Modelo de detección de intrusiones en sistemas computacionales,realizando selección de características con chi square, entrenamiento y clasificación con ghsom

Instruders detection computer systems model, chisquare, training and classification using shsom, features selection criteria

Johan Mardini
Magíster en ingeniería de sistemas y computación Universidad del Atlántico Barranquilla-Colombia. Correo electrónico: johanmardini@mail.uniatlantico.edu.co

Alberto Egea Colmenares
Especialista en Gerencia de Proyectos. Universidad de la Costa (C.U.C.) Barranquilla-Colombia. Correo electrónico: albertegea@gmail.com

Información del artículo: recibido: julio de 2016, aceptado: septiembre de 2016
https://doi.org/10.17081/invinno.5.1.2614

Resumen

Dado que la información se ha constituido en uno de los activos más valiosos de las organizaciones, es necesario salvaguardarla a través de diferentes estrategias de protección, con el fin de evitar accesos intrusivos o cualquier tipo de incidente que cause el deterioro y mal uso de la misma. Precisamente por ello, en este artículo se evalúa la eficiencia de un modelo de detección de intrusiones de red, utilizando métricas de sensibilidad, especificidad, precisión y exactitud, mediante un proceso de simulación que utiliza el DATASET NSL-KDD DARPA, y en concreto las características más relevantes con CHI SQUARE. Esto último a partir de una red neuronal que hace uso de un algoritmo de aprendizaje no supervisado y que se basa en mapas auto organizativos jerárquicos. Con todo ello se clasificó el tráfico de la red BI-CLASE de forma automática. Como resultado se encontró que el clasificador GHSOM utilizado con la técnica CHI SQUARE genera su mejor resultado a 15 características con precisión, sensibilidad, especificidad y exactitud.

Palabras clave:

DATASET KDD NSL DARPA; IDS (sistema de detección de intrusiones); GHSOM (mapas auto organizativos jerárquicos); reconocimiento de patrones.

Abstract

For organizations, information has become one of their most valuable assets, that is whyit is necessary to safeguard it, by using different strategies of protection, in order to avoid intruders access or any incident caused by data damage and misuse. Then, this paper aims to assess the efficiency of a proposed model related to network intruders detection, using metrics of sensitivity, specificity, precision and accuracy. This model uses DATASETNSL-KD DARPA, by selecting the most relevant featureswith CHI SQUARE and training a neural network,through a simulation process which use a non-supervised learning algorithm based on hierarchical organization maps, in order to classify BI-CLASSnetwork automatically. As a result, it was showed that using the GHSOM classifier with CHI SQUARE as features selection criteria, generate its best result : 15 features with precision, sensitivity, specificity and accuracy.

Keywords:

DATASET KDD NSL DARPA; IDS (intruders detection system);GHSOM (hierarchical organizational maps); pattern recognition.

INTRODUCCIÓN

Actualmente los sistemas informáticos gestionan gran cantidad de datos, ya sean de empresas o particulares. Este crecimiento ha provocado a su vez un aumento en los accesos no autorizados y la manipulación de datos. La gran conectividad de la que se dispone hoy en día, además de proporcionarnos acceso a datos, propicia un aumento en las intrusiones de red con las consecuentes violaciones de seguridad. Los atacantes están cada vez más preparados, llegando a ser muchos de ellos expertos analistas de las debilidades de los sistemas. A todo esto, hay que sumarle los problemas de configuración y la falta de recursos para instalar los parches de seguridad necesarios. En definitiva, se hace patente la necesidad de concienciar y enseñar en torno a la seguridad informática.

La seguridad informática tiene como objetivo principal disminuir los riesgos a que se exponen los sistemas informáticos. En este sentido, es claro que las vulnerabilidades de software constituyen el factor principal de los problemas de seguridad de Internet desde hace mucho tiempo, y esto es aprovechado por los hackers para apropiarse de información, intercambiarla o distribuirla a otras personas, así como para alterar o modificar el comportamiento de las aplicaciones. Todo esto ha originado técnicas, métodos y sistemas inspirados en estrategias de defensa, ataque y contraataque, cuyo propósito general es proteger la información almacenada e instalada en los medios de cómputos. En este marco, el presente trabajo tiene como propósito aplicar técnicas de selección de características, métricas de aprendizaje y reducción de dimensión en sistemas de detección de intrusos, utilizando los datos almacenados en el DATASET NSL-KDD, el cual contiene 225.000 registros de conexiones en una red de computadores con 41 características.

