Diseño de un banco de pruebas híbrido para la investigación de seguridad y resiliencia en redes industriales

Mikel Iturbe, Unai Izagirre, Iñaki Garitano, Ignacio Arenaza-Nuño, Urko Zurutuza, Roberto Uribeetxeberria

Dpto. de Electrónica e Informática
Escuela Politécnica Superior
Mondragon Unibertsitatea
Email: {miturbe, igaritano, iarenaza, uzurutuza, ruribeetxeberria}@mondragon.edu
unai.izagirre@alumni.mondragon.edu

***Resumen*—Las redes industriales son entornos interconectados dirigidos a controlar equipamiento físico en entornos industriales. Dada la dificultad de realizar investigación de seguridad en entornos industriales reales, principalmente por el potencial peligro derivado de la interacción con equipamiento físico, los bancos de pruebas son ampliamente utilizados. En estre trabajo, presentamos un banco de pruebas híbrido, que utiliza diferentes técnicas de implementación: implementación en hardware, emulación, simulación y virtualización. Para ello, se vale del software Emulab y el proceso simulado Tennessee-Eastman (TE). Emulab permite la emulación dinámica de la capa de red y virtualiza los nodos que lo componen, mientras que el proceso TE posibilita estudiar los efectos de diferentes situaciones en el plano físico de forma segura. Además, el banco de pruebas cuenta con soporte para redes definidas por software e integra un módulo de análisis de datos a gran escala, lo que permite el diseño de experimentos de ciberseguridad relacionados con estas tecnologías.**

***Palabras clave*—banco de pruebas, seguridad en entornos industriales, redes definidas por software, redes industriales**

Tipo de contribución: *Investigación en desarrollo*

I. Introducción

Las redes de control industrial son sistemas de equipamiento interconectado que automatizan, monitorizan y controlan equipamiento físico en entornos industriales. Como tales, las redes de control industrial están detrás de gran parte de los procesos automatizados de hoy en día, muchos de los cuales corresponden a infraestructuras críticas, como la generación y el transporte de energía, el tratamiento de aguas o la fabricación crítica.

Al operar equipamiento físico, es posible que ataques contra este tipo de redes vayan más allá de la capa lógica y tengan un impacto en medios físicos. En el pasado, varios incidentes de seguridad han dejado de manifiesto la magnitud del problema, desde la destrucción de equipamiento industrial como en el caso de Stuxnet [1], Aurora [2] o el incidente de seguridad en la acería alemana [3]; hasta la contaminación del medio ambiente como en el caso de las fugas de agua de Maroochy [4]. Además, debido a que estas redes manejan información sensible acerca de la naturaleza del proceso controlado (p.ej. recetas de fabricación), también son el objetivo de software maliciosos que realizan espionaje industrial. Duqu [5] o Flame [6] son algunos de los ejemplos importantes de estas herramientas avanzadas de espionaje.

Los sistemas de control industrial (SCIs) y especialmente, los controladores lógicos programables (PLC, por sus siglas en inglés) forman el núcleo de las redes industriales, ya que son los dispositivos responsables del control del equipamiento. Desde su desarrollo en la década de 1960, los SCIs y las redes industriales han evolucionado de ser entornos aislados, con hardware, software y protocolos de comunicación propietarios a utilizar tecnologías estándares que han facilitado su interconexión con las redes tradicionales de tecnologías de información (TI) con el objetivo de una mejor integración y reducción de costes.

En cambio, las tecnologías que utilizan las redes industriales no fueron diseñadas para estar interconectadas y por ello, generalmente carecen de mecanismos de seguridad presentes en redes TI, como el cifrado de las comunicaciones, cortafuegos o sistemas de detección de intrusiones (SDI).

Por ello, dada la tradicional vulnerabilidad de este tipo de redes y su importante papel en el correcto funcionamiento de sociedades modernas, es necesario desarrollar mecanismos de seguridad destinados a redes industriales. Sin embargo, a la hora de diseñar este tipo de soluciones, hay que tener en cuenta que las redes TI e industriales tienen objetivos y necesidades de seguridad diferentes [7], [8].

Mientras que las redes industriales tienen como objetivo el control de equipamiento físico, en las redes TI el objetivo es el transporte de datos. Si se tiene en cuenta la tradicional tríada de seguridad (Confidencialidad, Integridad y Disponibilidad), en las redes TI es la confidencialidad la que prima sobre el resto de cualidades. Es preferible no tener los datos disponibles en un lapso de tiempo que hacerlos públicos.

En contrapartida, en entornos industriales prima la disponibilidad, mientras que la confidencialidad y la integridad quedan en un segundo plano [7]. La red y sobre todo, el equipamiento controlado deben mantenerse disponibles y en funcionamiento aún en condiciones adversas y a base de sacrificar los otros dos principios de seguridad.

Por lo tanto, la latencia en las comunicaciones o la disminución de la capacidad de cómputo disponible que introducen diversas herramientas de seguridad o las auditorías como los tests de intrusión en las redes industriales no son admisibles, más aún en los procesos críticos que funcionan continuamente y donde no es aceptable una pausa del servicio o producción.

