Entrenamiento bajo demanda en competencias de ciberseguridad en redes sociales

Mario Fernández-Tárraga, Alejandro David Cayuela-Tudela, Pantaleone Nespoli, Joaquin Garcia-alfaro, Félix Gómez Mármol

Entrenamiento bajo demanda en competencias de ciberseguridad en redes sociales

Mario Fernández Tárraga[1], Alejandro David Cayuela Tudela[1], Pantaleone Nespoli[1,2], Joaquín García Alfaro[2], Félix Gómez Mármol[1]
[1]Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, 30100, Murcia, España
{mario.fernandezt,alejandrodavid.cayuelat,pantaleone.nespoli,felixgm}@um.es
[2]SAMOVAR, Télécom SudParis, Instituto Politécnico de París, 19 place Marguerite Perey, 91120 Palaiseau, Francia
joaquin.garcia_alfaro@telecom-sudparis.eu

***Resumen*—En la última década las redes sociales han sufrido un crecimiento exponencial dada su popularidad y sus beneficios de comunicación, conexión y difusión de contenido. Pero también presentan una serie de desventajas, riesgos y amenazas que pueden ser explotadas para fines malintencionados, aprovechándose del desconocimiento de los usuarios. Para solucionar estos problemas planteamos una propuesta de herramienta que permite la formación en competencias de ciberseguridad de alta calidad en el contexto de redes sociales, tanto para la concienciación para los sectores sociales vulnerables como para el entrenamiento de perfiles profesionales concretos y/o militares. Esta propuesta permite automatizar la generación y configuración de escenarios simulados en redes sociales a distintos niveles tanto de contenido, como dificultad y realismo, permitiendo generar multitud de situaciones hiperrealistas en un amplio rango de posibilidades tanto para el ámbito social como para ámbitos profesionales o militares.**

***Index Terms*—Cyber Range, Ciberseguridad, Ciberdefensa, Simulación de redes sociales, Entrenamiento de competencias de ciberseguridad en redes sociales, Educación digital**

Tipo de contribución: *Formación e innovación educativa*

I. Introducción

Desde la segunda mitad del siglo XX la tecnología ha evolucionado exponencialmente hasta llegar a la situación tecnológica y digital actual. Uno de los principales exponentes de esta evolución son las redes sociales (RRSS), cuya hiperconectividad ha facilitado la globalización de información y comunicación con miles de personas, entre otras muchas ventajas. Aún teniendo en cuenta el uso diario de las RRSS no todos los usuarios son conscientes de las desventajas, amenazas y riesgos que suponen el uso de estas plataformas, donde distintos actores con fines y objetivos malintencionados participan directa o indirectamente.

La difusión, propagación y alcance instantáneo de contenido representa un gran desafío cuando se utiliza información incorrecta o manipulada, englobados en la "desinformación". Estas publicaciones pueden ser no malintencionadas (como el humor), aunque la mayoría terminan con repercusiones negativas, como en las imágenes virales de marzo de 2023 del abrigo blanco del Papa[1] y las del arresto de Trump[2]. En escalas mayores aparecen los conceptos de *dominio cognitivo* y *guerra cognitiva* [1], junto con grandes actores como países y empresas, cuyo objetivo es influir en la mentalidad y conducta de la sociedad. Uno de los casos más recientes en España fue las elecciones generales españolas de 2019, donde se utilizaron las RRSS y los bots para influir sobre los votantes [2]. Además, en la guerra cognitiva se busca controlar una narrativa favorable, desmintiendo y combatiendo la desinformación para aunar aliados y disuadir a enemigos neutrales u hostiles.

Por otro lado, el Instituto Nacional de Ciberseguridad (INCIBE) declara que las RRSS se encuentran dentro del cuarto vector de ataque más utilizado por ciberdelincuentes [3], pues permiten tanto recolectar información para preparar ataques más elaborados como atacar a usuarios vulnerables con ingeniería social (*phishing*) o la subida de software malicioso. La mayoría de sus éxitos ocurren por distracciones y negligencias de los usuarios, ya que muchos ataques son identificables y/o previsibles si se conocen. Otros muchos problemas, como la privacidad, información filtrada, la identificación de delitos, ciberbullying, bots, cuentas falsas o robadas, etc., se podrían reducir con una buena formación digital.

Todos estos desafíos convergen en que el factor humano es la parte más vulnerable de la cadena de ciberseguridad, apoyado por muchos estudios como [4], [5] o [6]. En este artículo se propone formación social y profesional para mitigar y/o acabar con los problemas mencionados, apoyándose en una estructura de los conocidos Cyber Ranges [7], mediante un sistema de entrenamiento en RRSS. Esta propuesta de formación, junto a la arquitectura necesaria para desplegarla, permite levantar y controlar ciberejercicios fácilmente, así como su automatización y personalización de retos extensibles, multidisciplinales e hiperrealistas. Y, adicionalmente, se ha identificado una lista resumen de casos de uso simulables en la propuesta para la formación de distintos perfiles sociales y profesionales o militares.

El resto del artículo tiene la siguiente estructura: la Sección II muestra la escasez de trabajo académico. La Sección III explica la propuesta con un primer apartado III-A, el simulador de RRSS, y un segundo apartado III-B, presentando la arquitectura. A continuación, en la Sección IV se listan aquellos casos de uso identificados. En las Secciones V y VI se encuentran respectivamente la implementación general de la propuesta y una demostración (PoC) sobre el caso de uso de desinformación. Finalmente, la Sección VII expone las conclusiones obtenidas, así como el trabajo futuro de la propuesta.

[1]https://elpais.com/tecnologia/2023-04-01/por-que-nos-hemos-creido-la-foto-del-papa-con-el-abrigo-blanco.html

[2]https://www.antena3.com/noticias/mundo/redes-detienen-donald-trump-fotos-falsas-virales-expresidente-arrestado-carcel_20230322641b30287262e50001b91c9a.html

Tabla I
ESTADO DEL ARTE - COMPARACIÓN DE CARACTERÍSTICAS.