SEGURIDAD INFORMÁTICA

La seguridad informática comprende el cumplimiento de las premisas de confidencialidad, integridad y disponibilidad en un sistema informático. Esta se fundamenta en una serie de elementos conceptuales que es necesario detallar para una mayor comprensión [1].

Premisas básicas de seguridad informática

Un sistema se considera seguro si cumple con las propiedades de integridad, identificación, control de acceso, no repudio, confidencialidad y disponibilidad de la información. Cada una de estas propiedades conlleva la implementación de determinados servicios y mecanismos de seguridad, que se describen a continuación:

- Integridad

Este principio garantiza la autenticidad y precisión de la información sin importar el momento en que se solicita, con otras palabras, es una garantía de que los datos no han sido alterados ni destruidos sin autorización.

- Confidencialidad

Se define como "el hecho de que los datos o la información esté únicamente al alcance de las personas, entidades o mecanismos autorizados, en momentos autorizados y de una manera autorizada" [2].

- Disponibilidad

La disponibilidad es el "grado en el que los datos están en el lugar, momento y forma en que es requerido por el usua-

rio autorizado", situación que se produce cuando se puede acceder a un sistema de información en un periodo de tiempo aceptable. La disponibilidad está asociada a la fiabilidad técnica de los componentes del sistema de información.

- Autenticación (identificación)

El sistema debe ser capaz de verificar que un usuario identificado, que accede a un sistema o que genera una determinada información, es quien dice ser. Solo cuando un usuario o entidad ha sido autenticado, podrá tener autorización de acceso. Se puede exigir autenticación en la entidad de origen de la información, en la de destino o en ambas [3].

- No repudio o irrenunciabilidad

Proporciona al sistema una serie de evidencias irrefutables de la autoría de un hecho. El no repudio consiste en no poder negar haber emitido una información que efectivamente se ha emitido y en no poder negar su recepción cuando ha sido recibida.

Taxonomía de los ataques informáticos e intrusiones

El modelo de análisis de incidentes planteado por computer emergency response team-cert y descrito en los ataques informáticos contenidos en el DATASET DARPA, comprende cuatro categorías [4]:

- Denegación de servicio

Denominado también por sus siglas en inglés denial of service (DOS), constituyen un conjunto de ataques que conllevan a detener el funcionamiento de una red, máquina, proceso o servicios a usuarios autorizados debido a la sobrecarga de los recursos computacionales de la víctima [5].

- Remote to local (r2l)

Se origina cuando un atacante informático que no posee acceso a alguna máquina, logra acceder a dicho equipo ya sea como usuario común o root, utilizando algún método de intrusión o programas [6].

- User to root (u2r)

Ocurre cuando un atacante, que tiene una cuenta en un sistema informático, adquiere privilegios superiores a los inicialmente establecidos sin autorización del administrador de ti, ejecutando alguna técnica de intrusión que se basa en una determinada vulnerabilidad del sistema informático

- Probing

Es un conjunto de ataques que se caracteriza por sondear la red de la víctima para recopilar información importante de los host que la incluye sin ser detectada, previéndole al atacante información necesaria para tener una lista de vulnerabilidades potenciales y llevar a cabo un ataque informático a los servicios como a las maquinas que lo ejecutan [7].

SISTEMAS DE DETECCIÓN DE INTRUSOS

Los IDS son una medida de seguridad que ayuda a identificar un conjunto de acciones malintencionadas que comprometen la integridad, confidencialidad y disponibilidad de los recursos de informáticos. La función principal de los IDS es proteger la información de las organizaciones frente a cualquier amenaza. Estas se han visto reflejadas en una creciente ola de ataques informáticos debido al incremento en la demanda del uso de las redes, el Internet y la dependencia de los sistemas de información [8].

Clasificación de los sistemas de detección de intrusos

Los IDS se clasifican de acuerdo con la

estrategia o tipo de análisis, fuente de información, arquitectura o estructura, respuesta o comportamiento o tipo de predicción. La Figura 1 detalla el esquema de clasificación de los IDS [8-9].

