Los sistemas inmunes artificiales en la detección de ciberataques

David **Limón Cantú**
Vicente **Alarcón Aquino**

La ciberseguridad es el área de estudio de problemas adversos en un contexto informático que afectan la confidencialidad, integridad y disponibilidad de los recursos informáticos. Estos ataques se originan por medio de vulnerabilidades inherentes en los programas y sistemas computacionales (no existe programa ni sistema perfecto), y pueden ser errores de programación, configuración o diseño.

El ciberatacante es aquel individuo (o grupo) que busca obtener acceso a algún espacio cibernético (como una red de computadoras) sin los permisos ni la autoridad adecuados. Los ciberatacantes maliciosos que obtienen acceso a alguna red de computadoras pueden causar grandes daños, así como obtener información confidencial.

De forma general, los ciberatacantes siguen una serie de pasos básicos para ganar acceso a una red: reconocimiento o exploración y escaneo, aprovechamiento de vulnerabilidades, establecimiento de puntos de espionaje y obtención de resultados.

Los ciberatacantes usan una serie de herramientas que aprovechan vulnerabilidades en las redes y equipos identificados. De forma general, los ataques más populares se dividen en cuatro categorías:

1) Escalado de privilegios**,** suplantar y escalar privilegios otorgados a usuarios legítimos.

2) Ataques a la aplicación, aprovechan vulnerabilidades en los programas que dan soporte a los servicios accesibles desde una red de computadoras.

3) Negación de servicio, busca saturar las redes de comunicaciones enviando grandes cantidades de solicitudes a los equipos de cómputo en la organización atacada.

4) Ataques a los protocolos de comunicación para afectar o infiltrarse en equipos de cómputo aprovechando sus vulnerabilidades de comunicación de la red.

En la actualidad, la mayoría de las empresas, organizaciones gubernamentales, cuerpos militares y dispositivos de cómputo de uso cotidiano (como teléfonos celulares, cámaras de vigilancia, dispositivos de Internet de las cosas, etc.) hacen uso de redes de comunicaciones y del Internet. Algunas de las tendencias del uso de ciberataques son el secuestro de datos (conocido en inglés como *ransomware*), la negación de servicios importantes (como servicios gubernamentales), el sabotaje de servicios públicos (suministros de agua y electricidad) y los chantajes por obtención de datos privados.

LOS SISTEMAS DE DETECCIÓN DE INTRUSOS

Los sistemas de detección de intrusos (*Intrusion Detection System*, IDS, por sus siglas en inglés) son sistemas computacionales que tienen la finalidad de detectar anomalías por medio del análisis y monitoreo de procesos computacionales y de redes de comunicaciones. Un sistema computacional se compone de distintos elementos de *software* (programas) y *hardware* (equipo de cómputo), combinados para efectuar una o más funciones con una finalidad específica.

En el contexto de los ciberataques, el objetivo de un IDS es facilitar la toma de decisiones para su identificación, mitigación y análisis. Los IDS se pueden clasificar en dos grandes grupos: los IDS basados en redes de comunicaciones (o *Network* IDS, IDS, por sus siglas en inglés), y los IDS basados en equipos de cómputo (*Host* IDS, HIDS, por sus siglas en inglés).

El objetivo de los NIDS es detectar anomalías a través de las comunicaciones en redes de computadoras. Los HIDS son sistemas que detectan anomalías por medio del análisis de distintos elementos presentes en el equipo de cómputo, como eventos del sistema operativo, sistemas de archivos, ejecución de procesos computacionales, entre otros.

Adicionalmente, los IDS pueden identificar ciberataques por medio de dos enfoques: 1) por medio de firmas, el cual consiste en comparar el comportamiento observado (ya sea en las redes de comunicaciones o en eventos del equipo de cómputo) con patrones previamente identificados. 2) Con base en anomalías, en donde se busca encontrar comportamientos que difieran significativamente de patrones previamente identificados como normales. De forma general, un NIDS se compone de seis fases denominadas fase de captura de paquetes, fase de extracción de características, fase de preprocesamiento, fase de detección, fase de clasificación y fase de alerta, como se muestra en la Figura 1.