Características	CYRAN	Somulator	Prevency	Propuesta
Contenido multimedia	NO	SI	SI	SI
Creación de usuarios	SI	SI	SI	SI
Automatización de usuarios	NO	NO	SI	SI
Configuración de usuarios	NO	SI	SI	SI
Autoconfiguración de usuarios	NO	NO	SI	SI
Publicaciones pregeneradas	SI	SI	SI	SI
Publicaciones autogeneradas	SI	SI	NO	SI
Autopublicación	SI	NO	NO	SI
Configuración de estadísticas	NO	NO	SI	NO
Configuración de ejercicios	SI	SI	SI	SI
Automatización de ejercicios	NO	NO	NO	SI
Configuración en caliente	NO	SI	SI	SI
Uso de plataformas reales	SI	SI	NO	SI
Control paralelo	NO	NO	NO	SI
Múltiples usuarios simultáneos	SI	SI	SI	SI
Perfiles psicológicos	NO	NO	NO	SI
Perfiles de comportamiento	NO	NO	NO	SI
Relaciones de usuario	NO	NO	NO	SI

II. ESTADO DEL ARTE

La simulación de RRSS no es un concepto novedoso. Se han realizado e implementando muchas soluciones sobre casos de estudio concretos desde la última década, como pueden ser el estudio de comportamientos o la polarización de usuarios sobre determinados tópicos, o la agrupación de comunidades, interacciones y relaciones generadas por los agentes en las simulaciones, como en [8]. Sin embargo, el concepto de simulación de plataformas de RRSS no ha sido explorado, y menos aún su uso para entrenar competencias, es decir, existen escasas soluciones de herramientas y/o propuestas para simular RRSS enfocadas en el entrenamiento de competencias en ciberseguridad. Las pocas soluciones propuestas han sido principalmente desarrolladas por empresas y entidades privadas, por lo que apenas existen artículos e investigaciones en el ámbito académico que fomenten su desarrollo.

En [9] (CYRAN Cyber Range Extension) se utilizan las primeras versiones disponibles de plataformas de RRSS de código abierto, con grandes limitaciones en funcionalidad. Estas son la imposibilidad de uso de contenido multimedia y la automatización de creación de usuarios, así como la configuración de perfiles de usuarios ya creados. La publicación de contenido se realiza mediante publicaciones previamente generadas y publicaciones generadas dinámicamente (autogeneradas), publicándose automáticamente según los parámetros introducidos en las plantillas para la generación del ejercicio. Los ejercicios pueden configurarse manualmente mediante dichas plantillas, pero no existen mecanismos que automaticen su proceso de creación. Además, la herramienta no soporta la configuración en caliente, y no permite controlar varias ejecuciones de manera simultánea. Aun así, la herramienta sí está pensada para soportar a varios usuarios reales o estudiantes simultánea y concurrentemente.

La segunda propuesta [10] (Somulator) está desarrollada por el Norwegian Defence Research Establishment (FFI), soporta contenido multimedia, permite la creación y configuración de usuarios, pero no automatiza dichos procesos. Las publicaciones usadas son pregeneradas y cargadas gracias a plantillas, y deben ser seleccionadas y publicadas manualmente. No permite la configuración de estadísticas, pero sí la configuración de ejercicios (y también en caliente durante la ejecución). La herramienta utiliza varias plataformas de código abierto, y está diseñada para que muchos usuarios reales o estudiantes la utilicen simultáneamente.

La tercera propuesta [11] (Prevency) también soporta contenido multimedia y permite crear usuarios tanto individualmente como colectivamente mediante la automatización de este proceso. La configuración de usuarios y perfiles también se puede hacer tanto de manera individual como automática al crear dichos perfiles colectivos. Las publicaciones son pregeneradas y se deben publicar deliberadamente. Además, permite la configuración de estadísticas tanto a los usuarios como a las publicaciones, y permite configuración en caliente. La herramienta está pensada para utilizarla con varios usuarios reales simultáneamente, pero como las demás, no está enfocada a tener un ejercicio desplegado en varias instancias individuales.

En la Tabla I se muestran las principales diferencias según las características de cada propuesta o herramienta, así como su comparación con la propuesta de este artículo. En las últimas líneas de la tabla aparecen los nuevos conceptos explicados en la Sección III, Subsección III-A, los cuales no han sido explorados por ninguna de las propuestas anteriores.

III. PROPUESTA

En esta sección se explica la propuesta de este artículo en el contexto de simulación de ejercicios en RRSS para el entrenamiento de cibercompetencias, así como la arquitectura usada para su planteamiento.

III-A. Simulador de RRSS para entrenamiento

La propuesta se basa en una herramienta que automatiza el proceso de creación de ciberejercicios simulados en RRSS en

el contexto Cyber Range, así como su configuración avanzada. De esta manera, se pueden generar simulaciones hiperrealistas para el entrenamiento de cibercompetencias en cualquier tipo de escenario y situación, preparando y cualificando a los distintos sectores sociales, profesionales y militares contra las amenazas inminentes y riesgos causados por, relacionados con o desarrollados en las RRSS.

