Distinguiendo entre perturbaciones de proceso e intrusiones en sistemas de control: caso de estudio con el proceso Tennessee-Eastman

Mikel Iturbe*, José Camacho†, Iñaki Garitano*, Urko Zurutuza*, Roberto Uribeetxeberria*

* Departamento de Electrónica e Informática
Escuela Politécnica Superior
Mondragon Unibertsitatea
Email: {miturbe,igaritano,uzurutuza,ruribeetxeberria}@mondragon.edu

† Departamento de Teoría de la Señal, Telemática y Comunicaciones – CITIC
Universidad de Granada
Email: jose.camacho@ugr.es

***Resumen*—Los sistemas de control de procesos (SCP) son los responsables de operar diversos procesos físicos, incluyendo infraestructuras críticas (ICs). La detección de anomalías en SCPs es un campo de investigación activo que tiene como objetivo asegurar el correcto funcionamiento de las ICs. Trabajos previos se han centrado en monitorizar el tráfico de red de los SCPs o bien no han considerado la existencia de perturbaciones de proceso, por lo que es posible malinterpretar dichas perturbaciones como anomalías de seguridad y viceversa. Presentamos un sistema de detección y diagnosis de anomalías basado en control estadístico multivariante de procesos (CEMP) que tiene como objetivo distinguir entre perturbaciones de proceso y anomalías de seguridad. Para este fin, extendemos el CEMP tradicional para monitorizar variables a nivel de proceso y controlador. Evaluamos nuestra propuesta utilizando el proceso Tennessee-Eastman. Los resultados muestran que nuestro enfoque se puede utilizar para distinguir entre perturbaciones e intrusiones.**

***Palabras clave*—Control estadístico multivariante de procesos (*Multivariate Statistical Process Control*), Detección de anomalías (*Anomaly Detection*), Sistemas de Control de Procesos (*Process Control Systems*),**

I. Introducción

Los sistemas de control de procesos (SCP) controlan, automatizan y monitorizan un gran número de procesos físicos de diversa índole, formando el núcleo de diversas infraestructuras críticas (ICs), que impulsan las sociedades modernas. Algunos ejemplos de ICs incluyen generación de energía, transporte, distribución y tratamiento de aguas etc. Por ello, es necesario proteger los SCPs y activos relacionados para asegurar el correcto funcionamiento de las sociedades modernas.

Esta necesidad se ha puesto de manifiesto con incidentes de seguridad directamente relacionados con los SCP donde atacantes alteraron el funcionamiento normal de este tipo de sistemas afectando también al entorno físico, algunos de ellos relacionados con ICs. Algunos ejempos de estos ataques con repercusión en el medio físico incluyen Stuxnet [1] o el incidente de la acería alemana [2].

En consecuencia, la seguridad en SCP es un campo de investigación activo que se ha enfocado principalmente en el desarrollo de nuevos mecanismos y técnicas. De estos mecanismos, los sistemas de detección de anomalías son especialmente relevantes, ya que la naturaleza estática y predecible de los SCP es apropiada para este tipo sistemas [3].

Sin embargo, a la hora de detectar eventos anómalos en redes formadas por SCPs, los factores que causan estas situaciones pueden ser diversos. Generalmente, estos factores pueden agruparse en dos tipos de conjuntos: las perturbaciones o fallas del proceso y los ataques e intrusiones.

Mientras que la mayoría de de las propuestas de la literatura en detección de intrusiones en redes de control se centran en el tráfico de red (muchas de ellas analizadas en la revisión de Mitchell y Chener [3]), otras propuestas, como ésta, se centran en la monitorización de las variables de proceso.

A la hora de tratar con las variables de proceso, las contribuciones se pueden dividir en dos grupos: por un lado, las propuestas en la que es necesario modelar el proceso monitorizado para detectar anomalías, y por otro, las propuestas en las que no es necesario tener un conocimiento profundo del proceso.

Entre las propuestas correspondientes al primer grupo destacan las contribuciones de McEvoy y Wolthusen [4] y Svendsen y Wolthusen [5]. Pese a ser eficientes para detectar anomalías, ambas propuestas necesitan modelos detallados del proceso monitorizado, lo cual dificulta extender su aplicación a diferentes tipos de procesos, siendo inviables en el caso de procesos complejos.

Por otro lado, están las propuestas de detección de anomalías en las que no es necesario modelar los procesos en detalle. En este campo, destacan las contribuciones de Kiss y col. [6] y Krotofil y col. [7].