Los retos de los IDS consisten en desarrollar sistemas capaces de efectuar la detección de una manera eficiente, con capacidad de manejar grandes volúmenes de comunicaciones y de ejecutarse en sistemas computacionales con recursos limitados (ya sea energéticos o de capacidad de cómputo).

LOS SISTEMAS INMUNES ARTIFICIALES

Los sistemas inmunes artificiales (*Artificial Immune Systems*, AIS, por sus siglas en inglés) son un conjunto de algoritmos (pertenecientes al área de algoritmos de aprendizaje automático) inspirados en el sistema inmune humano (*Human Immune System*, HIS, por sus siglas en inglés).

A diferencia de otros algoritmos inspirados en funciones biológicas o fisiológicas (como los algoritmos genéticos y las redes neuronales), la motivación para el desarrollo de los AIS reside en el deseo de imitar características importantes del HIS, como la adaptabilidad, la tolerancia a errores, la descentralización y la robustez.

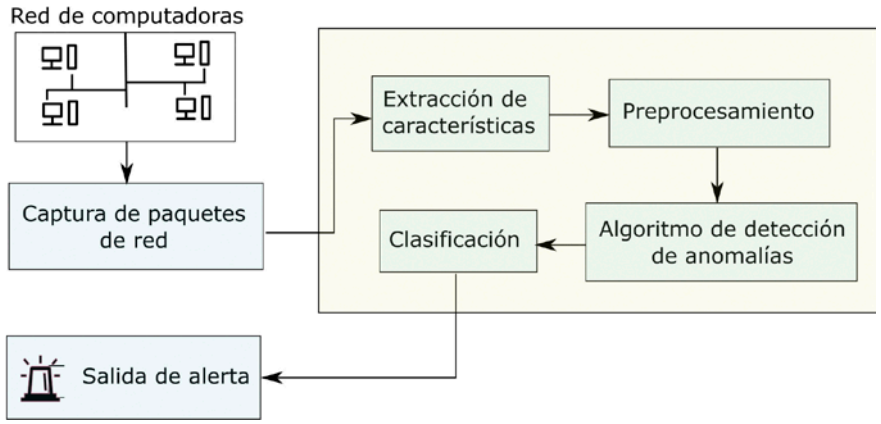

Figura 1. Modelo genérico de un IDS (Dwivedi *et al.*, 2020). El flujo indicado por las flechas representa el orden de las fases que se ejecutan en un IDS. La mayoría de los IDS buscan dar una alerta de ataque al administrador de la red computacional.

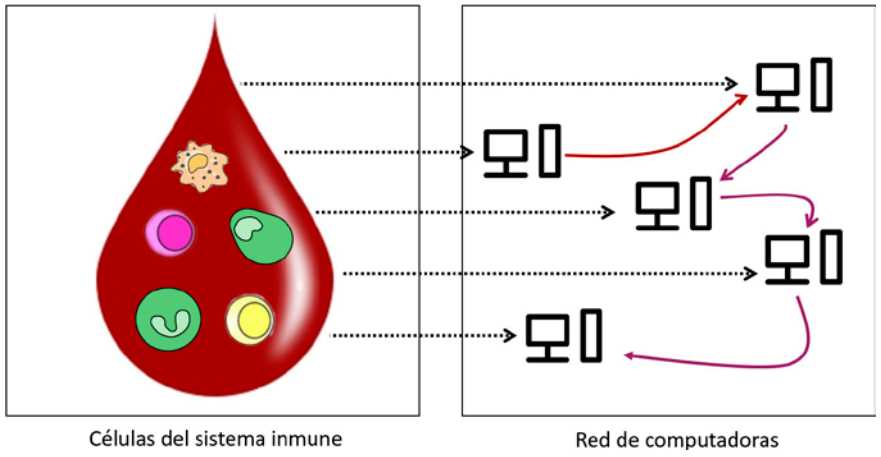

Figura 2. Analogía entre el HIS y los AIS (Limon-Cantu y Alarcon-Aquino, 2021). Los modelos AIS imitan el comportamiento de células del sistema inmune innato y adaptivo. Los modelos computacionales se ejecutan en una o más computadoras, las cuales forman parte de una red de comunicaciones.