La solución propuesta también hace uso de RRSS de código abierto, específicamente y únicamente Mastodon, un espejo de Twitter. Mantiene muchas bases comunes propuestas anteriormente enfocándolas en la automatización, así como los conceptos de usuario, publicación y contenido. Además, también introduce los nuevos conceptos de personalidad, comportamiento, relaciones, y la configuración de perfiles. Estos nuevos conceptos, junto con los conceptos base mejorados, permiten dotar a las simulaciones de configuraciones mucho más avanzadas, generando escenarios hiperrealistas para cualquier tipo de situación. La personalidad dota a los usuarios de descripciones psicológicas y gustos. Sirve para interpretar y clasificar contenido entrante y generar respuestas y publicaciones acordes a la personalidad. El comportamiento dota a los usuarios de acciones e interacciones. Sirve para que el usuario pueda llevar a cabo acciones concretas de publicación, de respuesta a notificaciones, o de control. Las relaciones dotan a los usuarios de relaciones con otros usuarios. Sirve para que el usuario pueda generar y aceptar relaciones, seguir, bloquear o silenciar usuarios, aceptar o rechazar peticiones de seguimiento. La configuración de perfil dota a los usuarios de configuraciones de perfil y de privacidad. Sirve para personalizar perfiles con biografías, alias, imágenes de perfil, etc., y con otras opciones de privacidad.

También se proponen varias mejoras mediante la automatización de los procesos. Estas automatizaciones se consiguen mediante el uso de superparámetros y de las plantillas base usadas para generar los ejercicios y las configuraciones necesarias para aprovisionar la simulación. Las plantillas o plantillas base son la guías o modelos en formato JSON que contienen elementos con la configuración y aprovisionamiento de una simulación (véase Sección V, Tabla III), y los superparámetros son parámetros usados para determinar la construcción y configuración automatizada de una plantilla base (véase Tabla II). Las automatizaciones propuestas son las siguientes:

1. Creación y configuración de ciberejercicios en RRSS: Se automatiza la creación de ciberejercicios, permitiendo que el instructor, con unos pocos superparámetros simples, obtenga una plantilla base configurada usable.
2. Configuración avanzada de ciberejercicios: Las plantillas base generadas pueden ser configuradas a distintos niveles de detalle, con infinitas posibilidades gracias a los nuevos conceptos propuestos.
3. Creación y personalización de usuarios: Se automatiza la creación de cuentas de usuario diferenciando entre usuarios relevantes y aleatorios. También se automatiza la configuración de perfiles, relaciones, personalidades y comportamientos.
4. Contenido de ciberejercicio: Se automatiza la generación de contenido y publicaciones de importancia necesarias para desarrollar el ciberejercicio, clasificándose en categorías y tipos según el tópico, el caso de uso o tipo de competencia que se quiera trabajar.
5. Contenido de Non Playable Character (NPC) simulado: Se automatizan las interacciones, generación de contenido y publicaciones usadas como tráfico de fondo para que la simulación tenga vida propia.

Adicionalmente, se debe destacar que la herramienta está orientada al contexto de un Cyber Range, por lo que debe trabajar con distintos ciberejercicios simultáneamente. Los despliegues de un mismo ciberejercicio deben de ser comunes para todos los estudiantes, siendo imperativo que la configuración base sea la misma, dejando como diferencia el tráfico de fondo dinámico. De igual manera, la herramienta debe permitir modificar e interactuar con las simulaciones mientras se desarrolla el ciberejercicio, tanto para despliegues individuales o como agrupados.

III-B. Arquitectura

La arquitectura propuesta está basada en la estructura típica de un Cyber Range ampliando la arquitectura propuesta en el Cyber Range COBRA [12] con el objetivo de centrar el foco en el desarrollo de las características propias de los módulos de entrenamiento en RRSS. En la Figura 1 se muestra de forma de gráfica la arquitectura de la propuesta y en los apartados III-B1, III-B2 y III-B3 se describen y detallan los elementos de la misma.

III-B1. Arquitectura front-end: Corresponde con la arquitectura de contenedores que se ha definido en el Cyber Range COBRA, de esta manera la arquitectura del front-end se abstrae de la inclusión de nuevos módulos. Las tecnologías que se pueden utilizar para dicho front-end son, entre otras, Django o AngularJS. En la Figura 1 se ejemplifica a que sistemas debe estar conectado.

III-B2. Sistema de creación de escenarios, retos y ciberejercicios: El sistema de creación escenarios, retos y ciberejercicios tiene como objetivo dotar de las configuraciones y capacidades necesarias a las máquinas virtuales (MV) para la realización del ciberejercicio y su posterior despliegue así como configurar los retos y ciberejercicios. Inicialmente el front-end se comunica con el gestor de creación que maneja las peticiones que se realizan respecto a la creación de los escenarios, retos y ciberejercicios y los envía al módulo de configuración de escenarios. Adicionalmente, se comunica con el módulo de generación de plantillas para recuperar la plantilla final del reto.

El módulo de configuración de escenarios es el encargado de crear escenarios con una o varias MV, pudiendo configurar parámetros cómo el sistema operativo, los servicios o aplicaciones, los ficheros que deben estar en la máquina y/o las características de la propia MV. Con el objetivo de que la configuración sea modular y ampliable, los servicios son contenedores *docker*, y es posible enviar cualquier tipo de fichero que sea necesario a la MV. Para el entrenamiento de competencias en RRSS es obligatorio que alguna de las máquinas contengan el servicio de Mastodon.

Tras la creación del escenario, el instructor crea y configura el reto mediante los denominados superparámetros presentados en la Sección V. El módulo de generación de plantillas devuelve una plantilla en JSON con la configuración completa del reto (plantilla e instrucciones de configuración de la