Figura 1. Clasificación de IDS

A. Las metodologías de detección de intrusos

Se dividen en 2 categorías: detección basada en abuso, detección basada en anomalías y análisis de protocolo de estado [8-9]:

- Detección basada en abusos: también se conoce como detección basada en el conocimiento, detección de uso incorrecto o detección basada en firmas. Este tipo de detección analiza la actividad del sistema en busca de uno o varios eventos que coincidan con un patrón predefinidos y que describan a un determinado ataque. Tales eventos son conocidos como firmas, las cuales son patrones o cadenas que corresponden a un ataque o amenaza conocida. En el caso de que se presente un ataque que coincida con una o varias firmas, se genera una alarma [9].

- Detección basada en anomalías (ad): una anomalía es una desviación a un comportamiento conocido, y los perfiles representan los comportamientos normales o esperados, que se derivan del seguimiento de las actividades regulares, conexiones de red, hosts o usuarios, durante un período de tiempo. Los perfiles se dividen en dos (2) tipos: estáticos o dinámicos. Ambos son registros del sistema que se alimentan de una serie de atributos o características, como son: el número de intentos fallidos de inicio de sesión, el uso del procesador, el recuento de los correos electrónicos enviados, entre otros.

Fundamentos relativos a la evaluación de los IDS

en la actualidad no existe un IDS 100% efectivo que clasifique a la perfección el tráfico normal del malicioso, debido a que hay una gran variedad de ataques que se incrementan con el pasar del tiempo, siendo cada vez más novedosos y desconocidos para los IDS. A esto se suman las malas prácticas en tecnologías de la información. Todo ello puede generar que un IDS tome decisiones incorrectas, en procesos de clasificación del tráfico de red. Por ejemplo: identificar ataques como tráfico normal. Para evaluar el desempeño de un IDS, se han identificado cuatro métricas asociadas a la naturaleza del evento (tráfico inofensivo o ataque) y al estado de la detección (normal o anómala), tal como se aprecia en la Figura 2

Figura 2. Matriz de confusión

Investigación e Innovación en Ingenierías, volumen V (número 1) ,Enero - Julio 2017 PP. 1 - 107 ISSN 2344-8652

Un IDS
Es más eficiente cuando durante el proceso de clasificación del tráfico de datos presenta mayores tasas de aciertos (es decir, que el porcentaje de verdaderos negativos y verdaderos positivos tiende a 100%) y, consecuentemente, presenta bajas tasas de fallos (es decir, el porcentaje de falsos positivos y falsos negativos tiende a 0%). A partir de lo anterior, se concluye que un IDS perfecto es aquel que detecta todo el tráfico de forma correcta, sin generar ninguna falsa alarma. De lo anterior se puede inferir [10]:

- verdaderos positivos (VP): ataque correctamente detectado como anomalía.

- falsos positivos (FP): tráfico inofensivo detectado de forma incorrectamente como anomalía.

- verdaderos negativos (VN): tráfico inofensivo correctamente identificado como tráfico normal.

- falso negativo (FN): ataque identificado incorrectamente como tráfico normal.

Métricas de desempeño
En esta investigación se usan métricas de desempeño estadísticas, para medir el comportamiento del IDS en relación al proceso de clasificación; tales métricas se definen a continuación [10]:

- Sensibilidad

Capacidad que tiene un IDS para identificar resultados "verdaderos positivos" [13]:

$$Especificidad \quad \frac{VN}{VN+FN} \qquad (2.2)$$

- Especificidad

Capacidad que tiene un IDS de medir la proporción de "verdaderos negativos" que se han identificado correctamente [13]:

$$Especificidad \quad \frac{VN}{VN+FP} \qquad (2.2)$$

- Exactitud

Grado de cercanía de las mediciones de una cantidad (x) al valor de la magnitud real (y); es decir, a la proporción de resultados verdaderos (tanto verdaderos positivos como verdaderos negativos). Una exactitud del 100% significa que los valores medidos son exactamente los mismos que los valores dados.

$$Exactitud \quad \frac{VP \quad VN}{VP+FP \quad FN \quad VN} \qquad (2.3)$$

- Precisión

Define la proporción de verdaderos positivos contra todos los resultados positivos:

$$Precision \quad \frac{VP}{VP+FP} \qquad (2.4)$$

Proceso de simulación aplicado a los IDS
Para la efectividad en un proceso de detección de tráfico malicioso en una red computacional implementando un sistema de IDS que utilice técnicas de selección de características, algoritmos de aprendizaje y la calidad de métricas, es idónea la evaluación mediante la simulación por software en laboratorio.

