UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

FACULTAD DE INGENIERÍA

PROGRAMA ACADÉMICO DE INGENIERÍA DE SISTEMAS DE INFORMACIÓN

Sistemas de control de supervisión y adquisición de datos para la deteccion de ciberataques en la industria minera

TRABAJO DE INVESTIGACIÓN

Para optar el grado de bachiller en Ingeniería de Sistemas de Información

AUTOR(ES)

Ore Huacles, Jonel Pelee (0000-0002-8483-0735)

Donaire Arbieto, Ilyan Eduardo (0000-0002-9540-1602)

ASESOR

Lomparte Alvarado, Romulo Fernando (0000-0003-3734-3966)

Lima, 09 de Mayo del 2021

RESUMEN

Los sistemas de control industrial son objetivos recurrentes de ciberataques con el fin de alterar procesos, detenerlos o secuestro de información. Existen distintas soluciones para contrarrestar estas actividades ilícitas, tales como los sistemas de detección de intrusos. Sin embargo, este tipo de soluciones no contemplan escenarios de trabajo específicos como plantas de procesamiento de minerales. El presente trabajo muestra un esquema a seguir para utilizar un modelo tecnológico capaz de utilizar tecnologías como machine learning y sistemas de detección de intrusos enfocados a plantas mineras.

Palabras clave: ciberataque; control industrial; minería; machine learning

ABSTRACT

Industrial control systems are recurring targets of cyberattacks in order to alter processes, stop them or hijack information. There are different solutions to counter these illegal activities, such as intrusion detection systems. However, these types of solutions do not contemplate specific work scenarios such as mineral processing plants. This work shows a scheme to follow to use a technological model capable of using technologies such as machine learning and intrusion detection systems focused on mining plants.

Keywords: cyber attack; industrial control; mining; machine learning

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS

1. INTRODUCCIÓN

En la industria, los sistemas de control industrial desempeñan un rol importante en la manufactura, gestión de productos y producción. Este tipo de sistema está conformado por una red de controladores programados (PLC) para la medición y regulación de las distintas máquinas en una planta, y un sistema de supervisión para los sensores que estos cuentan (SCADA). De esta forma, la data que fluye a través de dichos sistemas adquiere una connotación crítica para la empresa y el negocio entero, y que puede ser de interés a personas ajenas, esto a través, por ejemplo, a través de los ciberataques, el cual consiste en penetrar el ciberespacio de una empresa y robar o modificar datos críticos de la misma.

Sin embargo, los sistemas de control industrial presentan, en la mayoría de los casos, falta de protección perimetral, identificación y autenticación de usuarios, bloqueo de acceso físico a la red y gestión de cuentas. Algunos ciberataques para los sistemas industriales son TRITON, Trisis y HatMan, para el sector energético en Europa y EE. UU es dado por el grupo hacker Dragonfly / Energetic Bear, para una planta nuclear es ransomware NotPetya en Ucrania.

Una alternativa para contener los ciberataques son los sistemas de detección de intrusos (IDS), que es definido por National Institute of Standards and Technology (NIST) como “software o hardware que automatizan el proceso de monitoreo de los eventos que ocurren en un sistema informático o red, analizando en busca de señales de problemas de seguridad”. Por ello la mayoría de los estudios de seguridad en control industrial se basan en IDS, y estas a su vez en el uso algoritmos de machine learning (redes neuronales, bayesianas, árboles de decisiones, etc). Sin embargo, cada implementación presenta su propia concepción sobre los componentes básicos, entrada de datos, salida y servicios.

2. Estado del Arte

Un sistema de control de supervisión y adquisición de datos (SCADA) es un sistema compuesto por hardware y software para el control de procesos industriales locales o remotos; monitoreo, obtención y procesamiento de datos en tiempo real y recolección de registros de eventos. SCADA está compuesto por el servidor principal, soportado por una base de datos estructurada y proporciona una interfaz gráfica en los computadores de los usuarios finales, por debajo se encuentran los controladores lógicos programables (PLC) que a su vez se conectan a sensores específicos instalados en distintas máquinas dentro de una planta industrial.