Kiss y col. [6] presentan un sistema de detección de anomalías basado en clustering, donde agrupan las observaciones

recibidas de los sensores para después examinarlas mediante siluetas. Sin embargo, sólo consideran los ataques como fuente única de situaciones anómalas, por lo que las perturbaciones de proceso se clasificarían erróneamente como intrusiones.

Krotofil y col. [7] proponen un método de detección de ataques de integridad sobre las señales de sensores. El método propuesto se basa en detectar señales inconsistentes respecto a un grupo de señales correlacionadas utilizando entropía. Aunque tienen en cuenta la existencia de perturbaciones, no hacen una comparación entre las perturbaciones y ataques relacionados.

En este artículo, analizamos las limitaciones y posibilidades de detectar y distinguir la naturaleza del origen (perturbación o intrusión) de las anomalías utilizando control estadístico multivariante de procesos.

El resto del trabajo está estructurado de la siguiente manera. La sección II introduce el control estadístico multivariante de procesos. La sección III presenta el sistema propuesto. La sección IV muestra los resultados experimentales obtenidos. Por último, la sección V extrae las conclusiones y dibuja posibles líneas de trabajo futuro.

II. Control estadístico multivariante de procesos

El control estadístico multivariante de procesos (CEMP) [8] es una metodología de control de procesos que utiliza gráficos de control multivariantes para detectar cambios inesperados en el proceso monitorizado.

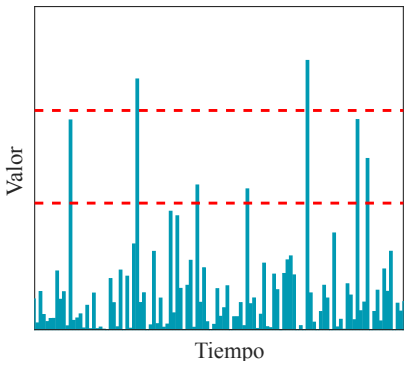


Figura 1. Ejemplo de un gráfico de control

La figura 1 muestra un ejemplo de un gráfico de control. Un cierto porcentaje (típicamente el 95 % o 99 %, dependiendo del cálculo del límite) de las observaciones en condiciones normales de operación (CNO) deben quedar debajo del límite de control con el nivel de confianza establecido.

En ese caso, consideramos que el proceso está en un estado de control estadístico. Es decir, solo existen causas comunes de variación en él [8].

La existencia de observaciones consistentemente fuera del límite de control establecido es atribuído a una causa de variación inesperada. En el caso de los SCP, es atribuible a perturbaciones o ataques, es decir, una anomalía.

Utilizando métodos como el análisis de componentes principales (PCA, por sus siglas en inglés), CEMP dispone de una metodología eficiente para monitorizar la magnitud de las variables y su relación con el resto de variables.

II-A. CEMP basado en PCA

Los datos históricos de proceso se pueden representar como un conjunto de datos de N observaciones de M variables, es decir, como una matriz $\mathbf{X} = N \times M$.

PCA transforma las variables originales a un nuevo conjunto de variables no correlacionadas llamadas componentes principales.

Para una X centrada en la media y típicamente auto-escalada[1] PCA utiliza la siguiente expresión:

$$\mathbf{X} = \mathbf{T}_A \mathbf{P}_A^t + \mathbf{E}_A \tag{1}$$

donde $\mathbf{T_A}$ es la matriz de puntuaciones, es decir las observaciones originales representadas según el nuevo subespacio; $\mathbf{P_A^t}$ es la matriz de cargas, representando la combinación lineal de las variables originales que forma cada uno de los componentes principales; por último, $\mathbf{E_A}$ es la matriz de residuos.

En CEMP basado en PCA, se monitorizan las puntuaciones y los residuos, cada uno en un gráfico de control diferente [9]. Por un lado, para representar las puntuaciones se utiliza el D-estadístico, también llamado el T^2 de Hotelling [10]. Por otro lado, en el caso de los residuos, el estadístico utilizado es el Q-estadístico o SPE [11].

Los estadísticos D y Q se calculan para cada observación en los datos de calibración, y se establecen los límites de control por cada uno de los dos gráficos. Después, estos estadísticos también se calculan para las observaciones entrantes y se dibujan en el gráfico de control. Cuando un cambio (o más) ocurre en una nueva observación, relativa a las M variables originales, al menos uno de los estadísticos superará el límite de control. Es decir, el problema de monitorización se simplifica de ser un problema M-dimensional a uno bidimensional.