Por ende, requiere la ejecución de varias fases, como se muestra en la Figura 3, a saber: elección de la colección de datos (DATASET), pre-procesamiento (parseo y normalización), selección de características, entrenamiento (training), clasifica-

ción (test) y evaluación de métricas.

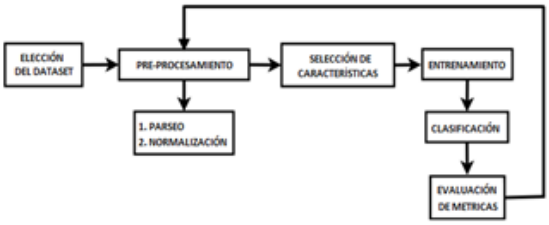

Figura 3. Fases de simulación

- Fase de elección de la colección de datos (DATASET)

En esta fase inicial se debe seleccionar la colección de datos que se va a utilizar para las siguientes fases. Existen distintas fuentes de datos utilizadas en sistemas de detección de intrusos, por ejemplo: PREDICT, DARPA KDD-NSL, caída, CRAWDAD, DRDC, NIST SAMATE y virtual DATASET REPOSITORY.

- Fase de pre-procesamiento

El pre-procesamiento es una fase previa a la selección o extracción de características en el proceso de simulación, en el cual se emplea una técnica de clasificación basada en redes neuronales artificiales. Esta fase permite homogenizar la presentación de los datos provenientes del DATASET e integrar esos datos contenidos en un formato diferente a la herramienta de simulación que se va a utilizar para el procesamiento de los datos.

- Fase de selección de características

Se define como el proceso de optimización que trata de encontrar el mejor subconjunto de características de un conjunto fijo de ellas [11].

Su objetivo es reducir el tamaño de los datos de entrada para facilitar el procesamiento y análisis, descartando datos que no contribuyen en mayor medida al posterior proceso de clasificación. Esto genera un ahorro de tiempo en el procesamiento de los datos, sin desestimar la generación de resultados óptimos.

- Fase de entrenamiento

En esta fase se procede a entrenar la red neuronal desde la implementación de un algoritmo de aprendizaje basado en los mapas auto-organizativos de KOHONEN -SOM, mapas auto-organizativos de jerarquía creciente-GHSOM y máquinas de soporte vectorial -SVM, entre otros, tomando como insumo todos los registros del DATASET KDD-TRAIN al 100%, proveniente de la aplicación de técnicas de selección de características.

- Fase de clasificación

Una vez entrenada la red neuronal, se procede con la fase de clasificación, la cual se realiza de forma autónoma, a partir de la implementación de un algoritmo de clasificación que determina el tráfico BI-CLASE (normal y anómalo), presentando la información de forma resumida y basada en fundamentos estadísticos. Una vez culminado el proceso anterior, se efectúa la prueba de aprendizaje, la cual se realiza usualmente con el DATASET KDD-TEST cargado al 100% de atributos.

- Fase de evaluación de métricas

en esta última fase, se realiza una evaluación de la calidad del modelo, mediante la implementación de un algoritmo que calcula cada una de las métricas utilizadas para el análisis de trafico de red (sensibilidad, especificidad, exactitud y precisión), basándose en los resultados obtenidos en la etapa de clasificación, con el propósito de conocer las características principales del modelo planteado.

Técnicas de selección de características en sistemas de IDS

La selección de características se refiere a un concepto utilizado en minería de datos, con el objetivo de reducir el tamaño de los datos de entrada para facilitar el procesamiento y análisis de dicha información. La selección de características no solo tiene en cuenta la disminución de la cardinalidad, es decir, el mantenimiento de un límite parcial o predefinido en la cantidad de atributos tenidos en cuenta al crear un modelo, también permite descartar de forma adecuada los atributos en función de la utilidad para la realización de un buen proceso de análisis.

