Primer premio al mejor trabajo de estudiante: Detección de anomalías en redes industriales guiada por datos

Mikel Iturbe
Mondragon Unibertsitatea
Goiru 2, Arrasate-Mondragón
miturbe@mondragon.edu

***Resumen*—Este artículo resume el trabajo realizado en una tesis doctoral en el campo de la ciberseguridad industrial. Más concretamente, el trabajo realizado se ha centrado en la detección de anomalías guiada por datos en redes industriales. Las redes industriales –entornos interconectados por sistemas dedicados a automatizar, monitorizar y controlar procesos físicos– han evolucionado enormemente, mientras que los mecanismos de seguridad aplicables no han evolucionado al mismo paso, bien porque no escalan correctamente, bien porque no han tenido en cuenta las particularidades de este tipo de redes. Esta tesis doctoral se centra en desarrollar sistemas de detección de anomalías (SDAs) que utilizan los datos intrínsecamente creados en este tipo de redes (mediciones de campo, tráfico de red, registros...) para detectar eventos de seguridad.**

***Index Terms*—redes industriales, detección de anomalías, análisis de datos**

Tipo de contribución: *Premio al mejor trabajo de estudiante*

I. Motivación

Desde el desarrollo de los primeros Controladores Lógicos Programables (PLC) en la década de 1960, los sistemas de control industrial (SCI) han evolucionado considerablemente. Desde las primitivas instalaciones aisladas, los SCI están más conectados entre sí, hasta formar los entornos interconectados complejos conocidos como redes industriales (RIs) de hoy en día. Los SCI son responsables de un gran número de procesos físicos, desde plantas industriales de fabricación hasta la generación de energía, incluyendo procesos que pertenecen a infraestructuras críticas (ICs). Por ello, el correcto funcionamiento de las redes industriales es vital para preservar la actividad cotidiana de sociedades modernas, ya que gran parte del tejido ecónomico y del bienestar de éstas se sustentan en este tipo de redes.

En el campo de la ciberseguridad, los sistemas de detección de anomalías (SDAs) han tenido un rol referencial en la protección de este tipo de redes. Estos sistemas monitorizan el comportamiento de las RI y/o SCI para detectar eventos o comportamientos que se alejan del funcionamiento normal del entorno. Generalmente, estas propuestas se han basado en el aprendizaje automático o el análisis de datos para cumplir su función. Sin embargo, la creciente complejidad de las RIs ha derivado en los que los datos intrínsecamente creados en las RIs han crecido en volumen (más capacidad de almacenamiento de datos históricos), variedad (más variables monitorizadas) y velocidad (más lecturas) convirtiéndose así en un problema *Big Data*. No obstante, los SDAs diseñados para trabajar en RIs no han evolucionado igualmente, y las propuestas recientes no han sido diseñadas para abordar esta complejidad de los datos, ya que no escalan correctamente o no analizan la mayoría de los tipos de datos creados en RIs. Además, en una revisión de la literatura en el campo de los SDAs a gran escala se comprobó que éstos no eran adecuados para su uso en RIs [1].

Esta tesis doctoral aspira a llenar ese vacío mediante dos propuestas principales: (i) un sistema visual de monitorización de red y (ii) un SDA multivariante, capaz de trabajar a gran escala y con datos heterogéneos (a nivel de red y proceso).

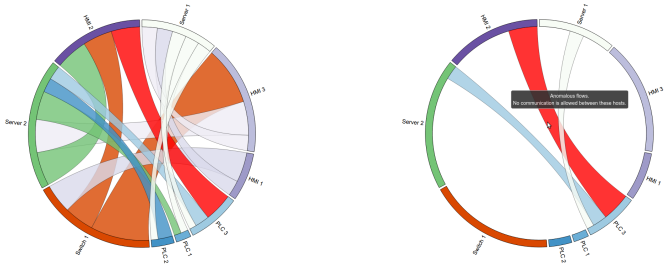
II. Contribuciones

Las contribuciones de esta tesis doctoral siguen un orden cronólogico: la construcción de un entorno de pruebas, la detección de anomalías a nivel de red y la detección de anomalías a nivel de red y de campo utilizando un modelo unificado escalable.

II-A. Banco de pruebas híbrido

En la investigación en ciberseguridad, es importante disponer de un entorno de pruebas en el que desarrollar y evaluar las diferentes propuestas. Los bancos de pruebas deben ser fieles al entorno que se está intentando replicar, y seguro, en el sentido que la realización de pruebas no afecte a activos fuera del banco de pruebas. Este último aspecto es aún más importante en el campo industrial, ya que la realización de pruebas puede comprometer la disponibilidad de los sistemas o tener un impacto físico no deseado en el entorno. Por ello, como paso preliminar a las principales contribuciones de la tesis doctoral, se ha construido un banco de pruebas en el que desarrollar y evaluar mecanismos de seguridad para redes industriales [2].

Este banco de pruebas es híbrido ya que utiliza diferentes técnicas de implementación: implementación en hardware, emulación, simulación y virtualización. Para ello se vale del software Emulab y del proceso simulado Tennessee-Eastman (TE). Emulab permite la emulación dinámica de diferentes topologías y escenarios de red, con la inclusión de nodos virtuales o físicos en diferentes configuraciones, mientras que el proceso TE, un proceso industrial de referencia, permite estudiar los efectos de diferentes situaciones en el plano físico de forma segura. La combinación de ambos, y la posibilidad de incluir diferentes tipos de procesos industriales hace que



(a) Flujo prohibido entre PLC 3 y HMI 2.