En este trabajo, consideramos que un evento es anómalo cuando tres observaciones consecutivas superan el límite de control con el nivel de confianza del 99 %.

Una vez que la anomalía ha sido detectada, utilizamos gráficos oMEDA [12] para diagnosticar la causa de la anomalía relacionando el evento fuera del límite de control con las variables originales.

Fundamentalmente, los gráficos oMEDA son diagramas de barras donde los valores extremos, tanto positivos como negativos, identifican las variables que más contribuyen con su valor a una desviación en un subespacio de un grupo de observaciones.

Por ello, cuando se calcula el diagrama oMEDA para un grupo de observaciones dentro de un evento anómalo, las variables más relevantes relacionadas con ese evento anómalo serán las que tengan las barras más largas, tanto positivas como negativas.

[1] Normalizada a una media de cero y con varianza unitaria

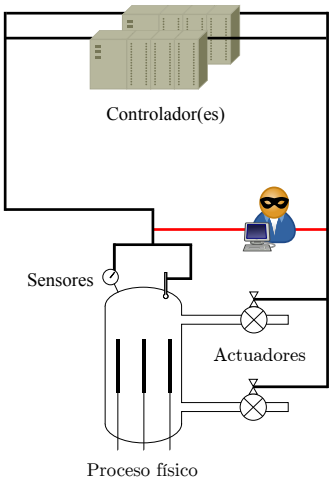


Figura 2. Ejemplo de un SCP con el modelo de ataque utilizado.

III. Sistema propuesto

La figura 2 muestra un ejemplo de SCP. El proceso físico reside en el núcleo del sistema, con un número fijo de sensores y actuadores. Estos sensores y actuadores son los dispositivos de entrada y salida que los controladores utilizan para obtener información del proceso e interactuar con él.

Sin embargo, las comunicaciones entre los controladores y el proceso ocurre generalmente sobre líneas de comunicación inseguras, utilizando protocolos de red sin cifrado o autenticación. Por ello, es posible que un atacante lea y/o modifique el tráfico de red entre los controladores y el proceso, realizando ataques *Man-in-the-middle*.

Esta realidad puede llevar a situaciones donde los datos recibidos por el controlador no se correspondan con los datos reales del proceso, o que las órdenes recibidas por los actuadores no hayan sido enviadas por el controlador.

En este trabajo utilizamos CEMP sobre un proceso industrial simulado, el proceso Tennessee-Eastman [13], para detectar y diagnosticar el origen de las anomalías, diferenciando entre factores naturales (perturbaciones) o inducidas (ataques).

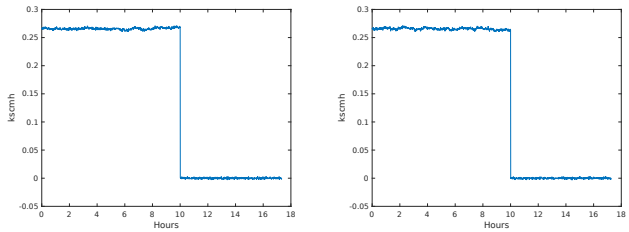
III-A. El proceso Tennessee-Eastman

El proceso Tennessee-Eastman (TE) es un modelo de un proceso químico real, presentado por Downs y Vogel [13] como proceso de referencia para evaluar diferentes estrategias de control. Sin embargo, el proceso también ha sido ampliamente utilizado en la investigación de ciberseguridad de los sistemas de control [4], [14], [7], [6].

En este trabajo utilizamos la estrategia de control descentralizado presentada por Ricker [15], junto con el modelo de aleatoriedad presentado por Krotofil y col. [7].

El modelo TE dispone de 41 variables medidas (XMEAS), 12 variables manipuladas y 20 perturbaciones de proceso (IDV). Estas variables y perturbaciones están descritas detalladamente en la publicación original de Downs y Vogel [13].

En el modelo, los controladores leen los XMEAS y fijan los valores de los XMV. Comparando con el diagrama simplificado de la figura 2, los XMEAS corresponden a las lecturas de los sensores y los XMV a los valores de los actuadores. Las perturbaciones del proceso son cambios inesperados, no deseados y a veces, inevitables en las condicones del proceso, que pueden afectar a la operación normal del proceso.



(a) IDV(6) (b) Ataque sobre XMV(3)

Figura 3. Comparación de la evolución de XMEAS(1) bajo la perturbación IDV(6) o un ataque de integridad sobre XMV(3).