- CHI-SQUARE

Es una técnica de análisis útil para determinar las reglas de asociación estadística, tomando en cuenta que las reglas de asociación son una técnica popular para producir calidad en las detecciones basadas en mal uso (MISUSED-BASED); sin embargo, estas reglas de asociación tienen se multiplican a menudo y reducen así el rendimiento de los IDS [12].

$$X = \sum \frac{(0-E)^2}{E} \quad [1]$$

Como se percibe en la ecuación [1], el cálculo estadístico de CHI-SQUARE depende de una pareja de variables. Esto implica la construcción de tablas de contingencia que se utilizan para examinar la relación entre las variables o para explorar la distribución de una variable categórica entre diferentes muestras.

Redes neuronales GHSOM (Growing Hierarchical Self Organizing Maps)

GHSOM es una estructura jerárquica y dinámica, desarrollada para superar las debilidades y problemas que presenta SOM. La estructura GHSOM consiste en múltiples capas compuestas de varias SOM independientes, cuyo número y tamaño se determinan durante la fase de entrenamiento. El proceso de crecimiento de adaptación es controlado por dos parámetros que determinan la profundidad de la jerarquía y la amplitud de cada mapa. Por lo tanto, estos dos parámetros son los únicos que deben fijarse inicialmente en GHSOM [13].

Este tipo de mapas son una versión mejorada de la arquitectura SOM y hay dos propósitos para su arquitectura [14]:

- SOM tiene una arquitectura de red fija, es decir. el número de unidades de uso, así como la distribución de las unidades, debe determinarse antes del entrenamiento.

- Los datos de entrada que son de naturaleza jerárquica deben representarse en una estructura jerárquica similar para mayor claridad de la representación. GHSOM utiliza una estructura jerárquica de varias capas, donde cada capa está formada por un número de SOM independientes. Pero solo se utiliza un SOM en la primera capa de la jerarquía.

Por cada unidad del mapa, un SOM podría añadirse a la siguiente capa de la jerarquía. Este principio se repite con el tercer nivel del mapa y las demás capas de la GHSOM, tal como se muestra en la Figura 4.

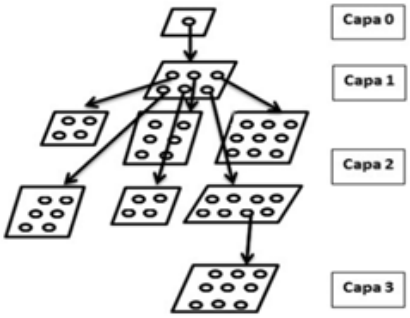

Figura 4. Estructura de una red ghsom

- Algoritmo de aprendizaje de GHSOM

La etapa de entrenamiento del algoritmo GHSOM empieza con la configuración inicial. En este paso se crea el "mapa" en el nivel 0 con una sola unidad [15].

El vector m0 de pesos de esta unidad es inicializado con la media de todos los vectores de entrada, y también se calcula el error medio de cuantificación mqe0. Con posterioridad a este procedimiento de inicialización, viene el proceso de formación y crecimiento del mapa.

El crecimiento de la estructura inicia con la creación de un nuevo SOM debajo de la capa 0 del mapa, con un tamaño inicial de 2x2 unidades. el proceso de crecimiento continúa hasta que el error medio de cuantificación, conocido como mqe, en mayúsculas, alcanza una cierta fracción τ1 del mqeu de la unidad correspondiente en la capa superior; es decir, la unidad que constituye la capa 0 del mapa para la primera capa de mapa.
La conceptualización descrita en los capítulos anteriores ha servido de fundamento para plantear diferentes escenarios de simulación con el propósito de definir un modelo de detección de intrusiones en sistemas de redes computacionales, basado en las fases de entrenamiento, clasificación y cálculo de métricas.

MODELO DE IDS BASADO EN TÉCNICAS DE SELECCIÓN Y CLASIFICACIÓN
El modelo que se usará para la detección de intrusos basados en anomalías de red comprende seis fases: (1) selección del conjunto de datos, (2) pre-procesamiento, (3) selección de características, (4) entrenamiento, (5) clasificación y (6) cálculo de métricas de desempeño. Para su aplicación se implementaron varios escenarios de simulación, variando la cantidad de características a evaluar en las fases de entrenamiento y clasificación y priorizando la escogencia de las características mediante la implementación de los métodos de selección de características CHI-SQUARE. La Figura 5 presenta el esquema funcional del modelo, cuyos elementos se describen a lo largo del capítulo.