El sistema inmune humano se compone de un conjunto de órganos y células cuya finalidad es proteger a nuestro cuerpo contra enfermedades, infecciones y el mal funcionamiento de otras células. El HIS se compone de tres capas: la física, la innata y la adaptiva. Como primera línea de defensa, la capa física protege a la mayoría de los órganos de nuestro cuerpo contra substancias foráneas (por ejemplo, toxinas, químicos, bacterias y virus). La capa física se compone de la piel y las membranas mucosas presentes en el tracto respiratorio, gastrointestinal y genitourinario. Cuando la primera capa de defensa es superada (por ejemplo, cuando nos cortamos o sufrimos una quemadura), las capas innata y adaptativa nos proporcionan la protección más importante en contra de microorganismos externos y dañinos. El sistema inmune innato incluye células (por ejemplo, células dendríticas) que son capaces de reconocer distintas amenazas, así como de presentar antígenos (patrones moleculares provenientes de organismos externos) al sistema inmune adaptativo. Las células de este último (por ejemplo, los linfocitos T y B) proporcionan la capacidad de reconocer antígenos no vistos antes, así como de recordarlos para generar rápidamente una respuesta inmune en el próximo encuentro.

Podemos asociar al HIS con los IDS a través de la detección de anomalías. Las células del sistema inmune tienen la función exclusiva de proteger nuestro cuerpo por medio de un proceso colaborativo entre las distintas células del sistema innato y adaptativo, como se muestra en la Figura 2. Ambos sistemas (innato y adaptativo) son inspiración para la creación de IDS por medio de la aproximación al comportamiento de una o más células inmunes. En el caso de los IDS, el comportamiento de las células inmunes tiene la finalidad de observar las comunicaciones en una red de computadoras (ver Figura 2).

ANTECEDENTES

El área de investigación de los AIS es un esfuerzo multidisciplinario que combina el estudio de inmunología, ciencias de la computación, estadística e ingeniería. Los AIS se han basado en principios de la inmunología como la selección negativa (Belhadj aissa *et al*., *2020*), las redes inmunes (Shi *et al*., 2017), la selección clonal (Elshafie *et al*., 2019), la teoría del peligro y el comportamiento de las células dendríticas (Farzadnia *et al*., 2020; Limon-Cantu y Alarcon-Aquino, 2021, 2022).

Una de las contribuciones más recientes en el área de los AIS son los algoritmos basados en la teoría del peligro, la cual establece que la respuesta inmune se activa en gran medida por medio de señales de peligro emitidas por células dañadas. De manera similar, las células saludables producen señales de calma. Dichas señales son recolectadas por células presentadoras de antígenos, como las células dendríticas, y son usadas para incitar o suprimir la respuesta inmune adaptiva.

La contribución más significativa basada en la teoría del peligro (Greensmith y Aickelin, 2008) se basa en el estudio de las células dendríticas, y se conoce como el algoritmo de células dendríticas (*Dendritic Cell Algorithm*, DCA, por sus siglas en inglés). El DCA es un algoritmo basado en el mecanismo de recolección de señales de las células dendríticas (*Dendritic Cells*, DC, por sus siglas en inglés), donde un conjunto de DC puede reprimir o estimular la respuesta inmune, proporcionando la información necesaria a las células del sistema inmune adaptativo, para proliferar la creación de células especializadas (T y B) capaces de reconocer las amenazas actuales.

El algoritmo resultante se compone de una población de DC artificiales que simulan su proceso de recolección y maduración. Posteriormente, se lleva a cabo un consenso para determinar el estado de anomalía y (en el caso de la seguridad en redes de computadoras) poder determinar la presencia de ciberataques.