Por ello, evaluar el rendimiento de dispositivos de seguridad donde puede ser necesaria la introducción de condiciones anómalas y, en especial, la realización de auditorías como tests de penetración en redes industriales puede resultar en una pérdida económica importante e incluso potencialmente peligroso para el entorno [9].

Estos riesgos hacen que sea necesaria la utilización de bancos de pruebas en los que se puedan realizar pruebas de seguridad de diferente índole sin que se den las situaciones potencialmente peligrosas que ocurren cuando se trabaja con redes industriales en producción.

En este trabajo presentamos el diseño de un banco de pruebas para redes industriales, modular, híbrido y personalizable que permita la realización de investigaciones en el campo de la seguridad en redes industriales sin riesgos pero también con rigor científico para garantizar la reproducibilidad de los experimentos realizados y su equivalencia a efectos de compararlos con redes y equipamiento industrial en producción. A diferencia de las propuestas anteriores, también dispone de soporte para redes definidas por software y un módulo de análisis de datos.

El resto del trabajo está organizado de la siguiente forma: la sección II analiza trabajos relacionados con esta propuesta, la sección III describe la arquitectura del banco de pruebas propuesto, la sección IV presenta un caso de uso aplicado a la ciberseguridad, la sección V extrae conclusiones del trabajo realizado y por último la sección VI muestra unas posibles líneas de trabajo futuras.

II. Trabajos relacionados

Dada la dificultad de realizar pruebas en entornos reales, la creación de bancos de pruebas de entornos industriales ha conseguido considerable atención por parte de la comunidad científica. Holm et al. [10] realizan un estudio exhaustivo de los bancos de pruebas relativos a sistemas de control existentes. Identifican 30 bancos de pruebas con el objetivo de realizar pruebas de ciberseguridad. Entre el uso principal de dichos bancos, destacan principalmente el análisis de vulnerabilidades, la educación y la creación de entornos para evaluar el rendimiento de mecanismos de defensa.

Di Pietro y Panzieri [11] desarrollan una taxonomía que permite clasificar los bancos de pruebas en diferentes categorías. Del mismo modo, realizan una comparación de siete bancos de pruebas diferentes acorde con la taxonomía presentada.

Basándonos en estos dos trabajos, se deduce que el criterio principal de clasificación de los bancos de pruebas reside en los métodos de implementación utilizados para su construcción. Holm et al. [10] definen estos métodos como virtualización, simulación, emulación o hardware.

La mayoría de los bancos de pruebas utilizan diferentes métodos de implementación a la vez. Por ejemplo, simulando el proceso industrial en cuestión, mientras que el controlador corre dentro de una máquina virtual.

De entre los diferentes tipos de implementación, destacan tres bancos de pruebas que pueden considerarse representativos de diferentes enfoques: el banco de pruebas presentado por Reaves y Morris [12], Candell et al. [13] y Siaterlis et al. [14].

Reaves y Morris [12] presentan un banco de pruebas centrado en dispositivos virtuales, que emulan a los controladores. Estos dispositivos virtuales son los responsables del control de un proceso físico simulado ya que tienen la lógica de control implementada y disponen de capacidad de comunicación utilizando Modbus. Dependiendo de la configuración, estos dispositivos pueden actuar como esclavos o maestros y son interoperables con sistemas de control reales. El banco de pruebas dispone de dos procesos simulados que corresponden a un depósito de agua y un gaseoducto a escala de laboratorio creados para este fin.

El banco de pruebas propuesto por Candell et al. [13] está principalmente basada en hardware. Los controladores y el resto de los agentes de la red son físicos. En el caso del proceso industrial, este banco de pruebas tiene dos variantes, la simulación del proceso Tennessee-Eastman [15] y la utilización de un brazo robótico real. A estos dos procesos se les añade un tercer componente que además de contener la electrónica de red es el responsable de la captura y modificación de los paquetes de la red.

La propuesta de Siaterlis et al. [14] conocida como EPIC, es la más flexible de entre las tres propuestas aquí mencionadas. La arquitectura de EPIC está fundamentada en el software Emulab [16], el cual utiliza para recrear la capa lógica de una red industrial (topología de red, nodos...). La flexibilidad de Emulab permite, de manera dinámica, utilizar controladores físicos y virtuales, y diferentes topologías y estados de red. Como en los casos anteriores, EPIC utiliza la simulación de diferentes procesos físicos reales de diversa índole.

Si bien los enfoques de los bancos de pruebas mencionados son fundamentalmente diferentes, prácticamente todas las propuestas anteriores comparten carencias de cara a la investigación en ciberseguridad en sistemas de control industrial. Por un lado, está la carencia de soporte de redes definidas por software. La potencialidad de este tipo de redes para las redes industriales ya ha sido mencionada anteriormente [17]. Más específicamente, en el campo de la ciberseguridad, las redes controladas por software pueden proporcionar resiliencia y mecanismos de seguridad adicionales a las redes industriales [18], [19]. Por ello, Dong et al. [18] defienden la inclusión del soporte de este tipo de redes en bancos de pruebas de redes industriales, proponiendo un caso de uso aplicado al Smart Grid.