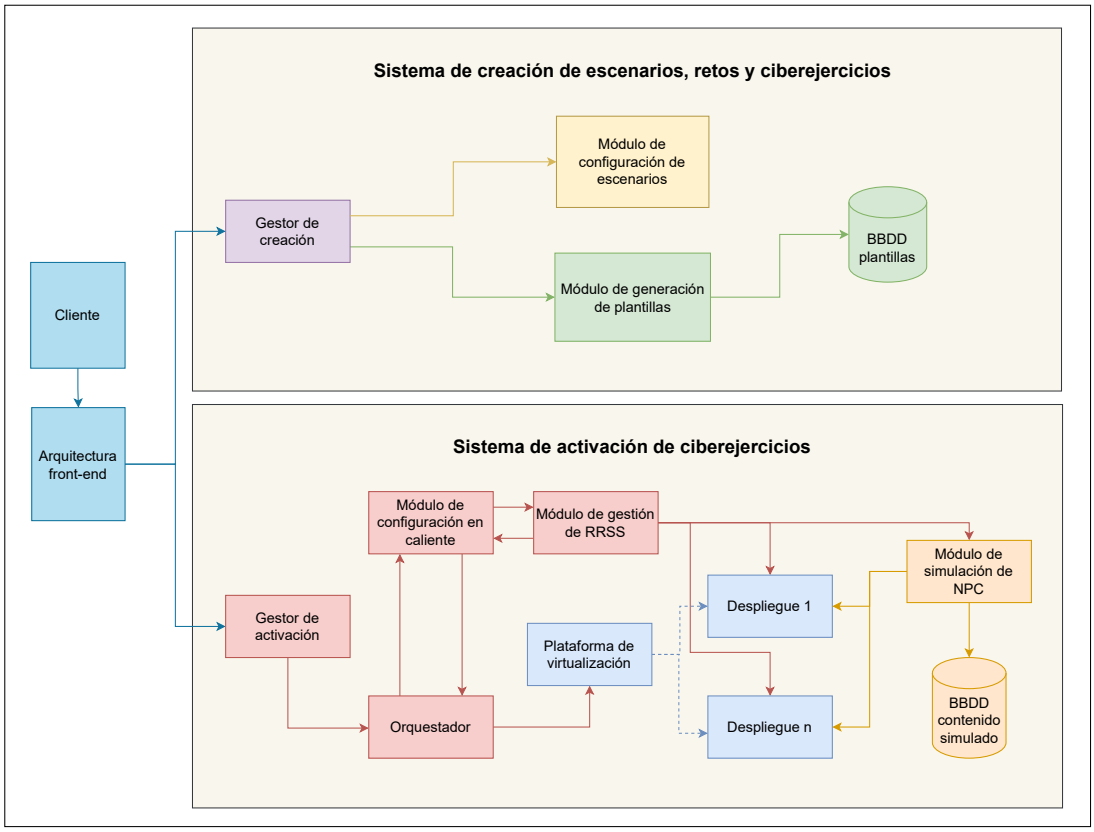

Figura 1. Arquitectura del simulador de RRSS.

instancia) que será cargado en Mastodon tras la activación. Para la generación de la plantilla el módulo cuenta con una base de datos de plantillas previas con contenido de RRSS previamente generado según tópicos, tipos y subtipos de ejercicios, objetivos, etc. También cuenta con configuraciones de personalidades y comportamientos, automatizaciones para parametrización y despliegue de usuarios, o la cantidad de publicaciones, entre otras muchas posibilidades.

III-B3. Sistema de activación de ciberejercicios: El sistema de activación tiene como objetivo el correcto despliegue, activación y realización de los ciberejercicios. A este sistema se puede acceder teniendo el rol de instructor o de estudiante. De forma similar al Gestor de creación el front-end se comunica con el Gestor de activación que maneja las peticiones realizadas y permite activar/desactivar los ciberejercicios al igual que permite la conexión entre el estudiante y el despliegue donde realizará el entrenamiento.

Si se accede como instructor se puede dar comienzo al ciberejercicio, donde el orquestador se comunica con la plataforma de virtualización, que realiza los despliegues de las MV y las aprovisiona, (se dota de sistema operativo y de servicios, como Mastodon). Tras el correcto despliegue el orquestador se comunica con el módulo de configuración en caliente para enviar la plantilla de Mastodon con el fin de preparar el estado inicial del ciberejercicio. Es importante destacar que cada despliegue cuenta con su propio módulo de gestión de RRSS y su módulo de simulación de NPC que se ejecutan localmente en la MV donde se encuentra Mastodon. El modulo de gestión de RRSS recibe la plantilla, la analiza, genera la configuración y la carga en su instancia local de Mastodon. Tras la carga, pasa a monitorizar el estado de la red social.

Si accede un estudiante, únicamente tendrá acceso al despliegue mediante una interfaz gráfica con el ciberejercicio y acceso a la máquina a través de aplicaciones de escritorio remoto. Tras la conexión con la máquina virtual el estudiante puede realizar el ciberejercicio para el entrenamiento de las competencias deseadas por el instructor, durante el tiempo estipulado en la configuración.

Además, para dotar de vida a la red social se cuenta con un módulo de simulación NPC que genera tráfico simulado, es decir, crea y publica contenido relacionado con las personalidades de los NPC, y realiza interacciones entre los usuarios de la simulación. Similar a la base de datos del módulo de generación de plantillas, el módulo de simulación de NPC cuenta también con una base de datos con contenido generado previamente.

IV. Casos de uso

Para enfrentarse a los posibles malos usos de las RRSS, se han investigado cuáles son los principales desafíos de las RRSS que podrían ser simulados en la propuesta de este artículo. Se han pensado y agrupado según tipos de ciberejercicios y competencias que se podrían desarrollar en desafíos desde las distintas perspectivas de la ciberseguridad. En las siguientes secciones se muestran algunos de los casos de uso simulables en esta propuesta para el entrenamiento de competencias de ciberseguridad en RRSS en el contexto de un Cyber Range. Estos casos de uso han sido identificados previamente por los autores durante el desarrollo de esta

propuesta.

IV-A. Detección de patrones de comportamiento y usuarios malintencionados

Dentro de las RRSS aparecen usuarios, interacciones y contenidos que pueden ser perjudiciales para la sociedad si se apoyan, se comparten o se confían en ellos, como por ejemplo los famosos y peligrosos ataques de phishing o la suplantación de identidad de celebridades. Muchas de estas amenazas pueden ser identificables por sus patrones de comportamiento, pero la mayoría de usuarios los desconoce. Para evitar y minimizar esas amenazas es necesario el desarrollo de una mente y pensamiento críticos para identificar aquellos patrones y elementos. En esta sección se identifican ciberejercicios desde la perspectiva de un usuario común enfocado en el desarrollo de competencias que permitan detectar amenazas para mitigarlas o eliminarlas. Algunos ejemplos de este caso de uso son descubrir bots, usuarios tóxicos, identificar ataques de phishing o ingeniería social, e incluso investigar cuentas que suplantan a otras personas reales.