Actualmente, la investigación alrededor del DCA ha sido incorporada en el estudio de los IDS, y se centra en resolver tres problemas: la automatización de la fase de extracción y selección de características, la implementación de mejoras en las fases de detección y de clasificación. Los objetivos principales de las mejoras en el estado de la ciencia son la reducción en la complejidad computacional, la incorporación de mecanismos robustos de aprendizaje automático, y la demostración de la efectividad del algoritmo, analizando ciberataques contemporáneos en escenarios realistas.

AVANCES EN EL ALGORITMO DE CÉLULAS DENDRÍTICAS

Este artículo analiza la implementación de dos modelos basados en el DCA. La importancia de dichos modelos ha sido la resolución de los tres problemas principales en el área. El primer modelo (Limon-Cantu y Alarcon-Aquino, 2022) se basó en la variación determinista del DCA y se propusieron algunas modificaciones. La primera fue la incorporación de un mecanismo automatizado para selección de características, basado en un proceso probabilístico conocido como información mutua. Este proceso analiza la entropía de las características obtenidas en la segunda fase del modelo general de un IDS (ver Figura 1), y ayuda a determinar si dichas características se pueden usar como analogías de las señales de peligro y señales de calma, establecidas en la teoría del peligro. Las señales seleccionadas para cada categoría se combinaron para generar nuevas características, y estas, a su vez, se usaron como entrada para el DCA determinista.

La segunda modificación del primer modelo fue la incorporación de un árbol de decisión como reemplazo de un umbral de clasificación, usado en la fase de clasificación del modelo general de un IDS (ver Figura 1). Los árboles de decisión permiten generar un conjunto de reglas jerárquicas en forma de decisiones lógicas. En el caso del modelo analizado, estas reglas permiten determinar la presencia o ausencia de ciberataques analizando la métrica de anomalía resultante del DCA en la fase

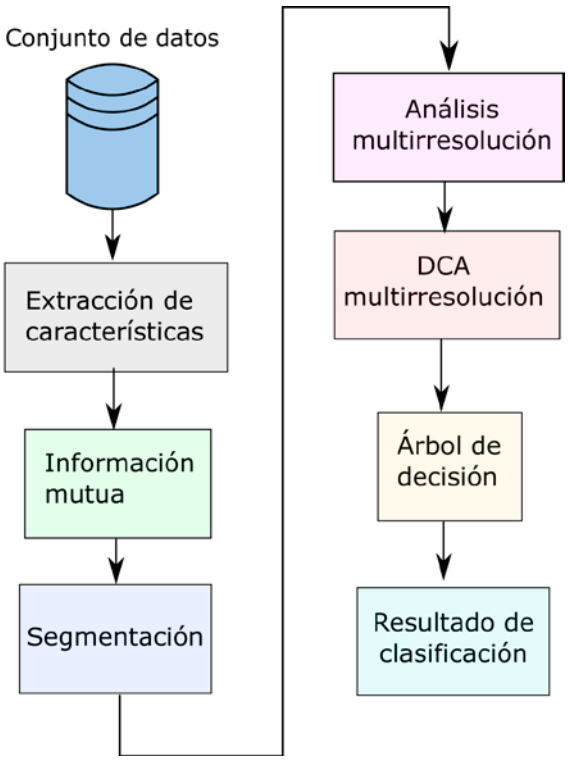

Figura 3. Fases de los modelos propuestos (Limon-Cantu y Alarcon-Aquino, 2021). Las fases de los modelos analizados son similares a las presentadas en la Figura 1. Las aportaciones se centran en las fases de selección de características (información mutua y segmentación), clasificación (análisis multirresolución), y clasificación (árbol de decisión). El resultado presentado indica si el dato analizado es un ataque o no.

de detección. Finalmente, este modelo fue evaluado y comparado usando conjuntos públicos de datos de ciberataques clásicos (conocido como NSL-KDD) y contemporáneos (conocido como UNSW-NB15), ambos publicados por el Instituto Canadiense de Ciberseguridad (CIC, por sus siglas en inglés).