Por otro lado, los bancos de pruebas analizados también carecen de un módulo de almacenamiento y procesamiento de los datos creados en él. Reaves y Morris [12] ya señalan esta carencia, y consideran la creación de un repositorio para el almacenamiento "una extensión natural" a su banco de pruebas. En el caso del banco de pruebas de Candell et al. [13] si bien se menciona la existencia de un módulo de almacenamiento de paquetes de red, el módulo no dispone de funcionalidad adicional.

La tabla I muestra una comparación entre los trabajos anteriormente mencionados con nuestra propuesta, evaluando la naturaleza modular o flexible de la propuesta, la existencia de soporte para SDN, módulo de análisis de datos y por último, la posibilidad de ejecutar diferentes tipos de procesos simulados en él.

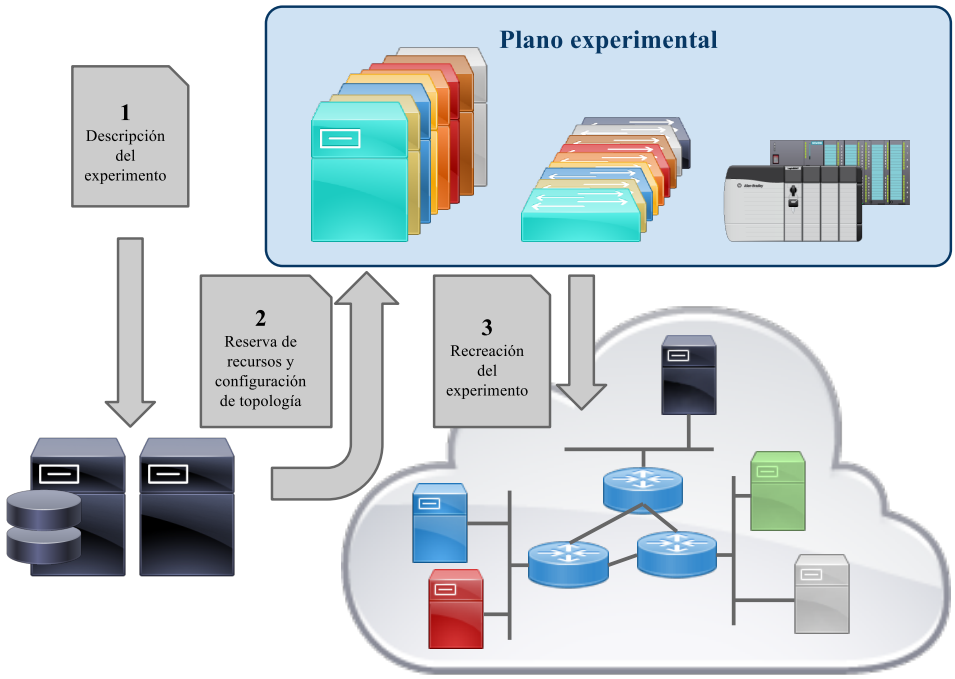

Figura 1. Flujo de trabajo en la creación de experimentos en Emulab

Tabla I
COMPARACIÓN DE TRABAJOS RELACIONADOS Y NUESTRA PROPUESTA.

Propuesta	**Modular**	**Soporte SDN**	**Módulo de análisis de datos**	**Procesos simulados**
Reaves et al. [12]	No	No	No	Sí
Candell et al. [13]	No	No	No	Sí
Siaterlis et al. [14]	Sí	No	No	Sí
Dong et al. [18]	No	Sí	No	Sí
Nuestra propuesta	Sí	Sí	Sí	Sí

III. ESTRUCTURA DEL BANCO DE PRUEBAS

En esta sección detallamos la estructura del banco de pruebas propuesto.

III-A. Visión de conjunto

La estructura propuesta del banco de pruebas se fundamenta en el uso del software Emulab [16]. Emulab es un sistema de software que proporciona una plataforma para la investigación, la educación, y el desarrollo de redes y sistemas distribuídos. Sus objetivos principales son la facilidad de uso, el control y el realismo, logrados mediante el uso sistemático de la virtualización y la abstracción.

La arquitectura básica de Emulab se compone de dos servidores de control, un conjunto de recursos físicos que se utilizan como nodos de experimento, y varios dispositivos de red que interconectan los nodos.

La figura 1 muestra el proceso de creación de un experimento en Emulab en diferentes pasos:

1. A través de un archivo NS (Network Simulator) el usuario especifica los detalles del experimento: el número de nodos utilizados, el sistema operativo utilizado por cada nodo, la topología de red deseada, las direcciones IP, el ancho de banda de cada red, etc.
2. Una vez definidos los detalles, el archivo se carga en los servidores que gestionan el banco de pruebas, donde a través del software Emulab se procede a la reserva y configuración de los recursos necesarios.
3. Transcurrido un tiempo se recrea la topología deseada, se configuran los nodos, se crean los usuarios y se le proporciona al usuario final la posibilidad de conectarse e interactuar con cada uno de los nodos disponibles.