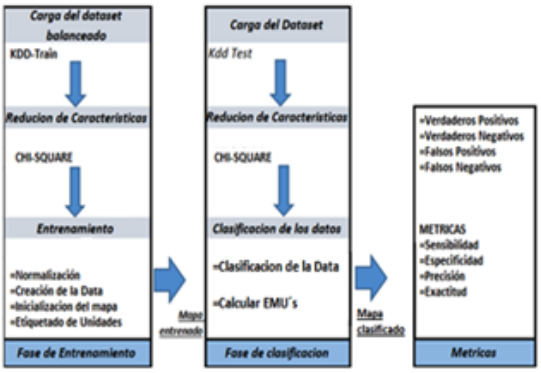

Figura 5. Modelo funcional propuesto

En la propuesta aquí descrita, las 41 características del conjunto de datos NSL-KDD son utilizadas en el algoritmo SOM y GHSOM, según el caso, para el cálculo de las distancias euclidianas. Por tanto, la escala de estas variables es muy importante para determinar la organización topológica del mapa. Si el rango de valores de una variable es mucho más grande que

el de las otras, esta probablemente dominará la organización del mapa.

El balanceo de datos se efectuó por tipo de conexión BI-CLASE (tráfico normal y ataque), con lo cual se buscó un equilibrio entre el número de conexiones que hacen referencia al trafico normal y a los ataques, dado que si en el entrenamiento a un algoritmo de aprendizaje se le suministra muchas más conexiones de un determinado tipo, es posible que durante la fase de clasificación tienda a dar mayor respuesta a un determinado tipo de conexión ya que aprendió la red neuronal más características de esa conexión en la etapa de entrenamiento. Cabe acotar que solo se balanceo el DATASET KDD-TRAIN, que fue utilizado para la fase de entrenamiento, por la razón anteriormente comentada.

Escenario experimental Tabla 1 (conjunto completo de características)
En este escenario se consideran las 41 características del DATASET KDD-TRAIN, a, y se realiza la clasificación utilizando el DATASET KDD-TEST con todas sus características.

Escenario	No. características	Exactitud	Sensibilidad	Especificidad	Precisión
GHSOM	41	60.27%	29.02%	93.31%	81.6%

Tabla 1. Resultados escenario experimental SOM y GHSOM sin selección de atributos

Escenario experimental tabla No 2 (conjunto de características seleccionadas, clasificando con GHSOM)
En este escenario experimental se utilizaron las técnicas de selección de características, CHI-SQUARE. Al aplicarlas se pudo efectuar reducción de características, obteniendo un DATASET más depurado, con el cual se realizó el entrenamiento del GHSOM. El DATASET específicamente usado para el proceso de entrenamiento ha sido el KDD-TRAIN al 100%. Se desarrolló una simulación tomando el DATASET KDD-TRAIN 100% y aplicando la técnica de selección de características INFO. GAIN. Con esta técnica, se identificó el orden de prioridad de las características del DATASET, lo cual permitió variar el número de ellas, generando pruebas con 5, 10, 11, 15-20, 30 y 41 características, y enfatizando en el intervalo de 15-20 características como lo sugiere la literatura. Como se observa en la Tabla 2, al combinar la técnica de INFO.GAIN con la técnica de clasificación GHSOM los mejores resultados en relación a las métricas exactitud 95,20%, y precisión 95,20%, se obtuvieron con 15 características. Pero, si bien la mejor tasa de especificidad, 94,94%, se obtuvo con 5 características, con una diferencia porcentual de +0,18% respecto a la obtenida con 15 características; y la mejor tasa de sensibilidad, 96,11%, se obtuvo con 20 características, con una diferencia porcentual de +0,52% respecto a la obtenida con 15 características; prevalece la mejor solución parcial de este conjunto de pruebas de simulación, esto es, la obtenida con 15 características, dado que el criterio de mayor relevancia es la métrica exactitud y que las diferencias porcentuales con las otras métricas no son tan significativas.

