Avances recientes en la aplicación de la ciencia de datos a la ciberseguridad industrial

Recent advances in the application of data science to industrial cybersecurity

Esteban Jove[1], José-Luis Calvo-Rolle[1], Daniel Urda[2], Álvaro Herrero[2], Urko Zurutuza[3] y Valentina Casola[4]

[1] Universidad de A Coruña (España)
[2] Universidad de Burgos (España)
[3] Mondragon Unibertsitatea (España)
[4] Università degli Studi di Napoli Federico II (Italia)

DOI: https://doi.org/10.6036/10178

A lo largo de las últimas décadas, el concepto de ciberseguridad ha ido ganando peso en infinidad de ámbitos, tales como la industria o las telecomunicaciones, entre otros. Se define la ciberseguridad como un conjunto de procesos y tecnologías diseñadas con el objetivo de proteger programas, ordenadores, redes de comunicación y también datos ante ataques, y/o accesos no autorizados, asegurando de esta manera la confidencialidad, integridad y disponibilidad de los sistemas. A pesar de que no es posible garantizar una seguridad total, la ciberseguridad tiene como objetivo evitar los ataques maliciosos, reducir la vulnerabilidad de la información consecuencia de errores propios y paliar los daños ocasionados como consecuencia de los mismos.

Dependiendo de sus características, se consideran varios tipos de incidentes a los que un sistema de ciberseguridad debe hacer frente:

- Acceso no autorizado de información de una red, sistema o conjunto de datos.
- Software malicioso diseñado para dañar un ordenador, servidor, red, etc.
- Denegación de servicio (DoS) de un sistema, forzando su inutilización.
- El conocido como *phishing,* que se nutre de la interacción fraudulenta de usuarios para obtener información sensible sobre cuentas bancarias, redes sociales, etc.
- Ataque día cero llevado a cabo sobre un aspecto vulnerable no conocido en el sistema de seguridad.

Ante el aumento exponencial de los incidentes de esta naturaleza experimentados a lo largo de los últimos años, se plantean una serie de puntos de partida ineludibles a la hora de implementar un sistema de ciberseguridad.

- Conocer el estado de la red. Incluye dispositivos de interconexión, puertos de transmisión de datos, equipos, etc.
- Conocer el estado de los servicios que se ofrecen.
- Conocer las actividades que se están desarrollando en cada momento.
- Tener una idea de las tendencias que se observan.
- Recibir alarmas ante eventos e incidencias relevantes.
- Almacenar el conjunto de datos en un histórico.
- Prever futuras necesidades/anticipar futuros problemas gracias a los datos de tendencia.

Uno de los ámbitos en los que más relevancia ha ido ganando el concepto de ciberseguridad es el de la industria. Los sistemas industriales clásicos estaban formados por una estructura piramidal con cinco niveles bien diferenciados (Figura 1): campo, control, supervisión, planificación y gestión. Si bien existía interconexión entre dos niveles contiguos, esta estructura jerárquica carecía de comunicación entre niveles alejados en la pirámide. De la misma manera, la conexión con el exterior era prácticamente inexistente, haciendo uso de aplicaciones específicas.

Sin embargo, la denominada cuarta revolución industrial ha puesto sobre la mesa una evolución de los sistemas industriales. Por una parte, éstos presentan mayor comunicación con el exterior, haciendo uso de aplicaciones más genéricas y software libre. Por otra parte, se ha avanzado hacia una topología distribuida, basada en una mayor conectividad y flexibilidad, favoreciendo una mayor adaptabilidad ante la demanda de un mercado cada vez más globalizado. En este nuevo paradigma, el empleo de herramientas de digitalización y el *Internet of Things* (IoT) favorecen la comunicación entre los elementos de los distintos niveles, ya sean operarios, sensores o personal de gestión, entre otros.

A pesar de los beneficios que implica el incremento de la interconectividad, el mayor flujo de datos supone un aumento del riesgo de sufrir cualquier tipo de ataque que ponga en peligro el correcto funcionamiento de un proceso. Con el fin de ayudar en la toma de decisiones y garantizar la correcta operación de una instalación, velando por la seguridad del mismo, durante la monitorización se han de registrar innumerables variables que entran en juego y que están caracterizadas por lo que se conoce como las '4 v': velocidad, volumen, variedad y veracidad. Estas

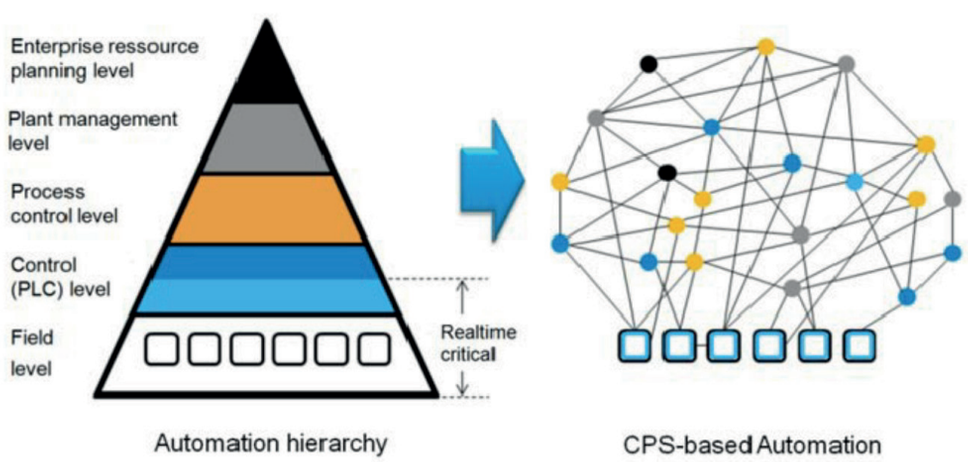

Figura 1: Evolución de la estructura de un proceso industrial

características, inherentes a la naturaleza de los datos monitorizados en cualquier proceso industrial, hacen que un observador humano encuentre gran dificultad a la hora de tomar decisiones que favorezcan la ciberseguridad.