Los aspectos personalizables de este sistema abarcan desde la elección del sistema operativo instalado en los nodos y las aplicaciones que se ejecutan en ellos, hasta el cambio en la carga, la velocidad y la pérdida de paquetes de cada enlace que conecta dichos nodos, pasando por el diseño de la topología de la red del experimento e incluso la posibilidad de trabajar tanto con ordenadores de sobremesa, PLCs, servidores, etc. como con máquinas virtuales.

Basándose en las directrices de la publicación NIST 800-82 [20] Holm et al. [10] contemplan cuatro aspectos que deberían estar presentes en un banco de pruebas para redes industriales (fig. 2):

1. **Arquitectura de comunicaciones**. Se refiere a los componentes que permiten la comunicación entre los diferentes dispositivos que existen en una red industrial: conmutadores, enrutadores, líneas de comunicación…
2. **Centro de control o dispositivos supervisores**. Engloba a los servidores y estaciones de trabajo que sirven para observar y controlar el proceso de forma remota. Historiadores, interfaces humano-máquina, servidores de control…
3. **Controladores y dispositivos de campo**. Son los dispositivos que unen la capa de red con la capa física. Aquí encontramos los controladores como los RTUs y los PLCs.
4. **Proceso físico**. Contempla la realidad física que los

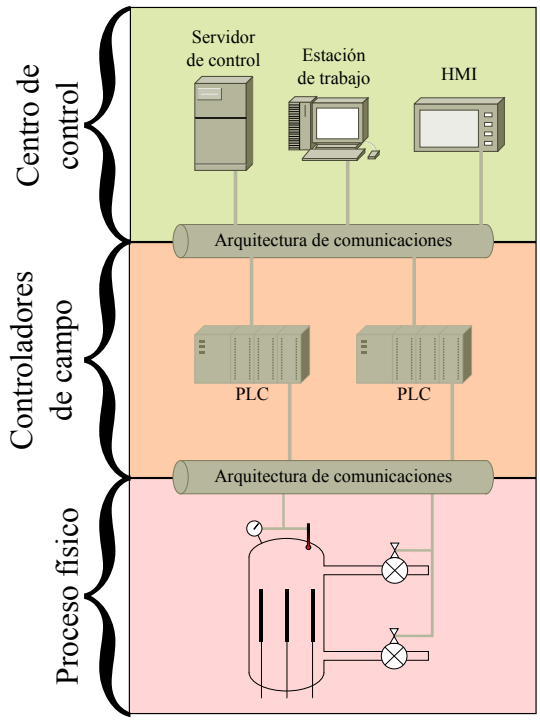

Figura 2. Estructura simplificada de un banco de pruebas

dispositivos de campo controlan y observan.

Además de los anteriormente mencionados, añadimos un módulo adicional de análisis de datos. Este módulo adicional, aunque no es parte de la red industrial, sí es un componente adicional que se encargará de la obtención, almacenamiento y procesamiento de los datos que se generan en los elementos que componen el banco de pruebas. De esta manera, se unifica todo el flujo de trabajo bajo un único entorno.

III-B. Arquitectura de comunicaciones

En el banco de pruebas, la arquitectura de comunicaciones está construída en torno a dos ejes: el software Emulab y las redes definidas por software (SDN, por sus siglas en inglés).

III-B1. Emulab: Anteriormente nos hemos referido a Emulab como el sistema de software con el que se ejecutan los experimentos diseñados. La figura 3 muestra la arquitectura del software.

Emulab se instala en dos servidores distintos. Estos dos servidores son los encargados de gestionar todo el hardware así como el software que se utiliza para recrear los experimentos. Mientras que uno de ellos se utiliza para la gestión del hardware, la interfaz de usuario, el servicio de nombres y el servicio de despliegue de las imagenes con el sistema operativo y las aplicaciones a ejecutar en los nodos, el segundo se encarga de la gestión y de los archivos de los usuarios, que típicamente se corresponden con los datos a utilizar en los experimentos y los datos generados en éstos. Del mismo modo, el primer servidor es el encargado de la validación de acceso y el gestor de las consolas de acceso para usuarios.

Tal y como se puede observar en la figura 3, los dos servidores están conectados a la red del Plano control de nodos. Esta red se utiliza para tener un control directo sobre los nodos de experimento. Así, cada nodo de experimento dispone de una interfaz de red reservada para su gestión por

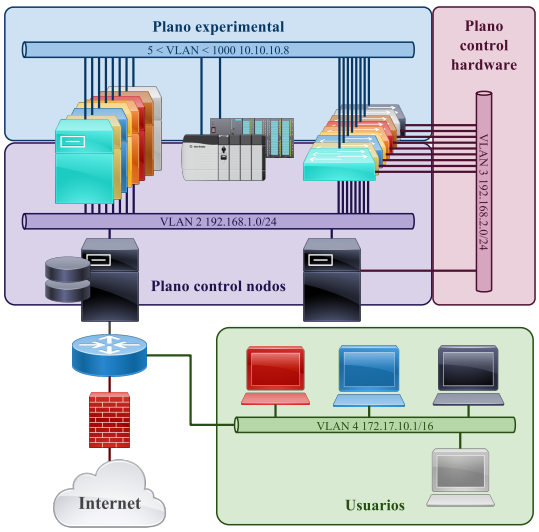

Figura 3. Arquitectura de Emulab

parte del software Emulab. Además, uno de los servidores está conectado a la red del Plano control de hardware. En esta red están conectados todos los conmutadores y enrutadores así como el gestor de energía. Mientras que los conmutadores y enrutadores se utilizan para configurar las Virtual Local Area Network (VLAN) que formarán la topología de red especificada por el experimento, el gestor de energía se utiliza para resetear los nodos que se hayan quedado bloqueados y para apagar los nodos que no se están utilizando en ningún experimento con el fin de ahorrar energía. Uno de los servidores se conecta a la red exterior mediante el enrutador con el objetivo de gestionar el archivo de descripción del experimento así como para dar acceso a los usuarios finales a los nodos de experimento.