En el caso del TE, la perturbación IDV(6) es una de las más difíciles de tratar. Modela una pérdida del flujo de entrada A.

XMEAS(1) es la variable medida asociada al flujo de entrada de A, mientras que XMV(3) es la variable manipulable que controla la válvula de entrada del flujo A. Por lo tanto, es de esperar que un ataque que tenga como objetivo cerrar la válvula de entrada asociada al flujo A y la existencia de la perturbación IDV(6) afectarán similarmente a XMEAS(1).

La figura 3 muestra ambas situaciones. Si se monitorizan los valores de XMEAS(1), no hay diferencias perceptibles en su evolución cuando se comparan la perturbación IDV(6) y un ataque de integridad en la variable XMV(3), donde el atacante cierra la válvula de entrada del flujo A. En ambos casos, el proceso realiza una parada de emergencia, ya que la falta del reactivo A conduce a una situación que puede ser potencialmente peligrosa si el proceso siguiese en marcha.

Teniendo una perturbación y un ataque asociado que se comportan de manera muy similar proporciona un entorno adecuado en el que evaluar técnicas de distinción entre intrusiones y perturbaciones.

III-B. Modelo de adversario

Los modelos de adversario utilizados en este escenario son los propuestos por Krotofil y col. [7].

Consideramos que el adversario es capaz de leer y manipular tráfico de red entre los controladores y el proceso físico, como muestra la figura 2.

Por lo tanto, el atacante es capaz de manipular las entradas tanto del proceso (valor XMV falsificado) como del controlador (valor XMEAS falsificado) mediante ataques de integridad, estableciendo el valor de la variable falsificada en un valor arbitrario de su elección.

IV. Resultados experimentales

Nuestra propuesta se evalua mediante un conjunto de experimentos en los que el modelo TE se ejecuta diez veces por situación anómala. La duración de cada simulación es de 72 horas, excepto en el caso en los que el proceso deja

de funcionar antes debido a que entra en límites inseguros para su operación. Los datos de calibración corresponden a 30 ejecuciones, con las que se construye el modelo CEMP utilizado luego para evaluar las anomalías.

Todas las situaciones anómalas empiezan a partir de la décima hora de simulación. Como se ha descrito anteriormente, un evento se categoriza como anómalo una vez que tres observaciones consecutivas sobrepasan el límite de control del 99 % en al menos uno de los estadísticos monitorizados.

Una vez detectada la anomalía, se calculan los gráficos oMEDA para el conjunto de las primeras observaciones fuera de los límites de control de cada una de las diez ejecuciones, detectada en cualquiera de los dos gráficos de control.

Por cada evento anómalo se crean dos gráficos, uno con los datos del proceso (datos recibidos y enviados por sensores y actuadores) y por otro con los datos del controlador. Ambos conjuntos de datos serán idénticos en escenarios libres de ataques.

Para el análisis de los datos y creación de diagramas se ha utilizado la herramienta MEDA [9].

Contemplamos tres escenarios diferentes: *a)* Perturbación IDV(6), *b)* Ataque de integridad sobre XMV(3) y *c)* Ataque de integridad sobre XMEAS(1).

En el caso de los ataques, el atacante establece el valor de las variables a cero, de modo que desde el punto de vista del controlador, la situación sería similar en los ataques y la perturbación, ya que percibiría una pérdida del flujo A.

Las figuras 4 y 5 muestran los gráficos oMEDA resultantes a nivel de controlador y proceso, respectivamente.

Las figuras 4a y 5a muestran el oMEDA en el caso de la perturbación. Al ser el flujo de entrada A mucho menor de lo habitual, XMEAS(1) sobresale como la variable que más contribuye a la situación anómala.

Las figuras 4b y 5b muestran los gráficos oMEDA en el que el atacante cierra la válvula de entrada del flujo de entrada A. En este caso, desde el punto de vista del controlador, la anomalía es similar a la perturbación. Pero, cuando examinamos los datos a nivel de proceso, observamos que la variable realmente relacionada con la anomalía, no es XMEAS(1), sino XMV(3), la variable que el atacante ha manipulado para cerrar la válvula.

Las figuras 4c y 5c muestran el escenario donde el atacante manipula la variable XMEAS(1) y la establece a cero. En consecuencia, el controlador recibe la información de que no hay flujo entrante en A. Eso es porque el atacante ha establecido el valor de XMEAS(1) en un valor menor del habitual. El algoritmo de control intenta solventar la situación percibida abriendo aún más la válvula de entrada XMV(3), permitiendo la entrada de más reactante A al proceso. Ésa es la razón por la que desde el punto de vista del proceso, se puede observar que tanto XMV(3) como XMEAS(1) tienen valores más altos que lo habitual.