IV-B. Identificación de delitos y crímenes en RRSS

Cada vez ocurren más delitos cibernéticos o relacionados con ellos, y las RRSS no son la excepción. Aunque no lo parezcan, las RRSS son un foco donde se desarrollan y generan actividades ilegales y delictivas a distintos niveles y escalas, por lo que es muy importante llevar un control férreo sobre los delitos que afectan y perjudican a la sociedad. Se ha identificado este caso de uso desde la perspectiva de perfiles profesionales jurídicos y judiciales como abogados, jueces, investigadores forenses o peritos informáticos, que les permitirían desarrollar competencias sobre delitos realizados o relacionados con RRSS. Algunos ejemplos son la investigación de casos de ciberbullying, amenazas de muerte ante un homicidio, publicaciones de imágenes privadas, etc., cuya detección podría ser vital en un caso.

IV-C. Detección de bulos y desinformación

La desinformación y los bulos representan una amenaza inminente en la sociedad actual y son principalmente propagados por las RRSS. Consisten en la generación de noticias y elementos falsos o manipulados con varios objetivos, casi siempre malintencionados, que son compartidos y propagados afectando a dimensiones de valor políticas, económicas y cognitivas que rigen la sociedad. En este tipo de ciberejercicios el estudiante debe aprender a identificar los elementos desinformativos y falsos de la plataforma, tratándolos mediante la búsqueda de fuentes verídicas, clasificándolos según el tipo de desinformación y/o el posible objetivo de la publicación. También se puede extender a la creación, planificación y lanzamiento de contramedidas para la minimización, mitigación y desactivación de los impactos, riesgos y amenazas de la desinformación.

IV-D. Búsqueda de información y datos personales

Todo lo que se hace en internet deja algún tipo de rastro. En el caso de muchas RRSS se recolectan y almacenan grandes cantidades de datos de los usuarios aunque estos no lo sepan. Además, los propios usuarios de las RRSS ofrecen información sobre ellos inconscientemente, ya sea por publicaciones o contenido que muestran su vida privada o porque dan información a usuarios desconocidos o falsos. Este tipo de ciberejercicios está centrado en la búsqueda y obtención de información desde una perspectiva atacante, permitiendo que los estudiantes se conciencien sobre estas técnicas. Las principales técnicas a usar son conocidas en el contexto de la ciberseguridad, como pueden ser el *Footprinting*, las técnicas *OSINT* (Open Source Intelligence) o la propia ingeniería social.

IV-E. Administración y moderación en RRSS

Gracias a las RRSS aparecen nuevos trabajos y modelos de negocio, de publicidad y de marketing, por lo que la administración y la moderación se han convertido en tareas críticas. Por un lado, están los trabajos de gestión y administración de RRSS, y por otro lado aparecen trabajos de gestión de cuentas profesionales y comerciales que representan a individuos u organizaciones. Este caso de uso está enfocado desde perfiles profesionales relacionados con las RRSS, donde se simulan entornos para gestionar problemas del ámbito de trabajo. Algunos ejemplos son la administración y gestión de crisis informáticas (nuevas cuentas falsas masivamente) o social (error de empresa), lanzamientos de productos o el desarrollo de campañas publicitarias o de concienciación, etc.

IV-F. Protección y privacidad en RRSS

Las RRSS presentan muchas amenazas y vulnerabilidades a nivel de usuario, tanto por una mala gestión del mismo como por atacantes externos que buscan aprovecharse. Muchas de estas vulnerabilidades se deben a la negligencia del usuario por no informarse o protegerse debidamente, normalmente por desconocimiento, negligencia o comodidad. Es necesario concienciar debidamente a los usuarios de las implicaciones de no aplicar correctamente las medidas de seguridad consideradas buenas prácticas. Para ello, se han identificado ciberejercicios relacionados con las buenas prácticas de ciberseguridad que proporcionan una mayor protección a los usuarios, como pueden ser el uso de la autenticación de doble factor, la generación de contraseñas seguras o la propia configuración de privacidad de las RRSS.

IV-G. Ataques a usuarios y plataformas de RRSS

Se reafirma que las RRSS son un vector de ataque bastante común en los ciberataques, siendo un problema de máxima prioridad en la sociedad actual debido a su alcance y fácil acceso. Este caso de uso está enfocado desde el punto de vista atacante, que debe usar la red social como vector u objetivo de ataque. En estos ejercicios el estudiante puede usar diversas herramientas, procedimientos y técnicas para aprovecharse de la plataforma y sus usuarios en varios ámbitos, ya sea por el envío de mensajes para phishing, virus, troyanos o worms, o ya sea por la explotación de vulnerabilidades de la propia plataforma y arquitectura propuestas para realizar *pentesting*.

V. IMPLEMENTACIÓN

En esta sección se explora la implementación general de la propuesta, siguiendo los procesos implementados desde la creación de escenarios y retos, y la activación de un ciberejercicio y el proceso de realización del mismo. En estos procesos

Tabla II
DESCRIPCIÓN DE SUPERPARÁMETROS.