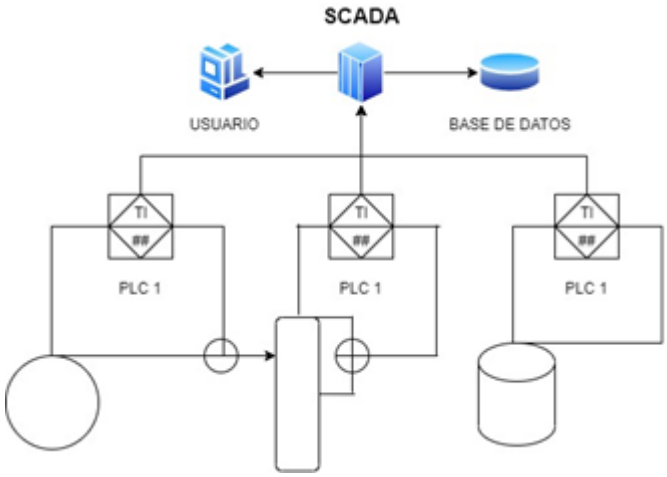

Ilustración 1: Arquitectura Scada

Para asegurar la seguridad de la información en plantas industriales que utiliza SCADA existen cuatro vías:

- **Software**: uso de antivirus, antimalware y antispyware. Esta solución es estrictamente reglamentaria en cualquier escenario de ciberdefensa y el objetivo es proteger el sistema de dispositivos USB maliciosos (“duckies”), virus por medio de acceso a internet y vectores de ataques basados en ransomware.
- **Arquitectura**: representado por 3 tecnologías: TSVs, arquitecturas C2 y S2A. El objetivo es aislar los componentes de menor nivel como los PLCs en un sistema de control industrial. Esto se logra agregando redes dedicadas de datos y aisladas del tráfico de datos convencional.

- **Sistemas de detección de intrusos (IDS):** el objetivo es aplicar un monitoreo constante a nivel de PLCs o del sistema SCADA, que sea capaz de determinar ciertos patrones predefinidos en cuanto a un ciberataque. Sin embargo, resulta complejo determinar cuál es el nodo vulnerado y rastrear el vector de ataque debido al gran volumen de datos.
- **Frameworks**: cuyo objetivo es buscar la resiliencia del sistema para lidiar con los ciberataques cotidianos, y es dado por políticas de uso de datos y de comportamiento de operarios, procedimientos sobre detección, tratamiento, respaldo y recuperación del ciberataque.

En contraste de esa actividad, se presentan soluciones en cuanto a ciberseguridad respecta para contrarrestar futuros ataques:

- Monitoreo orientado a nivel de procesos: el diseño de esta solución presenta una jerarquía para monitorear la actividad general a cada nivel del sistema de control industrial.
- Monitoreo orientado a procesos utilizando hardware-in-the-loop (HITL): esta tecnología propone crear diferentes emulaciones de procesos vulnerados o con comportamiento anómalo. La propuesta radica en colocar en un bucle cada componente del sistema de control industrial para poder comparar su comportamiento con el comportamiento histórico de incidentes de ciberseguridad. De esta forma es posible determinar si el sistema está siendo vulnerado, identificar el vector de ataque y el nodo atacado.

Los algoritmos de machine learning se dividen en 2 grupos, y ambos grupos contienen modelos determinísticos y probabilísticos:

- **Algoritmos supervisados:** necesitan ser entrenados constantemente para corroborar su eficacia. Uno de los algoritmos, y el más utilizado posiblemente, son las redes neuronales. Las ventajas de dicho algoritmo es el rápido desarrollo y definición de este, además de que se beneficia a mayor volumen de datos. El uso de este algoritmo es diverso, desde reconocimiento de datos codificados, hasta supervisión de redes y aplicaciones a nivel global.

- **Algoritmos no supervisados:** no necesitan ser entrenados, pues utilizan técnicas de clustering y auto codificadores, apoyando todo su análisis de machine learning. El algoritmo más utilizado en este grupo es el evolutivo, permitiendo mejorar en cada iteración. Básicamente este tipo de algoritmos se auto entrenan para lograr un objetivo concreto. A pesar de que los algoritmos son eficaces, no son eficientes en cuanto a tiempo y los resultados nunca son en tiempo real.

Por otro lado, se presenta el estudio de 6 algoritmos de machine learning:

- SVM: de todos los algoritmos, el más robusto y preciso en cuanto a detección de ciberataques. Es compatible con múltiples conjuntos de datos para su entrenamiento. Sin embargo, es un algoritmo de alta demanda computacional y relativamente lento en comparación con los demás algoritmos.
- KNN: preciso en cuanto a detección de ciberataques. Sin embargo, el set de datos es clave para la eficiencia de este algoritmo. Mientras mayor es el volumen de datos, el algoritmo se vuelve más pesado y la complejidad es exponencial.
- Árbol de decisión: algoritmo basado en clasificación de los datos y entrenamiento por paquetes de datos pequeños. Tiene la misma complejidad y problema que el algoritmo KNN, por lo que es descartado en un ambiente de continuo flujo de datos de PLC.
- DBN: es un modelo probabilístico, por lo que se define por funciones específicas que se construyen de acuerdo con los datos ingresados. Además, su clasificación es más precisa ya que no necesita un modelo de etiquetas para los datos.
- RNN: redes neuronales recurrentes. Se caracteriza por las múltiples entradas y variables que permite ingresar el sistema. De esta forma, el algoritmo es multipropósito. La complejidad computacional, al igual que el DBN, es ligera a comparación con las dos primeras opciones.
- CNN: redes neuronales convolucionales. De gran popularidad en los últimos años por su uso en el reconocimiento de lenguaje humano y procesamiento de imágenes. Es similar al anterior, pero tiene mayor capacidad para relacionar los datos ingresados.

