de una conexión al puerto 80 del servidor. El protocolo HTTP tiene asociados varios métodos los cuales permiten distintas operaciones descritas en la Tabla 2.1.

Tabla 2. 1: Funciones HTTP

GET	**Solicita al servidor el envío de una página web**
HEAD	Solicita lectura del encabezado de una página web
PUT	Solicita el guardado de una página web
POST	Envía datos a ser procesados por el servidor
DELETE	Remueve la página web
TRACE	Imprime en el cliente las peticiones recibidas por el servidor
CONNECT	Lleva a cabo la conversión de la petición de conexión a un túnel TCP/IP
OPTIONS	Devuelve los métodos http soportados

Los ataques que tienen como herramienta el navegador web comúnmente involucran el uso de *cookies, javaScript, Java y cross site scripting* [CIA08].

Cookies: Debido a que el protocolo HTTP no maneja estados, cualquier información introducida en una página web no es retenida cuando se lleva a cabo otra petición en el navegador. Esto ocasiona que cada vez que un usuario vuelva a una página visitada anteriormente el servidor no recupere la información relacionada, la solución a este

problema es que el servidor guarde un archivo de texto en el cliente que contenga toda la información del usuario para recuperarla en visitas posteriores. El archivo es llamado cookie. Las *cookies* pueden clasificarse en *cookies* de origen y *cookies* de terceros. Las *cookies* de origen se refieren a los archivos utilizados por los servidores que las crean y las *cookies* de terceros es cuando un servidor intenta acceder a archivos que no fueron creados por él. Las *cookies* de terceros pueden ser utilizadas para establecer hábitos de navegación o hábitos de compra, esta información puede inferir que tipos de productos le interesan al cliente y enviar publicidad especializada.

JavaScript: Los *scripts* residen dentro del código HTML, cuando un servidor que usa JavaScript es accesado, el documento HTML junto con el código *JavaScript* es bajado en el cliente. Posteriormente en la máquina del cliente el navegador web utiliza un intérprete de *Java* para ejecutar el código *JavaScript*. Debido a que el código de los *scripts* puede recolectar y enviar información sin la autorización del usuario, los *scripts* de *Java* representan una amenaza de seguridad.

Java: A diferencia de los *scripts*, los *applets* de *java* son programas separados del documento HTML. Los *applets* son almacenados junto con los documentos HTML en el servidor y posteriormente bajados en la máquina del cliente. Una vez que se encuentran en la máquina del cliente todo el procesamiento se lleva a cabo sin la intervención del servidor.

Active-X: Son un conjunto de controles los cuales pueden ser desarrollados en varios lenguajes como C++, *Visual Basic* o *Java*. Los controles *Active X* tiene un acceso completo a las características del sistema operativo, pueden crear directorio, borrar archivos o formatear unidades.

Cross Site Scripting (XSS): esta técnica involucra scripts que son diseñados para extraer información del usuario y pasarla al atacante. XSS es un tipo de ataque que tiene como objetivo sitios web que generan dinámicamente páginas que despliegan entradas de usuario que no han sido apropiadamente validadas. Un atacante que usa XSS puede comprometer información confidencial, manipular cookies o ejecutar código malicioso en el cliente.

2.6. Redes neuronales

El concepto de red neuronal está inspirado en el proceso biológico del cerebro humano, el cual procesa de manera paralela y distribuida las tareas por medio de la conexión de varias unidades llamadas neuronas. Una red neuronal artificial tiene unidades de procesamiento llamadas neurones, la tarea del neurón es recibir un estímulo de entrada de una fuente externa o de otro neurón y computar nueva información direccionándola a su salida. Una

vez que se ha procesado la nueva información esta puede ser enviada a otros neurones o representar la salida del sistema (ver Figura 2.3). Las conexiones entre los neurones tienen asociados pesos que son representados por números reales pequeños, estas conexiones representan canales de comunicación por medio de los cuales se envían estímulos de un neurón a otro o bien estímulos provenientes del exterior. Los valores asociados a las conexiones se ajustan a través de los algoritmos de entrenamiento, el proceso de ajuste de pesos sinápticos es una tarea que se lleva a cabo en paralelo lo que quiere decir que varios neurones pueden ajustar sus valores de manera simultánea.