El segundo modelo (Limon-Cantu y Alarcon-Aquino, 2021) tuvo la finalidad de mejorar la fase de detección en un IDS, y consistió en proponer modificaciones al algoritmo determinista del DCA, incorporando el análisis multirresolución como herramienta en la fase de preprocesamiento y en la fase de detección (ver Figura 3). El uso del análisis multirresolución permitió realizar un análisis en el espacio tiempo-frecuencia de las señales de peligro y las señales de calma generadas en la fase de extracción de características. Este modelo propone cuatro modificaciones. La primera consistió en la implementación de la transformada *wavelet* en el análisis multirresolución, con la cual se obtuvo una descomposición de las señales de peligro y calma. Las señales descompuestas fueron usadas como entrada para el DCA modificado.

La segunda consistió en analizar las señales descompuestas, resultantes del análisis multirresolución, dentro de la fase de detección. La tercera modificación fue la incorporación del enfoque segmentado para el DCA (Gu *et al*., 2013), que propone la idea de separar los datos analizados en segmentos de tamaño consistente para efectuar el proceso de detección y clasificación para cada segmento individual.

La segmentación fue incorporada en la fase de detección. Finalmente, se analizaron dos conjuntos de datos de ciberataques (NSL-KDD y UNSW-NB15), así como dos conjuntos de datos publicados en colaboración con el Establecimiento de Seguridad en Comunicaciones de Canadá (CSE, por sus siglas en inglés), abreviados como CIC-IDS2017 y CSE-CIC-IDS2018.

En general, los modelos analizados tienen cuatro limitaciones: la selección de los parámetros de las DC artificiales, la presencia y orden de los ataques en los conjuntos de datos analizados, la selección del tamaño de segmento, y la dificultad para realizar la clasificación usando datos no vistos. Estas limitaciones residen principalmente en la definición del DCA, ya que este fue modelado usando escenarios más simples en el área de los IDS. Actualmente, el DCA no cuenta con un mecanismo inherente de generalización, lo cual presenta un área de oportunidad para el desarrollo de versiones futuras. La Figura 3 muestra las fases de los modelos analizados.

CONCLUSIONES

Debido a la creciente dependencia de las redes de comunicaciones en sistemas computacionales, la detección de ciberataques es de gran importancia para mantener la seguridad e integridad de dispositivos que contienen información sensible. El desarrollo de sistemas de detección de intrusos facilita la toma de decisiones y la detección temprana de amenazas y ataques maliciosos.

© **Enrique Soto**. Serie "Mofles", 2007.

Uno de los retos más importantes del desarrollo de los sistemas de detección de intrusos es la capacidad de adaptarse a las tendencias emergentes en el campo de los ciberataques.

La creación de algoritmos computacionales basados en el comportamiento del sistema inmune humano ha sido de gran interés en la detección de ciberataques. Los algoritmos de detección de intrusos, basados en la teoría del peligro, han establecido un cambio fundamental en el estudio y modelado del sistema inmune humano para la detección de ciberataques.

La contribución más significativa se conoce como el algoritmo de células dendríticas (DCA) basado en el estudio de la interacción de señales emitidas por las células del cuerpo humano. El objetivo de este artículo de divulgación es dar a conocer los avances en la resolución de tres problemas principales en el desarrollo del DCA: la automatización de la extracción y selección de características, las mejoras en la fase de detección y las mejoras en la fase de clasificación. El primer modelo analizado se basó en la variación determinista del DCA. El segundo modelo analizado consistió en proponer modificaciones al algoritmo determinista del DCA, incorporando el análisis multirresolución.

Las limitaciones del último trabajo analizado residen principalmente en el estudio del algoritmo de células dendríticas, ya que no cuenta con un mecanismo inherente de clasificación en línea (aprendizaje en tiempo de ejecución), lo cual presenta un área de oportunidad para el desarrollo en versiones futuras.