(b) Detalle del flujo anómalo

Figura 1: Diagrama de cuerdas de flujos de red con flujo de red anómalo.

el banco de pruebas desarrollado sea una plataforma versátil para la investigación en ciberseguridad para redes industriales.

II-B. Sistema visual de monitorización de seguridad de flujos de red industriales

Después de un análisis preliminar en el que se ha constatado la falta de visualizaciones de seguridad específicas para entornos industriales, se ha propuesto un sistema de monitorización visual de monitorización diseñado para facilitar la operación de la red y detectar anomalías de flujo [3]. Esta propuesta está basada en listas blancas, y diagramas de cuerdas. En una fase preliminar, el sistema construye las listas blancas en base al tráfico de red existente.

Una vez formadas las listas blancas, el sistema monitoriza tráfico real y lo visualiza en diagramas de cuerdas que muestran las diferentes relaciones entre los nodos de la red. En el caso de encontrar una conexión no contemplada en las listas blancas, ese flujo de red se muestra destacado sobre el resto, tal y como muestra la Figura 1. La escalabilidad de esta propuesta viene dada por la utilización de un servidor de búsqueda para el almacenamiento de las trazas de red y la generación de registros y de alertas, para los casos en los que la inspección visual no es viable. La propuesta ha sido validada con tráfico de una red industrial real.

II-C. Sistema de detección de anomalías multinivel

Las redes industriales se dividen en dos níveles lógicos principales. Por un lado, está el nivel de campo, compuesto por la realidad física que es monitorizada y los dispositivos (sensores y actuadores) que interactúan con ella. Por otro lado, está la capa de red, también conocida como la capa *cíber*, que engloba los dispositivos supervisores (servidores de control, interfaces humano-máquina...) de las redes indutriales. En términos generales, es el controlador industrial el que hace de puente entre ambos niveles, conectando la realidad física con los dispositivos supervisores.

La mayoría de las propuestas en la literatura solo se centran en una de las dos capas para detectar anomalías. En esta tesis, se ha desarrollado un marco en el que unificar ambos niveles en un único modelo para detectar anomalías. Para ello, utilizamos el Control Estadístico Multivariante de Procesos (CEMP, también conocido como MSPC, por sus siglas en inglés). CEMP ha sido utilizado previamente como método de detección de anomalías para redes IT [4]. Sus principales ventajas:

1. Basado en análisis de componentes principales, CEMP proporciona un modelo robusto de detección de anomalías y es capaz de definir en dicho modelo la información de un gran número de variables, teniendo en cuenta todas las variables existentes en una red industrial.
2. Mediante lo que se conoce como gráficos de contribución, CEMP permite analizar la contribución que han tenido las diferentes variables en la causa de una anomalía.

Utilizando CEMP, ya hemos demostrado que es posible obtener información acerca de la causa de una anomalía, pudiendo discernir en ciertas condiciones de si la causa de una anomalía se debe a una falla de proceso o la actuación de un atacante externo, solamente analizando las variables correspondientes al proceso industrial, es decir, magnitudes físicas [5]. Sin embargo, para obtener más información acerca de la causa del ataque e identificar los vectores de ataque utilizados, no basta con analizar la realidad física del proceso industrial; es indispensable añadir la información contenida en el nivel cíber o de red en el modelo CEMP para obtener esa información.

Para ello, se ha implementado un método basado en el conteo de ciertos eventos de red (registros, flujos activos...) en un lapso de tiempo dado, propuesto por [4] y se han añadido estas variables a las variables de proceso para la construcción del modelo CEMP. Para asegurar la escalabilidad de la propuesta, se ha desarrollado sobre Apache Spark.

La validación de este SDA multinivel se ha realizado utilizando el banco de pruebas descrito en la sección II-A, con tráfico de red real, el proceso TE y los registros creados por el SDA descrito en la sección II-B. Los resultados muestran que el SDA multinivel es capaz de detectar anomalías que afectan a uno o ambos niveles de la red industrial.

III. Transferencia tecnológica

El SDA visual se ha realizado en un proyecto de colaboración con MSIGrupo, transfiriendo el SDA visual a la gama de productos ofrecidos por la empresa.

Referencias

[1] M. Iturbe, I. Garitano, U. Zurutuza, and R. Uribeetxeberria, "Towards large-scale, heterogeneous, anomaly detection systems in industrial networks: A survey of current trends," *Security and Communication Networks*, vol. 2017, 2017.

[2] M. Iturbe, U. Izagirre, I. Garitano, I. Arenaza-Nuño, U. Zurutuza, and R. Uribeetxeberria, "Diseño de un banco de pruebas híbrido para la investigación de seguridad y resiliencia en redes industriales," in *Actas de II Jornadas Nacionales de Investigación en Ciberseguridad (JNIC 2016)*, 2016, pp. 3–10.

[3] M. Iturbe, I. Garitano, U. Zurutuza, and R. Uribeetxeberria, "Visualizing Network Flows and Related Anomalies in Industrial Networks using Chord Diagrams and Whitelisting," in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 2, 2016, pp. 99–106.

[4] J. Camacho, A. Pérez Villegas, P. García Teodoro, and G. Maciá Fernández, "PCA-based multivariate statistical network monitoring for anomaly detection," *Computers & Security*, vol. 59, pp. 118–137, 2016.

[5] M. Iturbe, J. Camacho, I. Garitano, U. Zurutuza, and R. Uribeetxeberria, "On the feasibility of distinguishing between process disturbances and intrusions in process control systems using multivariate statistical process control," in *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshop (DSN-W)*, 2016, pp. 155–160.