Superparámetro	Definición	Objetivo	Requerido	Dependencias	Valores
Tópico	Tema principal del ejercicio usado para el contenido relevante	Construir una plantilla tematizada para el ejercicio	X	X	Texto
Tipo	Tipo de ejercicio o caso de uso a generar	Crea una plantilla base para un caso de uso concreto	V	X	Caso de uso
Subtipo	Subtipo de ejercicio o caso de uso concreto a generar en la plantilla	Crea una plantilla base para la implementación de un caso de uso concreto	X	Tipo	Caso de uso concreto
Cantidad de usuarios	Cantidad de usuarios simulados	Genera una cantidad de usuarios aleatorios	X	X	Entero mayor a 0
Tráfico simulado	Frecuencia y cantidad de tráfico simulado	Configura la cantidad y frecuencia de interacciones generadas por NPC	X	X	Porcentaje 0-100
Divergencia de tópico	Divergencia entre el tópico del tráfico simulado y el tema principal	Configura la aparición de temas separados del tópico principal	X	X	Porcentaje 0-100
Divergencia de usuarios	Divergencia entre el comportamiento y personalidad de usuarios	Genera usuarios de diferente comportamiento y personalidad	X	X	Porcentaje 0-100
Nivel de configuración	Configuración automática de la plantilla	Genera plantillas a varios niveles de configuración	X	X	Porcentaje 0-100
Cantidad de bots	Porcentaje de bots automatizados	Genera un porcentaje de usuarios bot	X	Nivel de automatización	Porcentaje 0-100
Nivel de automatización	Porcentaje de humanidad de un bot	Configura bots con comportamientos de humanos o de programa.	X	Cantidad de bots	Porcentaje 0-100

intervienen los módulos comentados en la Sección III, cuya funcionalidad permiten automatizar y simplificar los procesos de generación y configuración de las instancias de las RRSS para generar ciberejercicios más eficientes, efectivos y realistas.

V-A. Módulo de generación de plantillas

El primer paso necesario es generar y configurar una plantilla base (Sección III-A) que servirá para aprovisionar a las instancias de la red social. Estas plantillas se pueden generar manualmente, pero requieren de grandes conocimientos sobre la herramienta debido a la extensa configuración posible. Para simplificar y automatizar este proceso de creación se proporciona un módulo que permite dicha generación de plantillas mediante el uso de unos pocos superparámetros (Sección III-A) que el instructor debe introducir. Este módulo está formado por una base de datos que mantiene fragmentos de plantillas, y que permite generar una plantilla base según dichos superparámetros. En la Tabla II se muestran los superparámetros junto a su definición, para qué sirven, si son requeridos u obligatorios, las dependencias entre otros superparámetros y el valor que debe ser asignado.

Según los superparámetros elegidos e introducidos por el instructor, el módulo de generación de plantillas de ciberejercicios recupera todos los fragmentos compatibles de una base de datos, y a continuación construye la plantilla base adaptándola a las especificaciones. Los elementos recurrentes de una plantilla base se muestran en la Tabla III, junto con su descripción y uso. Estas plantillas deben ser revisadas por el instructor, que decidirá si es válida o si debe ser modificada manualmente para el ciberejercicio que quiere desplegar (apartado III-B2).

V-B. Módulo de gestión de RRSS

Cuando el instructor genera y configura una plantilla base para el ciberejercicio solo queda activarlo. Una vez activado el se envían y cargan las plantillas designadas configuradas por el instructor a los despliegues seleccionados. El módulo de gestión de RRSS recibe y mapea las plantillas a objetos que almacenan y trabajan con la información proporcionada. También ejecuta las acciones y procedimientos necesarios para aprovisionar la red social con las configuraciones de las plantillas y actualizar la información para la gestión interna.

El procedimiento que sigue al cargar una plantilla se separa en dos fases. La primera fase consiste en recibir y tratar la plantilla base, recorriéndola junto sus elementos para convertirla en objetos. Se cargan las listas de comportamiento y personalidad, y se mapean los usuarios relevantes y aleatorios a objetos Usuario a los que se asocian otros objetos de Configuración de perfil, Relaciones, lista de Publicación, Comportamiento y Personalidad. En la segunda fase se recorren todos los objetos Usuario y se aprovisiona a la simulación con la configuración asignada. El proceso de aprovisionamiento de la segunda fase se divide en varias etapas:

1. Creación de usuarios y configuración de perfiles: Se crean a los usuarios y se generan sesiones con Mastodon individuales. Después se configuran los perfiles.
2. Generación de relaciones: Se recorren los usuarios y realizan las peticiones de follow, y se bloquean y silencian a otros usuarios. A continuación se vuelven a recorrer y se aceptan o rechazan las peticiones y notificaciones de follow, actualizando las relaciones.
3. Publicación de contenido: Se recorren los usuarios y se envía el contenido pregenerado o publicaciones.
4. Simulación de usuarios: Para cada nuevo usuario se lanza un hilo que ejecuta la simulación de tráfico de fondo en función de la personalidad y comportamiento

Tabla III
ELEMENTOS DE LA PLANTILLA BASE.

Elemento de plantilla	Definición	Uso
Plantillas predefinidas	Lista de otras plantillas base previamente creadas	Carga plantillas base de ejercicios configurados preexistentes
Usuarios aleatorios	Lista de configuraciones para la creación de usuarios	Configura y crea distintas cantidades de usuarios NPC con características comunes con elementos de plantilla
Usuarios relevantes	Lista de configuraciones de creación de usuarios relevantes	Crea usuarios relevantes configurados con otros elementos de plantilla
Configuración de Perfil	Configuración de perfil concreta para un usuario	Asigna la configuración de perfil de usuario, alias, imagen o privacidad
Relaciones	Configuración de relaciones del usuario a generar	Genera las relaciones base con otros usuarios como seguir, bloquear y silenciar
Publicaciones	Lista de publicaciones con el contenido necesario para generarlas	Crea las publicaciones pregeneradas de usuarios y las envía
Comportamiento	Configuración de las acciones realizables	Dota a usuarios de acciones y comportamientos concretos para la publicación e interacción
Funciones de comportamiento	Acciones que pueden realizar los NPC	Permiten asignar acciones a los usuarios NPC
Personalidad	Descripción psicológica de un usuario	Dota a usuarios de personalidad para la interpretación y generación de contenido

asignados. La ejecución de estos hilos se incluye en el módulo de simulación NPC.