Un caso de estudio hace uso de un algoritmo de aprendizaje dirigido por los datos que recibe (data-driven). En este sentido, las pruebas se basan en la robustez de dicho algoritmo frente al ruido en los datos como el consumo energético, el poder actual del reactor, etc. La metodología de este esquema es el siguiente:

Ilustración 2: Metodología del esquema

Por otro lado, un framework presenta 3 requerimientos para el sistema y 4 soluciones de ciberseguridad para la industria. Para poder probar la eficacia de este framework, se puso a prueba ataques en paralelo por todo el sistema de control industrial y la red de datos de la planta. Se utilizaron 6 máquinas (portátiles) con Kali Linux. Aprovechando el conjunto de herramientas de esta distribución de GNU/Linux, se apreciaron diversos vectores de ataque para cada intruso.

TC	Test Case Description	Attack Detail
TC-01	Kali Linux replaces PV smoothing usecase simulation	Nmap scan on TAC
		Data fuzzing of the usecase simulation
		Running DoS against the usecase in the presence of TAC
TC-02	Kali Linux replaces the microgrid controller of the MCS	Nmap scan on the controller's OT interface
		Information leak during Nmap scan
		Running DoS against controller's OT interface in the presence of Denelis
TC-03	Kali Linux inside OT VLAN of Field switch	Nmap scan on the controller with ACL
		Nmap scan on the controller without ACL
		DoS against the controller in the presence of TAC
		Running DoS against the controller without TAC
TC-04	Kali Linux inside OT VLAN of Bus switch	Nmap scan on the controller with ACL
		Sending duplicate IP address logs to the controller
		Running DoS against the controller in the presence of TAC
TC-05	Kali Linux replaces the controller in OT VLAN	Create TCP connection with the controller's application server
TC-06	Kali Linux accesses management interface of the controller's application server	TCP communication with controller's application server
		Nmap scan against controller's application server
		Running DoS to prevent legitimate logins to the application server
		Running Ncat listener and connecting to application server

Ilustración 3: Casos de prueba diseñados para evaluar los sistemas de control de micro red (MCS por sus siglas en inglés)

Uno de ellos es el estudio de CockpitCI que propone un framework de ciberseguridad para la detección de intrusos en sistemas de adquisición de datos

El módulo OCSVM que permite a los sistemas distribuidos operar de manera conjunta bajo un modelo de sistema de detección de intrusos (IDS).

El siguiente sistema propone un avance sobre los sistemas de detección de intrusos (IDS) para sistemas de control industrial, específicamente propone el uso de filtros para la red del sistema, utilizando restricciones temporales y combinatorias para mejorar la ciberseguridad en una planta industrial.

Por último, el siguiente sistema propone el uso de algoritmos de deep learning (un tipo de algoritmo de machine learning no supervisado) para poder detectar ciberataques en sistemas de IoT. Si bien no es específico para sistemas de control industrial, la similitud de implementación es alta y supone una característica principal de la industria 4.0.

Como se puede apreciar, las soluciones presentadas si bien logran su objetivo como IDS, en conjunto no representan un modelo a seguir para futuros proyectos. Esto se ve reflejado en los componentes básicos y servicios que cada uno presenta no se comparten entre los demás.

3. Aporte

Según lo investigado en relación con ciberseguridad para los sistemas de control industrial en el Perú aún se mantienen en el ámbito teórico o en estudio. Se presentan algoritmos de machine learning como posible respuesta ante el crecimiento de ataques cibernéticos, como se indica en la investigación dichos algoritmos presentan una robustez frente al ruido en los datos. También se han propuesto algunos modelos tecnológicos que ayudan a prevenir o detectar posibles escenarios de ciberataques los cuales se dividen en 4 vías distintas: Software, Arquitectura, Framework y Sistemas de detección de intrusos (IDS).

Basados en la investigación realizada se propone un modelo tecnológico que permite la detección de ciberataques en tiempo real y con ayuda de las redes neuronales se obtiene un modelo con aprendizaje continuo no supervisado, además de brindar las herramientas de gestión, como los reportes, indicadores y un Dashboard, para que el usuario pueda tomar decisiones rápidas. El modelo propuesto trabajaría en conjunto con el sistema SCADA y cuenta con un bus de datos dedicado para el flujo estándar del sistema mencionado.