Una característica común de las redes neuronales artificiales es que se pueden organizar en capas, en la Figura 2.3 se muestra una red neuronal con tres capas.

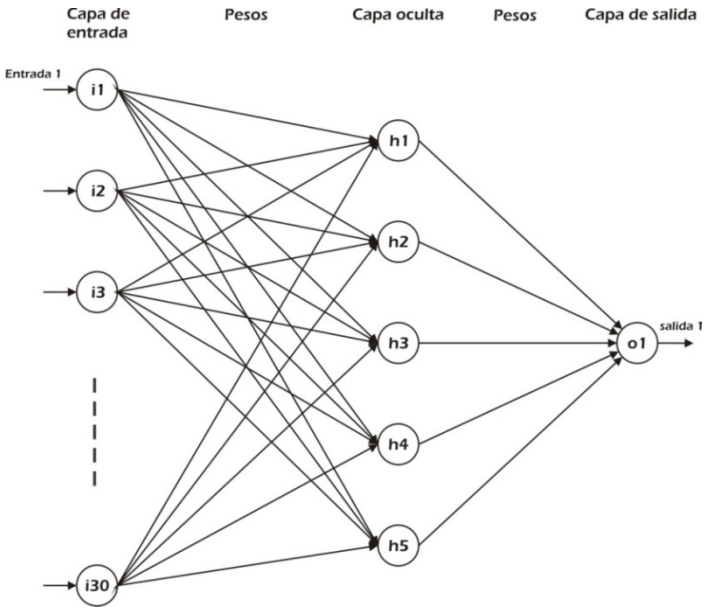

Figura 2.3: Red neuronal con alimentación hacia adelante.

La primera capa es la capa de entrada, esta capa es encargada de recibir los datos provenientes del exterior que van a ser procesados por la red neuronal. La segunda capa, también llamada capa oculta, procesa los datos enviados desde la capa de entrada y envía su salida hacia la última capa. Finalmente la tercera capa llamada capa de salida toma los datos enviados desde la capa oculta, los procesa y obtiene la salida final del sistema. Las redes neuronales son entrenadas a partir de un conjunto de datos representativo y durante esta etapa de aprendizaje los parámetros libres de la red (pesos sinápticos) son modificados para producir una salida más cercana al objetivo. Existen distintos modelos de arquitectura de rede neuronales, cada modelo describe una topología la cual establece el esquema de interconexión entre los neurones, el número de capas, el algoritmo de entrenamiento aplicable, los métodos de inicialización de parámetros libres.

2.6.1. Redes neuronales con alimentación hacia adelante

En las redes neuronales con alimentación hacia adelante todas las conexiones van en una dirección desde la capa de entrada hasta la capa de salida. Un neurón con una sola entrada es mostrado en la Figura 2.4, tiene una entrada escalar denotada por p , la cual es multiplicada por el valor del peso sináptico w, este producto (wp) es sumado con el peso de b que determina un estimulo constante igual a uno. La suma es conocida como entrada a la red, finalmente se produce la salida al aplicar la función de transferencia a esta suma.

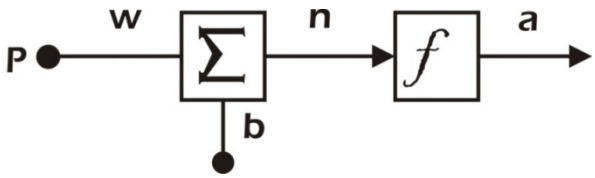

Figura 2. 4: Modelo de Neurón

Típicamente una red neuronal tiene más de un neurón y más de una entrada, cada entrada tiene un peso sináptico distinto con el neurón destino, la operación se puede describir por medio de una matriz de pesos que contiene todos los pesos asociados a cada neurón:

La salida del sistema queda establecida por la ecuación:

$$a = f(Wp + b) \tag{2. 1}$$

donde W es la matriz de pesos sinápticos y p es el vector columna que representa las entradas presentadas a la capa de entrada.

Una vez que hemos definido la estructura básica que define a una capa podemos establecer una arquitectura común en las redes neuronales conocida como perceptrón multicapa el cual tiene varios neurones por capa, la salida de cada capa alimenta la entrada de la capa subsecuente (ver Figura 2.5).