V-C. *Módulo de simulación de NPC*

Este módulo permite simular a los usuarios dadas sus personalidades y sus comportamientos asignados. Las personalidades dictan los temas que gustan y disgustan al usuario, son las que determinan el tema y tipo de contenido que publican. Los comportamientos dictan las acciones que pueden realizar para generar el tráfico de fondo. El módulo consiste en la ejecución de hilos por cada usuario, que realizan un proceso en bucle, seleccionando acciones por categorías según la funcionalidad requerida en el comportamiento. Primero se realizan las acciones de publicación, tanto públicas como privadas. Después se realizan las acciones relativas a las notificaciones recibidas. A continuación se tratan las publicaciones de la línea temporal pública, y finalmente se ejecutan las acciones de control y actualización.

V-D. *Módulo de configuración en caliente*

Mientras que el ciberejercicio está en ejecución y los despliegues están operativos puede requerirse realizar cambios sobre dichas instancias, para ello necesitamos un módulo de configuración en caliente. Este módulo permite modificar las simulaciones de RRSS en ejecución. Para ello se deben desarrollar varios mecanismos que permitan añadir nuevos elementos mediante el uso de plantillas (nuevas personalidades, comportamientos, usuarios y publicaciones) a despliegues concretos o agrupados. También debe permitir enviar acciones de control que actúen sobre dichas simulaciones. Algunas acciones de control pueden ir directamente sobre el funcionamiento de red social, el tráfico de fondo, los usuarios NPC o relevantes, etc.

VI. DEMOSTRACIÓN DE FUNCIONAMIENTO

En esta sección me muestra el resultado del aprovisionamiento de una plantilla configurada con un caso de desinformación que podría causar una crisis a escala mundial en distintos ámbitos. La plantilla base está formada principalmente por una plantilla de comportamiento, una plantilla de personalidad, un usuario relevante configurado con publicaciones desinformativas y un perfil concreto configurado, y dos usuarios aleatorios autoconfigurados. En la Figura 3 se muestra

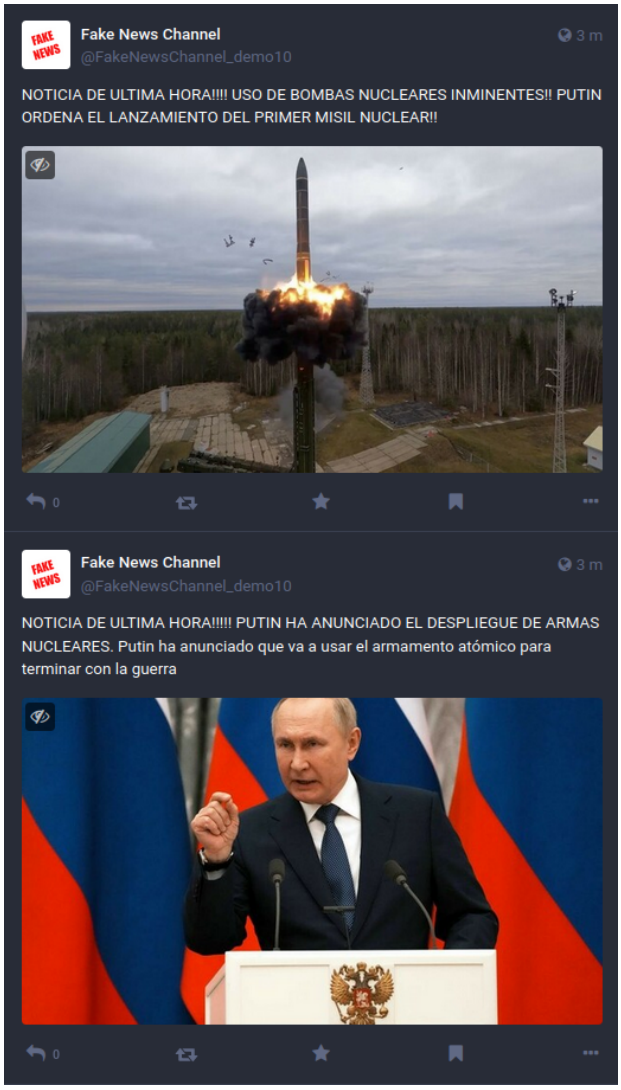

Figura 2. Desinformación - Putin ordena el uso de misiles nucleares.

que los usuarios NPC *andrew99* y *ublevins* se han creado y registrado, y que han obtenido alias y avatares propios (entre otros). También se puede observar como los usuarios *ublevins* y *andrew99* impulsan y dan me gusta respectivamente a publicaciones creadas previamente por el administrador

(*MegaAdmin* o *tfmastodonAdmin*) de la instancia, mediante el módulo de simulación de NPC.

En la Figura 3 también aparece el usuario relevante, Fake News Channel, que publicará las noticias de desinformación. Tiene asignado un fragmento de plantilla configuración de perfil donde se le asigna un fondo, un avatar y un alias. Para las publicaciones se le asigna un fragmento de plantilla de publicaciones con una lista de dos noticias con texto e imágenes sobre Vladímir Putin anunciando el uso de armamento nuclear para la guerra, tal y como se muestra en la Figura 2.