REFERENCIAS

Belhadj aissa N, Guerroumi M and Derhab A (2020). NSNAD: negative selection-based network anomaly detection approach with

relevant feature subset. *Neural Computing and Applications 32*(8): 3475-3501. https://doi.org/10.1007/s00521-019-04396-2.

Dwivedi S, Vardhan M and Tripathi S (2020). Incorporating evolutionary computation for securing wireless network against cyberthreats. *The Journal of Supercomputing 76*(11):8691-8728. https://doi.org/10.1007/s11227-020-03161-w.

Elshafie HM, Mahmoud TM and Ali AA (2019). Improving the Performance of the Snort Intrusion Detection Using Clonal Selection. *2019 International Conference on Innovative Trends in Computer Engineering* (*ITCE*) 104-110. https://doi.org/10.1109/ITCE.2019.8646601.

Farzadnia E, Shirazi H and Nowroozi A (2020). A New Intrusion Detection System Using the Improved Dendritic Cell Algorithm. *The Computer Journal 64*(8):1193-1214. https://doi.org/10.1093/comjnl/bxaa140.

Greensmith J and Aickelin U (2008). The Deterministic Dendritic Cell Algorithm. En PJ Bentley, D Lee y S Jung (Eds.), *Artificial Immune Systems* (pp. 291-302). Springer Berlin Heidelberg.

© **Enrique Soto**. Serie "Mofles", 2012.

Gu F, Greensmith J and Aickelin U (2013). Theoretical formulation and analysis of the deterministic dendritic cell algorithm. *Biosystems 111*(2):127-135. https://doi.org/10.1016/j.biosystems.2013.01.001.

Limon-Cantu D and Alarcon-Aquino V (2021). Multiresolution dendritic cell algorithm for network anomaly detection. *PeerJ Computer Science 7*:e749. https://doi.org/10.7717/peerj-cs.749.

Limon-Cantu D and Alarcon-Aquino V (2022). Network Intrusion Detection Using Dendritic Cells and Danger Theory. *Technology, Science and Culture: A Global Vision 3*(1):89. https://doi.org/10.5772/intechopen.99973.

Shi Y, Peng X, Li R and Zhang Y (2017). Unsupervised Anomaly Detection for Network Flow Using Immune Network Based K-means Clustering. En B Zou, M Li, H Wang, X Song, W Xie y Z Lu (Eds.), *Data Science* (pp. 386-399). Springer Singapore.

GLOSARIO

Recursos computacionales: componente físico o informático (virtual) de un sistema computacional.

Sistema computacional: sistema que almacena y procesa información. Se compone de la parte física (hardware) y virtual (software).

Redes de comunicaciones: conjunto de elementos que se comunican entre si a través de un medio físico en común.

Sistema operativo: programa que conecta y administra los componentes físicos y virtuales de una computadora.

Sistema de archivos: método por el cual un sistema operativo almacena y recupera información.

Algoritmo: conjunto finito de operaciones que resuelven un problema en específico.

Aprendizaje automático: conjunto de técnicas que permiten a las computadoras aprender a través de experiencia analizando datos.

Complejidad computacional: estudio de los recursos (tiempo y espacio) requeridos por los programas computacionales.

Información mutua: proceso que mide cuanto conocimiento se puede obtener de una variable con respecto a otra.

Entropía: concepto que cuantifica el estado de desorden o incertidumbre de un sistema.

Análisis multirresolución: proceso que descompone una señal en componentes cuya suma produce la señal original.

Transformada wavelet: herramienta matemática que analiza las características en la frecuencia de una señal en el tiempo, permitiendo cuantificar los diferentes componentes de frecuencia a través del tiempo.

David Limón Cantú
Vicente Alarcón Aquino
Departamento de Computación Electrónica y Mecatrónica
Universidad de las Américas Puebla
vicente.alarcon@udlap.mx

© **Enrique Soto**. Serie “Mofles”, 2006.