El Plano experimental se utiliza para recrear todos los experimentos definidos. Cada nodo de experimento cuenta con al menos dos interfaces de red, una de ellas utilizada para su control y el resto conectados a los conmutadores experimentales. El software de Emulab cuenta con una base de datos en la que se establece cuál de las interfaces de cada nodo está conectado a cada puerto del conmutador. Esta información es realmente útil cuando se necesita configurar las VLAN.

Además de recrear diferentes arquitecturas, con Emulab también es posible recrear diferentes condiciones de la red. Es decir, puede emular pérdida de paquetes, latencia, ancho de banda de diferentes conexiones... En un estudio de Siaterlis et al. [21] los autores concluyen que Emulab es una herramienta eficiente y realista para emular las condiciones de la red.

III-B2. Redes definidas por software: Las redes definidas por software (SDN) constituyen un paradigma de red el cual tiene como objetivo superar las limitaciones existentes de las actuales infraestructuras de red [22].

Este tipo de redes separa el plano de control de la red del plano de transimisión de datos. Es decir, los agentes que configuran la red no son los mismos encargados de la transmisión de datos. De esta manera, los conmutadores de red se convierten en simples reenviadores y la lógica de control se implementa en un controlador central, simplificando de esta

manera la reconfiguración y evolución de la red.

La separación entre los planos de datos y control se puede realizar mediante una interfaz de programación (API) bien definida. En este sentido, una de las más notables APIs es OpenFlow [23]. Los dispositivos capaces de comunicarse a través del API OpenFlow como pueden ser los conmutadores OpenFlow, disponen de una o más tablas de direccionamiento de paquetes. Las tablas se componen de reglas en las que cada una describe un subconjunto de tráfico y una acción que bien puede ser eliminar, reenviar y/o modificar paquetes. Así, dependiendo de las reglas que un controlador SDN establece sobre un dispositivo de red SDN, como bien puede ser un conmutador SDN, el conmutador actuará como un reenviador, un cortafuegos, o realizará otras acciones como el balanceo de carga y/o conformado de tráfico.

En este banco de pruebas se utiliza OpenFlow como API para la construcción de diferentes redes definidas por software.

Para evitar conflictos entre el software Emulab y el controlador SDN, se identifica el nodo correspondiente al controlador como no gestionado, por lo que la interfaz de conexión con el controlador será independiente. Adicionalmente, se configura cada nodo de experimento en una VLAN diferente (además de la de control) para forzar el tráfico a través del controlador, de no ser así, el conmutador controlado por Emulab redireccionaría los paquetes directamente al nodo de destino sin consultar previamente con el controlador SDN.

De este modo, el tráfico se administra mediante los filtros y reglas definidas en el controlador SDN.

III-C. Centro de control

El centro de control de las redes industriales engloba el conjunto de dispositivos que realizan un control supervisor del proceso y sirven para controlarlo de forma remota: servidores de control, estaciones de trabajo, interfaces humano máquina... A diferencia de los controladores y dispositivos de campo, los dispositivos supervisores se ejecutan sobre hardware y sistemas operativos estándares, presentes en redes TI.

Por ello, la virtualización de estos dispositivos es una opción viable a la hora de realizar experimentos. Para poder utilizar dispositivos supervisores virtualizados en el banco de pruebas, es necesario generar las imágenes con el software necesario a partir de las imagenes de sistema operativo que dispone Emulab.

A la hora de elegir el software, hay dos principales alternativas: Utilizar software comercial, generalmente, del mismo fabricante que los controladores, o bien, desarrollar un software que realice esa función.

Si la fidelidad es la cualidad más importante la primera alternativa es más recomendable, ya que así se pueden utilizar componentes reales, con las ventajas que ello supone: p.ej. poder realizar un análisis de vulnerabilidades en el producto o no tener que implementar capacidad de comunicación con los protocolos generalmente propietarios de cada fabricante.

Sin embargo, ya que la creación de imágenes para los experimentos está en manos de los usuarios, es posible utilizar uno, otro o ambos enfoques a la hora de elegir software de control.

III-D. Controladores de campo

Los controladores de campo son el núcleo de las redes industriales. Con entradas y salidas múltiples y con una capacidad de cómputo limitada, los controladores de campo actúan sobre el proceso partiendo de las entradas recibidas y el algoritmo de control que tienen implementado. Además envían los valores de las distintas variables a los aparatos supervisores del centro de control. Como tal, son los actores que unen las capas física y virtual o cíber de las redes industriales.