Si se representa el vector de entradas externas con R elementos, el número de neurones en la primera capa por s^1. La primera entrada de red de la primera capa como n_1^1, la segunda entrada de red de la primera capa como n_2^1, es decir el superíndice indicando el número de capa, la salida final queda dada por:

$$a^3 = f^3(W^3 f^2(W^2 f^1(W^1 p + b^1) + b^2) + b^3 \tag{2. 2}$$

2.6.2. Redes neuronales recurrentes

Las redes neuronales recurrentes tienen conexiones de retroalimentación o ciclos que regresan a neuronas en otras capas u otras unidades. Un posible esquema de interconexión recurrente se puede formar conectando las neuronas en capa de salida con las neuronas de la capa de entrada. Otro posible escenario es el de la conexión desde neuronas en la capa oculta con neuronas en la capa de entrada. Las redes neuronales recurrentes ofrecen características que permiten la representación de un estado interno que sea adaptativo en lugar de ser fijo, y de esta manera permitir la creación de estructuras de memoria que puedan preservar el estado durante un periodo de tiempo. Un aspecto importante que se debe tomar en cuenta en las redes neuronales recurrentes es el de la estabilidad. La capacidad de las redes neuronales recurrentes es que responde temporalmente a una señal externa aplicada a la capa de entrada. El uso de conexiones de retroalimentación permite representar estados dentro de la red lo cual las hace adecuadas a problemas como predicción no lineal, ecualización y procesamiento de voz [HAY98].

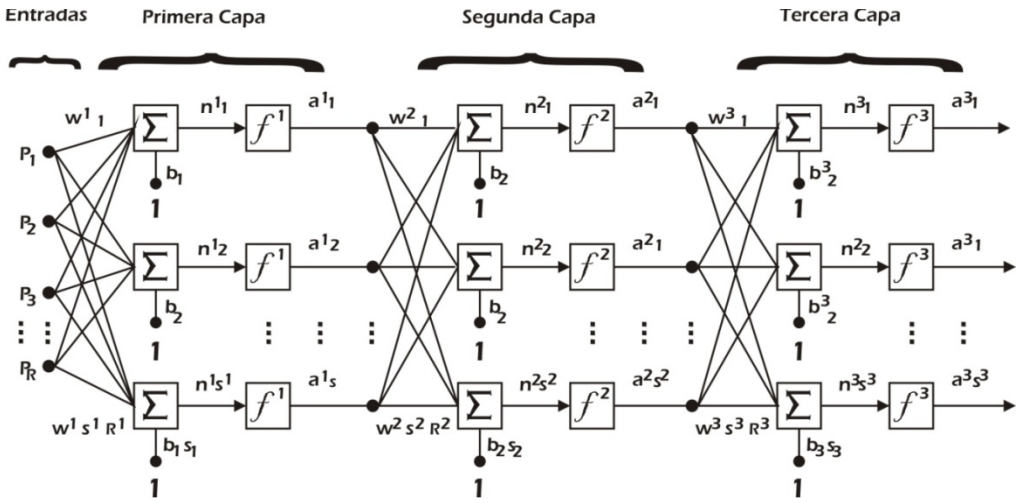

Figura 2. 5: Red neuronal con múltiples capas

Modelo Recurrente Input Output.

En éste modelo de arquitectura se considera una sola entrada a la cual se le aplica una serie de unidades de retraso, también se considera una sola salida la cual es conectada a la entrada pasando antes por un número q de unidades de retraso. El resultado de todas las unidades de retraso es presentado también a la capa de entrada del perceptrón. El valor de entrada presentado al modelo se denota por $u(n)$ y la salida se denota por medio de $y(n+1)$. De esta manera el vector presentado a la capa de entrada del perceptrón consiste en una ventana de datos que se construye a partir de:

- valores presentes y pasados de entrada que se consideran la influencia externa o exógena al sistema
- valores retrasados de la salida

Esta arquitectura se conoce como auto regresiva no lineal con entradas exógenas (NARX), la conducta dinámica del modelo está dada por la ecuación:

$$y(n+1) = F(y(n), \dots, y(n-q+1), u(n), \dots, u(n-q+1)) \quad (2.3)$$

En la Figura 2.6 se muestra el esquema de este modelo.

Modelo espacio estado.