Ante esta tesitura, se recurre al concepto multidisciplinar conocido como *data science*, que abarca la recopilación de datos, el análisis estadístico y procesado de los mismos, la detección de patrones, extracción de información valiosa y ayuda en la toma de decisiones. Este novedoso concepto, capaz de lidiar con el volumen, velocidad y variedad de los datos, se presenta como una herramienta de gran utilidad en el ámbito de la ciberseguridad.

A la hora de emplear técnicas inteligentes, el conocimiento previo acerca del sistema a proteger da lugar a tres tipos de enfoques posibles. Por una parte, se puede considerar la existencia tanto de datos correspondientes al correcto funcionamiento del sistema, como de los eventos anómalos a detectar. En ese caso, el empleo de técnicas de aprendizaje supervisado se aplica para determinar la naturaleza normal o anómala de los datos. Otro posible enfoque comúnmente utilizado se lleva a cabo cuando sólo se dispone de información acerca del funcionamiento habitual del sistema, en cuyo caso se emplean técnicas de tipo semisupervisado, capaces de etiquetar las instancias anómalas aun sin tener previo conocimiento acerca de las mismas. Este enfoque se desarrolla a partir del uso de técnicas de tipo one-class. Finalmente, el empleo de técnicas basadas en aprendizaje no supervisado pretende etiquetar los datos en ausencia de conocimiento previo acerca del comportamiento de los mismos.

Son muchos los retos que se plantean en cuanto a ciberseguridad de cara al futuro. La apertura de los sistemas industriales es un aspecto inaplazable, en el que se están introduciendo todas las prerrogativas que concede la informática de consumo. Sin embargo, no se puede dejar a merced de actitudes despiadadas y casi siempre injustificables, las acciones de ataques informáticos, que pudieran recaer en sectores tan críticos y/o estratégicos como pueden ser el nuclear o, desde un punto de vista más general, el energético.

La competitividad en el sector industrial indudablemente va a pasar por abandonar la actitud relativamente conservadora que la ha caracterizado en los últimos tiempos, y que era, en cierta medida, un mecanismo de seguridad. Los proveedores de productos industriales ya incorporan en sus catálogos la protección desde un punto de vista de ciberseguridad. Estos servicios han de incluir sistemas de protección escalables, disponiendo de mecanismos de identificación con una robustez infranqueable y adaptable en el tiempo. Por supuesto han de incorporar todas las prestaciones de los diferentes avances que se vayan produciendo, y han de ser lo suficientemente ágiles para progresar de forma pareja a la informática más avanzada del momento.

Todo lo anterior va a venir de la mano de implantación de políticas en las que se incorpore la ciberseguridad desde los estados más incipientes. No es una cuestión que pase únicamente por contar con tecnologías avanzadas en este sentido, si no de que éstas estén embebidas en los propios procesos de gestión, sin dejar de lado todos los factores y actores implicados en las instituciones. Con el objetivo de conferir alguna de las características anteriormente mencionadas, es vital por ejemplo para una actualización constante, el manejo de toda la información, así como su uso. El tratamiento se ha de hacer de una forma inteligente y que permita escalabilidad ante nuevas tecnologías, que con toda seguridad se consolidarán, como puede ser el uso del *cloud* en todos los procesos en que sea posible. Por supuesto, esta afirmación conlleva que la inteligencia de los sistemas, así como el conocimiento del que están provistos, se vaya adaptando a las nuevas exigencias, y que lo haga con la rapidez suficiente que permita garantizar que las instalaciones industriales no sean vulnerables a posibles ataques que se dan o puedan dar en el futuro. Es necesario hacer especial énfasis en esta última parte, dada la celeridad con la que aparecen nuevas amenazas, independientemente de la peligrosidad que puedan entrañar.

No cabe duda que para poder efectuar un mantenimiento y prevención óptimo de ataques desde un punto de vista digital es primordial poder contar con conocimientos especializados, tanto técnicos como estratégicos. De esta forma se podrá realizar una evaluación fidedigna del riesgo existente y las metas a perseguir, realizando actualizaciones periódicas con ese fin, tratando de ser y, por supuesto, parecer seguros desde el punto de vista de la ciberseguridad.

REFERENCIAS

[1] Givehchi, O., & Jasperneite, J. (2013). Industrial automation services as part of the Cloud: First experiences. Proceedings of the Jahreskolloquium Kommunikation in der Automation–KommA, Magdeburg.
[2] Jove, E., Casteleiro-Roca, J. L., Quintián, H., Méndez-Pérez, J. A., & Calvo-Rolle, J. L. (2021). A new method for anomaly detection based on non-convex boundaries with random two-dimensional projections. Information Fusion, 65, 50-57.
[3] Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data, 7(1), 1-29.
[4] Thames, L., & Schaefer, D. (2017). Cybersecurity for industry 4.0. Heidelberg: Springer.
[5] Vega Vega, R. A., Chamoso-Santos, P., Briones, A. G., Casteleiro-Roca, J. L., Jove, E., del Carmen Meizoso-López, M., ... & Calvo-Rolle, J. L. (2020). Intrusion detection with unsupervised techniques for network management protocols over smart grids. Applied Sciences, 10(7), 2276.