III-D1. Controladores físicos: El banco de pruebas dispone la posibilidad de utilizar controladores de campo físicos, del mismo tipo que se utilizan en redes industriales en producción. Desde Emulab no es posible administrar los controladores físicos de la forma en que se administran las máquinas virtuales y la topología virtual. Sin embargo, sí es posible añadir nodos estáticos a Emulab, es decir, nodos experimentales que no estén directamente administrados por el software, de manera similar al controlador de SDN mencionado anteriormente.

En el caso de los controladores esta posibilidad presenta ventajas como poder realizar análisis de vulnerabilidades a un dispositivo real, o poder trabajar con controladores que utilizan software y protocolos de red propietarios que son difíciles de replicar en un entorno virtual o emulado.

En cambio, utilizar controladores físicos también tiene desventajas: tiene unos costes más altos, y es necesaria el desarrollo de la aplicación de control lo cual para procesos más complejos puede no ser viable.

III-D2. Controladores virtuales: La alternativa a utilizar controladores físicos es la virtualización de las funciones del controlador. En estos momentos, no es posible la virtualización completa de controladores de campo, especialmente en el caso de los controladores propietarios.

Pero por otro lado, sí que es posible emular las funciones de un controlador de campo mediante software, una práctica que se ha utilizado anteriormente en bancos de pruebas [12], [14].

En esta propuesta, se utiliza el entorno propuesto por Genge et al. [24]. Este entorno es el encargado de comunicar la capa física con la capa virtual del banco de pruebas. Para ello, esta plataforma, unifica directamente las variables de un proceso simulado a registros de memoria de un controlador o varios, de manera similar al que ocurre con controladores físicos.

De esta manera, es posible desarrollar controladores virtuales propios, utilizando librerías de protocolos de red industriales estándares existentes como Modbus/TCP, u OPC. La comunicación entre el controlador y el proceso quedaría abstraída.

III-E. Proceso físico

La gran mayoría de bancos de pruebas existentes recurren a la simulación del proceso físico. La simulación simplifica el uso de procesos físicos complejos en el banco de pruebas y permite utilizar procesos estándares que permitan replicar los experimentos realizados. Adicionalmente, el uso de la simulación proporciona un entorno seguro y fiel en el que investigar situaciones potencialmente peligrosas en los procesos industriales reales sin tener que interactuar directamente con un proceso físico.

Frente a las diferentes alternativas de software de simulación, Thornton y Morris [25] realizan una comparación entre simulaciones de Python y Simulink, demostrando que el último es el que más se asemeja a un proceso real.

Por ello, es también el software de simulación que se utiliza en este banco de pruebas.

El proceso físico utilizado en este banco de pruebas es el célebre Tennessee-Eastman (TE), originalmente planteado como problema de control por Downs y Vogel [15]. El proceso TE ya ha sido previamente utilizado con fines de investigación de seguridad en entornos industriales [26], [27].

Específicamente, utilizamos la variante llamada DVCP-TE que utiliza el control descentralizado de Ricker [28] para el control del proceso TE y varios módulos de ataques en Simulink que permiten la realización de ataques en la capa física desarrollado por Krotofil y Larsen [29].

Sin embargo, el componente de simulación y el software de controlador virtual permite utilizar cualquier modelo de proceso escrito en Simulink.

III-F. Análisis de datos

Muchos de los procesos son continuos o de larga duración, por lo que los experimentos asociados a procesos largos pueden crear potencialmente una gran cantidad de datos (registros del control, capturas de paquetes de red, registros...) intrínsecamente. Estos datos capturados pueden utilizarse posteriormente para la monitorización de procesos, evaluación de sistemas de detección de anomalías, optimización...

Pero la naturaleza de los datos capturados puede suponer un problema Big Data, ya que en una ventana de tiempo larga, pueden crearse un gran volumen de datos, a gran velocidad (las lecturas de procesos críticos se realizan cada pocos milisegundos) y de gran variedad (observaciones del proceso, tráfico de red...)

Las tecnologías Big Data se fundamentan sobre arquitecturas y sistemas operativos estándares, por lo que de manera similar a los componentes del centro de control, pueden crearse imágenes de los nodos de procesamiento que se unirían a la red experimental.

En este banco de pruebas el software utilizado para el análisis Big Data es Apache Spark [30]. Spark permite la realización de procesamiento por lotes y procesamiento contínuo, por lo que se puede utilizar para construir modelos de datos en reposo o procesar datos entrantes.

El objetivo principal del módulo es la detección de anomalías mediante el análisis de datos de gran volumen, aunque también puede ser utilizado para otros fines. Abre la posibilidad de estudiar la seguridad de las comunicaciones y el almacenamiento desestructurado de datos en grandes volúmenes.

IV. Caso de uso aplicado a la ciberseguridad

En esta sección presentamos un caso de uso aplicado a la ciberseguridad. La figura 4 muestra la visión en conjunto de un caso de uso en el que se emula una red industrial que tiene como objetivo controlar el proceso Tennessee-Eastman.