En éste modelo las neuronas ocultas definen el estado del modelo, la salida de la capa oculta es conectada a la capa de entrada por medio de un banco de unidades de retraso. La capa de entrada consiste en los nodos de retroalimentación y los nodos fuente (entrada). La red neuronal está conectada al ambiente externo por medio de los nodos fuente. Si denotamos $u(n)$ como el vector de entrada, $x(n)$ el vector de salida de la capa oculta en el tiempo n, entonces el modelo queda representado por medio de las siguientes ecuaciones:

$$x(n+1) = f\big(x(n), u(n)\big) \quad (2.4)$$

$$y(n) = Cx(n)$$

donde f representa una función no lineal y C es la matriz de pesos sinápticos de las unidades de salida, por lo tanto la capa de unidades ocultas no es lineal, pero las unidades de salida son lineales. Este modelo de incluye varias arquitecturas como la red recurrente simple propuesta por Elman, en la cual las unidades de salida pueden no ser lineales y se omiten las unidades de retraso en la capa de salida.

Perceptrón multicapa recurrente.

Esta arquitectura tiene una o más capas ocultas, cada capa oculta tiene conexiones de retroalimentación que regresan a su capa de entrada, si denotamos como $x_a(n)$ a la salida de las unidades ocultas en la capa a dentro del perceptrón, la respuesta al vector de entrada $u(n)$ queda dada por las siguientes ecuaciones:

$$x_1(n+1) = f(x_1(n), u(n)) \quad (2.5)$$

$$x_2(n+1) = f(x_2(n), x_1(n+1))$$

$$\dots$$

$$x_0(n+1) = f(x_0(n), x_k(n+1)$$

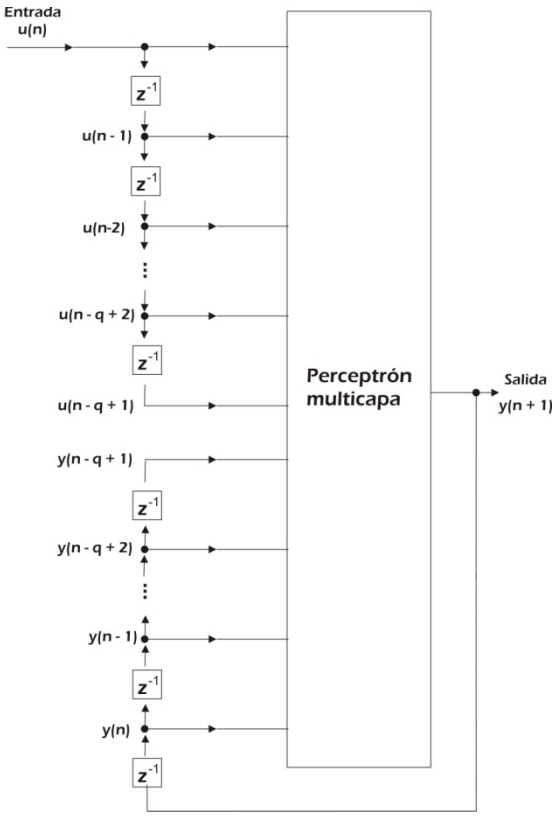

Figura 2. 6: Red neuronal NARX

Investigaciones han comparado el desempeño obtenido por redes neuronales con distintos esquemas de interconexión y algoritmos de aprendizaje aplicados a la detección de intrusos, dentro de los resultados obtenidos se ha logrado un mejor índice de detección e identificación con redes neuronales recurrentes [SAN05], [AQU08], [MEJ03], [XUE04].

En la Figura 2.7 se muestra el diagrama del perceptrón multicapa recurrente:

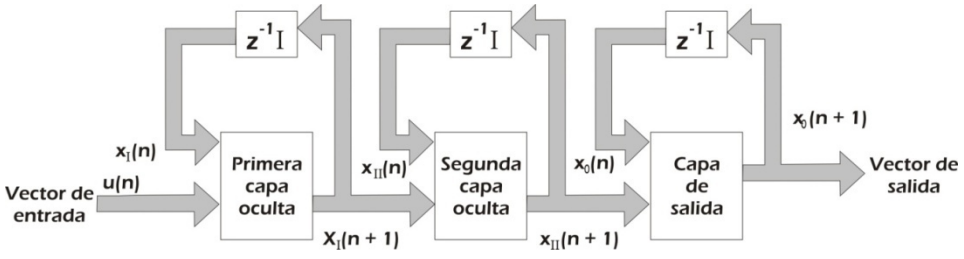

Figura 2. 7: Perceptrón multicapa recurrente

2.6.3. Algoritmos de aprendizaje

El algoritmo de aprendizaje establece la manera en que los parámetros libres de la red y pesos sinápticos van a ser ajustados, se puede decir que estas reglas dictan la manera en que las redes neuronales aprenden de los estímulos presentados por el ambiente externo. Si el proceso de aprendizaje es llevado a cabo correctamente la red neuronal puede desarrollar las capacidades para generalizar y exhibir características predictivas. En la definición del proceso de aprendizaje implica una secuencia de eventos:

1. La red neuronal es estimulada por el ambiente externo, es decir, se le presentan instancias del problema a la red neuronal.
2. Como resultado del estímulo los parámetros libres son cambiados.
3. La red neuronal responde de una nueva manera a los estímulos externos como resultado de los cambios en su estructura interna.

En la actualidad existe un gran número de algoritmos de entrenamiento, la diferencia sustancial subyace en la manera en que se alteran los pesos sinápticos y en la relación que tienen los neurones en su ambiente.

En las redes neuronales existen dos paradigmas de aprendizaje, el primero se llama aprendizaje supervisado y también es conocido como aprendizaje con maestro. El segundo paradigma de aprendizaje recibe el nombre de no supervisado o aprendizaje sin maestro. Esta trabajo contempla la utilización de un aprendizaje supervisado.

El aprendizaje supervisado puede ser visto como una instancia llamada maestro la cual posee conocimiento sobre el ambiente en un formato de ejemplos de entrada y salida, el maestro da ejemplos de instancias a la red neuronal sobre lo que se considera en este trabajo como un patrón de tráfico normal o intrusivo, lo que significa que clasificará conductas como intrusivas o no intrusivas. Finalmente la red neuronal usa las reglas generadas en el entrenamiento para clasificar nuevas instancias presentadas y emite una alerta si representa un patrón malicioso.

En contraste con el aprendizaje supervisado, en el aprendizaje no supervisado no existe el maestro que indique cuando se trata de tráfico intrusivo o tráfico normal. El aprendizaje no supervisado tiene la habilidad de tomar instancias sin clasificar y crear nuevas clases automáticamente. El primer paso en el aprendizaje no supervisado es utilizar un algoritmo de agrupamiento sobre los datos, después se lleva a cabo un proceso de etiquetado sobre los vectores de pesos que finalmente pueden ser usados para clasificar nuevas instancias.

Existen dos modos de entrenamiento para las redes neuronales recurrentes: basado en épocas y continuo, estos modos se describen a continuación. En el entrenamiento basado

en épocas se establece un periodo de tiempo en el que la red recurrente empieza a funcionar, desde un estado inicial hasta que alcanza un nuevo estado, en este punto el entrenamiento se detiene y se restablece la red neuronal al estado inicial correspondiente al siguiente periodo. El estado inicial no tiene que ser el mismo para cada periodo de entrenamiento, lo que resulta importante es que el estado inicial en un periodo nuevo sea distinto del alcanzado en el periodo anterior. Como se describe en [GOL05] se debe distinguir entre el entrenamiento basado en épocas de las redes neuronales recurrentes y los entrenamientos: por lote (batch) e incremental utilizados en las redes neuronales con alimentación hacia adelante. La principal diferencia es el momento en el que se lleva a cabo la actualización de pesos. El enfoque basado en lotes actualiza los pesos solamente después de que se complete un ciclo de presentaciones de patrones. En el enfoque incremental se lleva a cabo un ajuste de pesos después de la presentación de un patrón a la red neuronal. En el contexto de redes neuronales recurrentes los ajustes de pesos basados en épocas pueden utilizar un enfoque incremental (los cambios en los pesos se llevarán a cabo al final de cada época) o basado en lotes (los cambios en los pesos se llevan a cabo después de varias épocas).

El método de entrenamiento continuo es ideal para situaciones donde no existen estados de restablecimiento disponibles o donde se requiere un aprendizaje en línea. La característica que define el entrenamiento continuo es que la red neuronal recurrente aprende mientras el procesamiento de señal es desarrollado por la red neuronal, es decir que el proceso de entrenamiento no se detiene. El entrenamiento continuo es ideal en situaciones donde la operación continua de la red no ofrece tiempo para detener el proceso y empezar con nuevos valores para los parámetros libres de la red.