Finalmente, se presentó un esquema como base a seguir para poder complementar el modelo desarrollado con futuras investigaciones, por ejemplo: Blockchain, el cual ayudaría a tener una red de confianza en la industria minera, fortaleciendo el flujo de datos en redes computacionales complejas (por ejemplo, redes con VPN para conectividad entre sedes distantes geográficamente). Por otra parte, el modelo presentado esta pensado en industrias que usan el sistema SCADA, el cual es muy común en el sector industrial, pero se debe desarrollar un modelo que se adapte a distintos sistemas de control, con mira a implementaciones en industrias tanto simples como complejas en su infraestructura.

4. Conclusiones

- Luego de realizar las investigaciones realizadas usando la técnica de benchmarking se determinó que las redes neuronales es el mejor algoritmo para que sea usado en el modelo tecnológico desarrollado, debido a que permite una respuesta en tiempo real y un aprendizaje continuo.
- A partir de la investigación realizada, se desarrolló un modelo tecnológico que usando redes neuronales permitirá detectar potenciales ciberataques en sistemas de control industrial para el sector minero.
- Se desarrolló el modelo tecnológico basado en machine learning para la detección de posibles ciberataques potenciales. Este modelo se encarga de procesar los datos en tiempo real, los cuales son obtenidos por medio de los componentes de bajo nivel (PLC y DCS), mediante redes neuronales. El modelo brinda un servicio de reportes que alerta de manera inmediata el ataque, así como presentar un Dashboard para ver el histórico de los posibles ciberataques detectados.
- Las validaciones del modelo se realizaron mediante una encuesta con expertos en las áreas que intervienen en el proyecto, así como un gerente de planta de una minera. La aceptación del producto presentado confirma la eficacia y capacidad de detectar ciberataques en sistemas de control industrial y alertar a usuarios claves para la rápida toma de decisiones.
- Un gerente de planta de una minera en Perú valido el modelo, confirmando que es aplicable al sector y además ayudaría en la seguridad ente posibles ciberataques.
- Como plan de continuidad se definió las buenas prácticas de ITIL, permitiendo a la mina mantener la solución disponible, así como la continuidad del negocio.
- Para la implementación del modelo desarrollado se deberá obtener la aprobación de la alta gerencia de la empresa

5. Referencias

- Hemsley, K. E., & E. Fisher, D. R. (2018). History of Industrial Control System Cyber Incidents. History of Industrial Control System Cyber Incidents. Published. https://doi.org/10.2172/1505628.
- Scarfone, K. A., & Mell, P. M. (2007). Guide to Intrusion Detection and Prevention Systems (IDPS). Guide to Intrusion Detection and Prevention Systems. Published. https://doi.org/10.6028/nist.sp.800-94.
- Rubio, J. E., Alcaraz, C., Roman, R., & Lopez, J. (2019). Current cyber-defense trends in industrial control systems. Computers & Security, 87, 101561. https://doi.org/10.1016/j.cose.2019.06.015.
- Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., & Wang, C. (2018). Machine Learning and Deep Learning Methods for Cybersecurity. IEEE Access, 6, 35365–35381. https://doi.org/10.1109/access.2018.2836950.
- I.A. (2018, 12 septiembre). What Is SCADA? Inductive Automation. https://www.inductiveautomation.com/resources/article/what-is-scada
- Cruz, T., Rosa, L., Proenca, J., Maglaras, L., Aubigny, M., Lev, L., Jiang, J., & Simoes, P. (2016). A Cybersecurity Detection Framework for Supervisory Control and Data Acquisition Systems. IEEE Transactions on Industrial Informatics, 12(6), 2236–2246. https://doi.org/10.1109/tii.2016.2599841.
- Cybersecurity for Control Systems: A Process-Aware Perspective. (2016, 1 octubre). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/7523254/
- Tonello, A. M., Letizia, N. A., Righini, D., & Marcuzzi, F. (2019). Machine Learning Tips and Tricks for Power Line Communications. IEEE Access, 7, 82434–82452. https://doi.org/10.1109/access.2019.2923321.
- A Multidimensional Holistic Framework for the Security of Distributed Energy and Control Systems. (2020, 1 marzo). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/8755282/
- Li, Y., Bertino, E., & Abdel-Khalik, H. S. (2019). Effectiveness of Model-Based Defenses for Digitally Controlled Industrial Systems: Nuclear Reactor Case Study.

Nuclear Technology, 206(1), 82–93. https://doi.org/10.1080/00295450.2019.1626170

- Sicard, F., Zamai, E., & Flaus, J. M. (2018). Filters based Approach with Temporal and Combinational Constraints for Cybersecurity of Industrial Control Systems. IFAC-PapersOnLine, 51(24), 96–103. https://doi.org/10.1016/j.ifacol.2018.09.541.
- Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761–768. https://doi.org/10.1016/j.future.2017.08.043.