Como se ha mencionado antes, Emulab crea la topología de red, y carga las imágenes en los nodos mediante las opciones definidas en un archivo NS. Estas imágenes se han creado previamente, y para simplificar el caso de uso, no

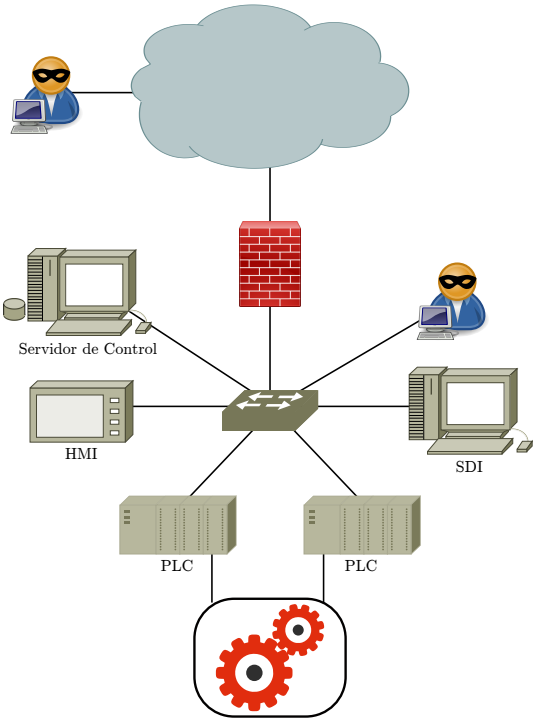

Figura 4. Caso de uso

consideramos nodos estáticos (PLCs y cortafuegos físicos), aunque podrían utilizarse en caso de ser necesario.

Así, cada nodo de la red se cargaría con un software y una configuración especializada dependiendo de su función. En primer lugar están los nodos que generalmente existen en redes industriales: PLCs, el proceso simulado, servidor de control y HMI. En segundo lugar, se introducen mecanismos de seguridad como los cortafuegos o SDIs. Por último, se cargan dos nodos que corresponderían a atacantes, interno y externo.

De este modo, esta configuración descrita permitiría evaluar los siguientes aspectos de ciberseguridad de una red industrial:

- Resiliencia de la red ante la pérdida o deterioro de la conexión entre diferentes nodos.
- Evaluación de la eficacia del cortafuegos para detectar ataques externos.
- Evaluación de la eficacia del SDI para detectar intrusiones.
- Técnicas de evasión de SDI en redes industriales.
- Identificación y realización de ataques críticos contra el proceso TE.
- Detección de anomalías en la red mediante el análisis a gran escala de los datos creados.

V. Conclusiones

Los bancos de pruebas para la investigación en ciberseguridad de redes industriales es un campo de investigación activo. Existen diferentes bancos de pruebas que se diferencian en los métodos de implementación utilizados para reproducir las funcionalidades de una red industrial.

Hemos presentado un banco de pruebas que mediante la emulación dinámica de la capa de red, su topología y los dispositivos que la componen y la simulación del proceso físico, puede construir las funcionalidades de una red industrial.

Para ello utiliza el software Emulab y el proceso Tennessee-Eastman, ambos utilizados previamente para la investigación de ciberseguridad en entornos industriales.

El banco de pruebas está diseñado además para poder funcionar con redes definidas por software, lo que permite la investigación de mecanismos avanzados de seguridad y resiliencia en redes industriales, aspecto que no se ha cubierto en contribuciones de bancos de pruebas anteriores. Además cuenta con un módulo de análisis de datos a gran escala integrado en el ecosistema que permite procesamiento de grandes volúmenes de datos en reposo y contínuos.

VI. Trabajos futuros

Una vez diseñado e implementado el banco de pruebas, el siguiente paso es el de la medición de la fidelidad de éste. En este sentido, se han propuesto varias métricas que sirven para este fin. Siaterlis et al. [14] proponen la utilización del tiempo de ejecución del modelo físico simulado y compararlo con los requisitos de latencia de éste. Reaves y Morris [12] identifican métricas relativas a las tramas de paquetes Modbus. Otras propuestas recogidas por Holm et al. [10] utilizan el cumplimiento de las directrices y estándares publicados por organismos de referencia.

Es interesante trabajar hacia una métrica de medición común, que pueda ser utilizada ampliamente para la comparación de resultados de bancos de pruebas de forma eficiente, ya que no hay un marco común. La integración de métricas de proceso, red y regulatorios pueden ser combinadas para crear este marco de evaluación de bancos de pruebas.

Por otro lado, la federación de bancos de pruebas puede ser un campo prometedor para trabajar la interdependencia y resiliencia de infraestructuras críticas. Permitiría estudiar el "efecto cascada", donde la interrupción del servicio de una infrastructura crítica pueden impactar en otras, así como medidas y mecanismos para hacerle frente. Especialmente, las redes definidas por software pueden tener un gran potencial en este campo.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Departamento de Innovación, Desarrollo Rural y Turismo de la Diputación Foral de Gipuzkoa en el marco del programa de "Red guipuzcoana de Ciencia, Tecnología e Innovación", bajo el proyecto KEA (56/15) y por el Departamento de Desarrollo Económico y Competitividad del Gobierno Vasco en el marco del programa Elkartek, bajo el proyecto BID3A (KK-2015/00080).