Nuestra propuesta detecta las tres situaciones anómalas presentadas.

Sin embargo, a la hora de diagnosticar una anomalía, las lecturas a nivel del controlador –en los que el CEMP tradicional se ha basado– no son suficientes para hacerlo de manera correcta.

Es cuando se monitorizan las variables reales de proceso, junto con las variables del controlador cuando se pueden reconocer las causas de la anomalía.

Por ello, hemos extendido el modelo CEMP para monitorizar tanto las variables de proceso como las del controlador.

V. Conclusiones y trabajos futuros

Hemos presentado una metodología para distinguir entre perturbaciones de proceso y ataques de comportamiento similar.

Esta metodología se basa en CEMP para la detección de anomalías y gráficos oMEDA para identificar sus causas. Hemos utilizado el célebre proceso Tennessee-Eastman para evaluar experimentalmente nuestra propuesta.

Distinguir intrusiones y perturbaciones en sistemas de control es una tarea compleja, especialmente si se comprometen todas las líneas de comunicación.

Hemos extendido el modelo tradicional de CEMP que permite monitorizar tanto las variables de proceso como las de los controladores, para monitorizar eficientemente los SCP. En algunos entornos de SCP, se asigna una variable medida a cada una de las variables manipuladas, por lo que es factible utilizar esta metodología para la detección y diagnóstico de anomalías. Este escenario también complica el trabajo del atacante, ya que tendría que falsificar tanto el valor de la variable manipulada como la medida asociada para pasar desapercibido.

A la hora de analizar perturbaciones de proceso o intrusiones, los gráficos oMEDA muestran claramente cuales son las variables implicadas en la anomalía.

Sin embargo, el modelo también presenta algunas limitaciones: hay escenarios en los que no es posible monitorizar las variables de proceso junto con las del controlador. Para estos casos, una línea de trabajo futura interesante para detectar anomalías es la inclusión de las variables de red al modelo CEMP, junto con las del controlador. Además, este enfoque puede ayudar a detectar ataques adicionales sobre SCP que no tienen efecto alguno sobre la capa física del proceso.

En este sentido, ya se han utilizado metodologías similares a CEMP para la monitorización de seguridad de redes utilizando únicamente variables de red [16].

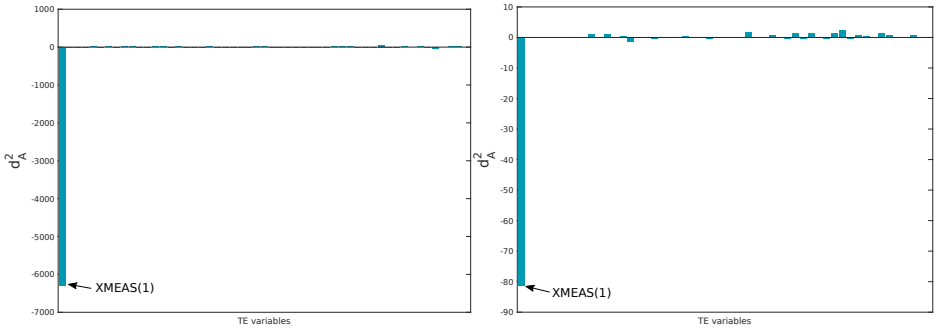
Agradecimientos

Este trabajo está parcialmente financiado por el programa Elkartek del Gobierno Vasco a través del proyecto KK-2015/0000080, el programa “Red guipuzcoana de Ciencia, Tecnología e Innovación” de la Diputación Foral de Gipuzkoa a través de la concesión 56/15 y el MINECO y fondos FEDER, a través del proyecto TIN2014-60346-R.

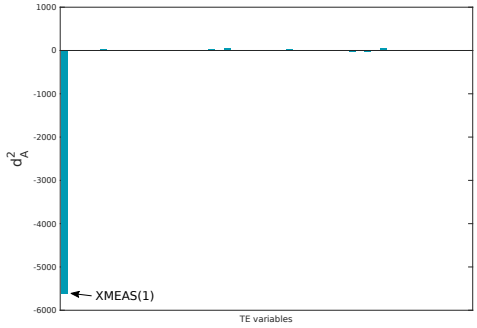
Referencias

[1] R. Langner, “Stuxnet: Dissecting a Cyberwarfare Weapon,” *Security Privacy, IEEE*, vol. 9, no. 3, pp. 49–51, May 2011.