2.6.4. Redes neuronales wavelet

El trabajo realizado por Marc Thulliard [THU02] remonta el origen de las redes neuronales basadas en wavelets al trabajo de Daugman en el año de 1988, en el cual se utilizaron wavelets Gabor para la clasificación de imágenes. También se señala que las redes wavelet se volvieron más conocidas después de varios trabajos realizados durante el año de 1992: Y. C. Pati presenta en su investigación una red wavelet con alimentación hacia delante, Zhang ocupa un modelo neuronal con wavelets para el control de un robot y Szu centra su investigación en clasificación de fonemas y reconocimiento de voz.

Las redes wavelet o *wavenets* son redes neuronales que combinan la teoría wavelet con el campo de las redes neuronales. Las redes wavelet han sido aplicadas a problemas de clasificación e identificación de manera exitosa que van desde monitoreo de procesos de

manufactura, procesamiento de voz, problemas de control y procesamiento de imágenes [THU02].

La arquitectura típica de las redes neuronales basadas en wavelet utiliza un conjunto de funciones wavelet que reemplazan a las funciones sigmoides, la combinación de teoría wavelet con redes neuronales tiene las siguientes ventajas [YUL07]:

1. Las unidades están basadas en el análisis teórico wavelet, lo que permite una mayor flexibilidad en las funciones de transferencia aplicadas a los vectores de entrada.
2. Las redes neuronales basadas en wavelets tienen menos parámetros que tienen que ser ajustados durante el proceso de entrenamiento.
3. Cada unidad que utiliza funciones wavelet tienen poca influencia sobre las otras unidades demás de ésta manera se incrementa la velocidad de entrenamiento.
4. El proceso de aprendizaje en las redes neuronales basadas en wavelet es un proceso de aproximación para obtener una solución global óptima en contraste con los puntos locales mínimos.

2.7. Redes neuronales aplicadas a los sistemas de detección de intrusos

Los IDS intentan detectar actividades inapropiadas y proveen un sistema de alarma activado en caso de intrusiones o ataques. Este sistema de alarma puede realizar acciones reactivas como la reconfiguración del firewall, ejecución de rutinas para manejar el evento, guardar información en la bitácora del sistema, terminar la sesión de comunicación y guardar información sobre el atacante [WAT08]. Dentro de la técnica de análisis basada en anomalías, se trata de detectar ataques o intrusiones a partir de desviaciones significativas del perfil de comportamiento normal, es decir, se trabaja en el principio de comparar las acciones de los usuarios contra un perfil de conducta normal o aceptable.

Las redes neuronales artificiales permiten a las computadoras el aprendizaje y la adaptación de distintas tareas que se les presentan, se inspiran en la forma en la que funciona el cerebro humano. Existen neuronas interconectadas las cuales, dependiendo del impulso de entrada, manifiestan una respuesta. El impulso que las neuronas reciben de entrada se traduce en una decisión como salida en la red neuronal. Para llevar a cabo decisiones, la red neuronal necesita una fase de entrenamiento iterativo, en la cual se aplican muestras de datos y se ajustan pesos hasta que el factor resultante se encuentre cerca del resultado deseado [HAG96]. Las redes neuronales poseen características predictivas y de reconocimiento de patrones que las hacen una herramienta atractiva para los sistemas de detección de intrusos. La investigación en esta área se enfoca en explorar diversas topologías y algoritmos de entrenamiento que permitan establecer mejores resultados de desempeño, tanto en la etapa de entrenamiento como en la de prueba. Debido

a su propiedad de generalización una red neuronal puede reconocer patrones no presentados durante la etapa de entrenamiento, de esta forma se detectan variantes de ataques entrenados que hayan traspasado otros niveles de protección como firewalls [KUK08].