Referencias

[1] N. Falliere, L. O. Murchu, and E. Chien, "W32.Stuxnet dossier," *White paper, Symantec Corp., Security Response*, 2011.

[2] M. Zeller, "Myth or reality–Does the Aurora vulnerability pose a risk to my generator?" in *Protective Relay Engineers, 2011 64th Annual Conference for*. IEEE, 2011, pp. 130–136.

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

[4] Jill Slay and Michael Miller, *Lessons learned from the Maroochy Water Breach*. Springer, 2007.

[5] B. Bencsáth, G. Pék, L. Buttyán, and M. Félegyházi, "Duqu: Analysis, detection, and lessons learned," in *ACM European Workshop on System Security (EuroSec)*, 2012.

[6] K. Munro, "Deconstructing Flame: the limitations of traditional defences," *Computer Fraud & Security*, vol. 2012, no. 10, pp. 8–11, 2012.

[7] M. Cheminod, L. Durante, and A. Valenzano, "Review of Security Issues in Industrial Networks," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 277–293, 2013.

[8] B. Galloway and G. Hancke, "Introduction to Industrial Control Networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 860–880, 2012.

[9] D. Duggan, M. Berg, J. Dillinger, and J. Stamp, "Penetration testing of industrial control systems," Sandia National Laboratories, Tech. Rep. SAND2005-2846P, March 2005.

[10] H. Holm, M. Karresand, A. Vidström, and E. Westring, "A Survey of Industrial Control System Testbeds," in *Secure IT Systems*, ser. Lecture Notes in Computer Science, S. Buchegger and M. Dam, Eds. Springer International Publishing, 2015, vol. 9417, pp. 11–26. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-26502-5_2

[11] A. Di Pietro and S. Panzieri, "Taxonomy of SCADA systems security testbeds." *International Journal of Critical Infrastructures*, vol. 10, no. 3, pp. 288–306, 2014.

[12] B. Reaves and T. Morris, "An open virtual testbed for industrial control system security research," *International Journal of Information Security*, vol. 11, no. 4, pp. 215–229, 2012.

[13] R. Candell, T. Zimmerman, and K. Stouffer, "An Industrial Control System Cybersecurity Performance Testbed," National Institute of Standards and Technology, Tech. Rep. NISTIR 8089, Nov 2015.

[14] C. Siaterlis, B. Genge, and M. Hohenadel, "EPIC: a testbed for scientifically rigorous cyber-physical security experimentation," *Emerging Topics in Computing, IEEE Transactions on*, vol. 1, no. 2, pp. 319–330, 2013.

[15] 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.

[16] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar, "An integrated experimental environment for distributed systems and networks," *ACM SIGOPS Operating Systems Review*, vol. 36, no. SI, pp. 255–270, 2002.

[17] J. Zhang, B.-C. Seet, T.-T. Lie, and C. H. Foh, "Opportunities for software-defined networking in smart grid," in *Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on*. IEEE, 2013, pp. 1–5.

[18] X. Dong, H. Lin, R. Tan, R. K. Iyer, and Z. Kalbarczyk, "Software-defined networking for smart grid resilience: Opportunities and challenges," in *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*. ACM, 2015, pp. 61–68.

[19] E. Molina, A. Astarloa, and E. Jacob, "Seguridad definida por software en subestaciones eléctricas," in *I Jornadas Nacionales de Investigación en Ciberseguridad (JNIC 2015)*, 2015, pp. 78–79.

[20] K. Stouffer, J. Falco, and K. Scarfone, "Guide to Industrial Control Systems (ICS) Security, Special publication 800-82," National Institute of Standards and Technology, Tech. Rep., June 2011.

[21] C. Siaterlis, A. P. Garcia, and B. Genge, "On the use of Emulab testbeds for scientifically rigorous experiments," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 2, pp. 929–942, 2013.

[22] H. Kim and N. Feamster, "Improving network management with software defined networking," *Communications Magazine, IEEE*, vol. 51, no. 2, pp. 114–119, 2013.

[23] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[24] B. Genge, C. Siaterlis, I. N. Fovino, and M. Masera, "A cyber-physical experimentation environment for the security analysis of networked industrial control systems," *Computers & Electrical Engineering*, vol. 38, no. 5, pp. 1146–1161, 2012.

[25] Z. Thornton and T. Morris, "Enhancing a Virtual SCADA Laboratory Using Simulink," in *Critical Infrastructure Protection IX*, ser. IFIP Advances in Information and Communication Technology, M. Rice and S. Shenoi, Eds. Springer International Publishing, 2015, vol. 466, pp. 119–133. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-26567-4_8

[26] 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.

[27] 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.

[28] N. L. Ricker, "Decentralized control of the Tennessee Eastman challen-

ge process," *Journal of Process Control*, vol. 6, no. 4, pp. 205–221, 1996.

[29] M. Krotofil and J. Larsen, "Rocking the pocket book: Hacking chemical plants for competition and extortion," *DEF CON 23*, 2015.

[30] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: cluster computing with working sets," in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, 2010, pp. 10–10.