[2] Bundesamt für Sicherheit in der Informationstechnik, “Die Lage der IT-Sicherheit in Deutschland 2014 (Technical Report),” Dec. 2014.

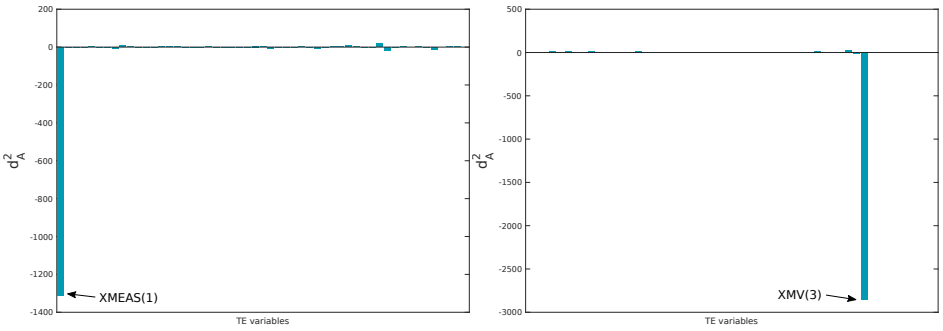


(a) IDV(6) (b) Ataque de integridad sobre XMV(3)

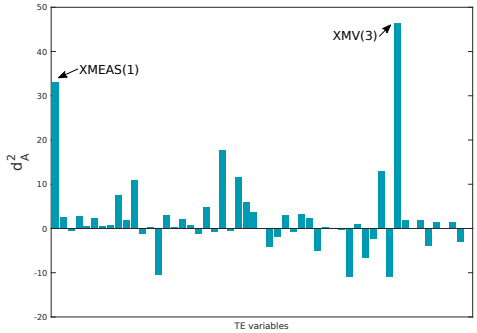


(c) Ataque de integridad sobre XMEAS(1)

Figura 4. Gráficos oMEDA de diferentes anomalías desde el punto de vista del controlador.



(a) IDV(6) (b) Ataque de integridad sobre XMV(3)



(c) Ataque de integridad sobre XMEAS(1)

Figura 5. Gráficos oMEDA de diferentes anomalías desde el punto de vista del proceso.

[3] R. Mitchell and I. Chen, "A Survey of Intrusion Detection Techniques for Cyber Physical Systems," *ACM Computing Surveys*, vol. 46, no. 4, April 2014.

[4] T. McEvoy and S. Wolthusen, "A plant-wide industrial process control security problem," in *Critical Infrastructure Protection V*. Springer, 2011, pp. 47–56.

[5] N. Svendsen and S. Wolthusen, "Using physical models for anomaly detection in control systems," in *Critical Infrastructure Protection III*. Springer, 2009, pp. 139–149.
[6] I. Kiss, B. Genge, and P. Haller, "A clustering-based approach to detect cyber attacks in process control systems," in *Industrial Informatics (INDIN), 2015 IEEE 13th International Conference on*, July 2015, pp. 142–148.
[7] M. Krotofil, J. Larson, and D. Gollmann, "The Process Matters: Ensuring Data Veracity in Cyber-Physical Systems," in *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, ser. ASIA CCS '15. ACM, 2015, pp. 133–144.
[8] J. F. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995.
[9] J. Camacho, A. Pérez Villegas, R. A. Rodríguez Gómez, and E. Jiménez Mañas, "Multivariate Exploratory Data Analysis (MEDA) Toolbox for Matlab," *Chemometrics and Intelligent Laboratory Systems*, vol. 143, pp. 49–57, 2015.
[10] H. Hotelling, "Multivariate quality control," *Techniques of statistical analysis*, 1947.
[11] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, no. 3, pp. 341–349, 1979.
[12] J. Camacho, "Observation-based missing data methods for exploratory data analysis to unveil the connection between observations and variables in latent subspace models," *Journal of Chemometrics*, vol. 25, no. 11, pp. 592–600, 2011.
[13] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.
[14] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and communications security*. ACM, 2011, pp. 355–366.
[15] N. L. Ricker, "Decentralized control of the Tennessee Eastman challenge process," *Journal of Process Control*, vol. 6, no. 4, pp. 205–221, 1996.
[16] 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.