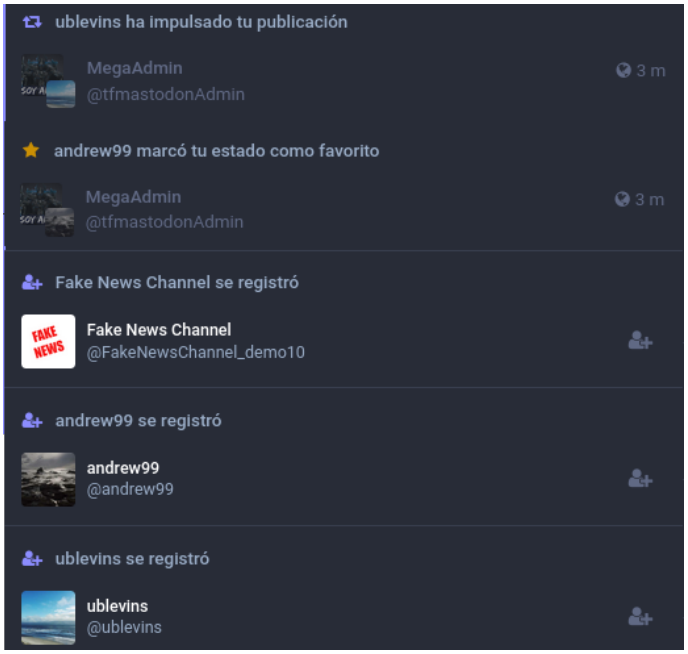

Figura 3. Ejemplo de un caso de desinformación - generación de usuarios.

VII. Conclusiones y trabajo futuro

En este artículo se ha presentado una extensión de un Cyber Range para entrenar competencias de ciberseguridad en RRSS. Es la propuesta de una herramienta de entrenamiento y formación para civiles, profesionales y militares, donde se muestra la arquitectura necesaria para el correcto despliegue así como la automatización en la generación y gestión de ciberejercicios. La identificación de los casos de uso para la preparación de competencias en el marco Cyber Range permiten que se realicen simulaciones con distintos grados de dificultad y objetivos de aprendizaje según tópicos y objetivos. La propuesta da la oportunidad de generar retos ultraconfigurables, extensibles y multidisciplinares, además de la generación automática de simulaciones utilizando las plantillas base mostradas en la Sección V como elemento básico o personalizable, y la posibilidad de añadir NPC con características específicas. Con el estado actual de la herramienta se ha llevado a cabo una demostración básica de un caso de uso de desinformación con el que se pretende mostrar la importancia y criticidad que puede tener contenido falso y su posible difusión en las RRSS.

Respecto a las lineas de investigación que se generan se destacan principalmente la posibilidad de integración de otras plataformas de código abierto para simular otras plataformas como Facebook, Youtube, Telegram, etc.; la creación de un módulo totalmente automatizado para la generación de contenido dinámico que permita obtener y/o crear contenido cuando se requiera; la creación de un módulo para la recolección y evaluación de datos sobre los estudiantes; la investigación y declaración de nuevos casos de uso simulables; se plantea la posibilidad de mejorar las simulaciones mediante el uso de inteligencias artificial (IA) para los usuarios en base a personalidades, comportamientos e interacciones para crear escenarios lo más realistas posible; y también se propone la conexión con otras herramientas de simulación basada en agentes de RRSS para comprobar el alcance y éxito que tendrían los estudiantes al crear una contramedida en algunos tipos de ejercicios como los de desinformación o gestión de crisis.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Universidades español vinculado a la Unión Europea a través del programa NextGenerationEU, con cargo a la beca postdoctoral Margarita Salas (172/MSJD/22).

Referencias

[1] STO-TR-HFM-ET-356, "Mitigating and Responding to Cognitive Warfare," NATO, STO TECHNICAL REPORT, Marzo 2023. [Online]. Available: https://www.sto.nato.int/publications/STO%20Technical%20Reports/STO-TR-HFM-ET-356/$$TR-HFM-ET-356-ALL.pdf.

[2] J. Pastor-Galindo, M. Zago, P. Nespoli, S. L. Bernal, A. H. Celdrán, M. G. Pérez, J. A. Ruipérez-Valiente, G. M. Pérez, and F. G. Mármol, "Spotting political social bots in twitter: A use case of the 2019 spanish general election," *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 2156–2170, 2020.

[3] "Los 10 vectores de ataque más utilizados por los ciberdelincuentes," https://www.incibe.es/empresas/blog/los-10-vectores-ataque-mas-utilizados-los-ciberdelincuentes, accessed: May 17, 2023.

[4] C. Colwill, "Human factors in information security: The insider threat–who can you trust these days?" *Information security technical report*, vol. 14, no. 4, pp. 186–196, 2009.

[5] L. Hadlington, "Human factors in cybersecurity; examining the link between internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours," *Heliyon*, vol. 3, no. 7, p. e00346, 2017.

[6] C. Geeng, S. Yee, and F. Roesner, "Fake news on facebook and twitter: Investigating how people (don't) investigate," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–14.

[7] R. Petersen, D. Santos, M. C. Smith, K. A. Wetzel, and G. Witte, "Workforce Framework for Cybersecurity (NICE Framework)," NIST, Tech. Rep., 2020. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181r1.pdf

[8] D. Pérez and E. Argente, "Simulating users in a social media platform using multi-agent systems," in *Hybrid Artificial Intelligent Systems*, E. A. de la Cal, J. R. Villar Flecha, H. Quintián, and E. Corchado, Eds. Cham: Springer International Publishing, 2020, pp. 486–498.

[9] S. Braidley, "Extending our cyber-range cyran with social engineering capabilities," *Master's thesis. De Montfort University, Leicester, England*, 2016.

[10] "Somulator - NTNU," https://www.ntnu.no/ncr/somulator, accessed: April 20, 2023.

[11] "Project NATO Stratcom COE," https://prevency.com/en/projects/project-nato-stratcom-coe/, accessed: April 20, 2023.

[12] P. Nespoli, M. Albaladejo-González, J. A. Pastor Valera, J. A. Ruipérez-Valiente, and F. Gómez Mármol, "Capacidades avanzadas de simulación y evaluación con elementos de gamificación," in *VII Jornadas Nacionales de Investigación en Ciberseguridad (JNIC '22)*, 2022, pp. 55–62.