Escenario No. características exactitud sensibilidad especificidad precisión

Escenario	No. características	exactitud	sensibilidad	especificidad	precisión
CHI SQUARE + GHSOM+ V.C.	5	94,87%	94,81 %	94,94 %	94,90%
	10	93,78%	94,41 %	93,06 %	93,80%
	11	93,81%	94,41 %	93,13 %	93,80%
	15	95,20%	95,59 %	94,76 %	95,20%
	16	95,13%	95,57 %	94,63 %	95,10%
	17	95,13%	95,61 %	94,59 %	95,10%
	18	94,35%	96,09 %	92,47 %	94,40%
	19	94,45%	96,10 %	92,68 %	94,50%
	20	94,39%	96,11 %	92,53 %	94,40%
	30	94,72%	95,14 %	94,25 %	94,70%
	42	94,72%	95,14 %	94,24 %	94,70%

Tabla 2. Resultados de las pruebas de simulación aplicando INFO.GAIN + GHSOM con validación cruzada

CONCLUSIÓN

Acorde con la aplicación de técnicas de selección de características, métricas de aprendizaje y reducción de dimensión en sistemas de detección de intrusos, utilizando los datos almacenados en el DATASET NSL-KDD, el cual contiene 225.000 registros de conexiones en una red de computadores con un total de 41 características, se infiere que al utilizar el clasificador GHSOM con la técnica CHI SQUARE como técnica de selección de características, su mejor resultado es a 15 características con precisión, sensibilidad, especificidad y exactitud.

REFERENCIAS

[1] Russell & Gangemi. Computer Security Basics, 1991

[2] R. Feiertag, C. Kahn, P. Porras, D. Schackenberg, S. Staniford – Chen and B. Tung. A common instrusión specification language, 1999

[3] M. Piattini & E. peso, Auditoría informática: un enfoque práctico, 2001

[4] A. Howard. Badland Morphology and Evolution: Interpretation Using a Simulation Model, 1997.

[5] D.J. Marchette. Computer Intrusion Detection and Network Monitoring: A Statistical. New York: Viewpoint Springer – Verlag, 2001.

[6] M. Sabhnani and G. Serpent, "Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context," in Proc the International Conf. on Machine Learning: Models, Technologies, and Applications, Las Vegas, vol. 1, pp. 209-215, 2003.

[7] S. T. Brugger, and J. Chow. An assessment of the DARPA IDS Evaluation Dataset using Snort. Technical Report CSE-

2007-1, University of California, Davis, Department of Computer Science, 2007.
[8] P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, and P. Tan. "Data mining for network intrusion detection". NSF Workshop on Next Generation Data Mining, 2002.
[9] R. Bace & P. Mell. Intrusion detection systems. NIST special publication in intrusión detection systems.
[10] J. Andersen, S. Glasdam y D. Larsen. "New Concepts of Quality Assurance in Analytical Chemistry: Will They Influence the Way We Conduct Science in General?". Chemical Engineering Communications, Vol. 203, No. 12, p. 1582-1590, 2016.
[11] H. Hota, and A.K. Shrivas. "Data mining approach for developing various models based on types of attack and feature selection as intrusión detection systems (IDS)". Intelligent Computing, networking, and informatics, pp. 845-851, 2014.
[12] A. F. Namik & Z. A. Othman. Reducing network intrusion detection association rules using Chi-Squared pruning technique. Conference on Data Mining and Optimization, 2011.
[13] E. Pampalk, A. Rauber, D. Merkl. "Content-based organization and visualization of music archives", 2002.
[14] D. Merkl, M. Dittenbach, A. Rauber. The Growing Hierarchical Self – Organizing Map. In: S. AMARL et al. (Eds): Proceedings of the International Joint Conference on Neural Networks, 2000.
[15] M. Dittenbach, D. Merkl, and A. Rauber. The Growing Hierarchical Self-Organizing Map. In Amari, S., Giles, C. L., Gori, M., and Puri, V., editors, Proc of the International Joint Conference on Neural Networks (IJCNN 2000), volume VI, pages 15 – 19, Como, Italy. IEEE Computer

Este artículo se cita:

J. Mardini y A. Egea, "Modelo de detección de intrusiones en sistemas computacionales, realizando selección de características con CHI SQUARE, entrenamiento y clasificación con GHSOM", Revista Investigación e Innovación en Ingenierias,vol. 5, n°1. pp. 24-35, 2017.