Dentro de la incorporación de redes neuronales a los IDS se ha experimentado con distintas arquitecturas y algoritmos de aprendizaje, se encuentran trabajos iniciales como [CAN98] donde se muestran algunas ventajas y desventajas de la aplicación de redes neuronales en los IDS. En [RYA02] se implementa un sistema de detección de intrusos llamado NNID (Neural Network Intrusion Detection) el cual usa como objeto de anomalías un conjunto de comandos ejecutados por el usuario y trata de que la red neuronal identifique las variaciones en los patrones de uso. En [LIM08] se desarrolla un prototipo llamado I-IDS (Intelligent Intrusion Detection System) el cual lleva a cabo un monitoreo sobre el flujo de paquetes y clasifica los eventos de red utilizando una red neuronal de perceptrón multicapa entrenada con el algoritmo de propagación hacia atrás que tiene como función de activación la tangente hiperbólica, otros trabajos que también utilizan MLP/BP se encuentran en [LIJ05], [GOL05], [FAN09], [SHU08].

La investigación desarrollada por Kukielka y Kotulski hace un estudio comparativo entre tres arquitecturas diferentes aplicadas a los sistemas de detección de intrusos: MLP/BP, Radial Basis Function y Self Organizing Maps [KUK08] . En esta investigación se toman como datos de entrenamiento los producidos en el proyecto KDD99[1].

En los distintos trabajos descritos anteriormente se analizan las medidas de desempeño: falsos positivos, falsos negativos, índices de detección y velocidad de convergencia del algoritmo de entrenamiento seleccionado. Como se plantea en [YUL07], los distintos experimentos revelan problemas comunes que reducen el desempeño de las redes neuronales aplicadas a los IDS como son: la baja velocidad de convergencia, zonas de mínimo local, la dificultad para determinar el número de capas ocultas y la cantidad de unidades en éstas capas, así como la cantidad y calidad de muestras necesarias durante el entrenamiento.

Para solucionar los problemas señalados anteriormente se han propuesto métodos que incluyen el uso del algoritmo *annealing* simulado junto con algoritmos genéticos para superar el mínimo local, la aceleración del proceso de entrenamiento mediante algoritmos como Levenberg – Marquardt y gradiente conjugado. Sin embargo las mejoras señaladas anteriormente se ven afectadas cuando se tiene un largo número de datos multidimensionales [YUL07], por consiguiente es necesario llevar a cabo investigación que proponga nuevos modelos de redes neuronales aplicadas a los IDS.

[1] Fifth International Conference on Knowledge Discovery and Data Mining, llevada a cabo en 1999, este conjunto de datos fue creado a partir de un programa de evaluación de detección de intrusos de DARPA(Defense Advanced Research Project Agency).

2.8. Discusión

Los sistemas de detección de intrusos utilizan distintas estrategias de análisis, el presente trabajo se concentra en la detección de anomalías al utilizar una red neuronal recurrente con wavelets en su funcionamiento. Al analizar el desempeño que tiene un IDS en particular, es necesario entender distintos fenómenos que se dan en este contexto: falsos positivos y falsos negativos, valores que representan parte importante de la etapa de análisis ya que a partir de ellos se calculan distintos porcentajes de rendimiento.

Las redes neuronales funcionan a partir de unidades básicas llamadas neurones, las cuales responden a estímulos externos a través de una función de transferencia que produce un valor numérico como salida. Existen distintos esquemas de interconexión dependiendo de la complejidad deseada en el modelo, los esquemas sencillos llamados de alimentación hacia adelante constituyen la estructura básica de conexión en las redes neuronales, mientras que la complejidad incrementa a medida que se añaden conexiones entre unidades y capas, dando lugar a esquemas recurrentes. Las arquitecturas recurrentes implementan conexiones recursivas de una capa a otra permitiendo de esta manera la introducción de términos de memoria que tomen en cuenta estados pasados durante la etapa de entrenamiento, es necesario puntualizar que los algoritmos de entrenamiento son distintos para cada arquitectura.

A lo largo de este capítulo se presentaron a manera de resumen distintas arquitecturas como el perceptrón multicapa, modelo espacio estado, perceptrón multicapa recurrente, etc. Todos estos modelos ayudan a entender la diversidad de esquemas disponibles al tratar un problema en específico, dentro de la implementación de IDS que utilicen redes neuronales, se han llevado a cabo con éxito diversas investigaciones, cada una utilizando distintas arquitecturas y algoritmos de entrenamiento.

Al igual que en otras arquitecturas, las redes neuronales que poseen unidades de procesamiento con funciones wavelet tienen distintos esquemas de interconexión y organización de capas, de especial atención en esta investigación son las redes recurrentes.