TRABAJOS DEL FORO CONTRA LAS CAMPAÑAS DE DESINFORMACIÓN

INICIATIVAS 2024

CAPÍTULO 4

INGENIERÍA DE LA DESINFORMACIÓN: INFRAESTRUCTURA TECNOLÓGICA DE LAS OPERACIONES DIGITALES EN CAMPAÑAS DE MANIPULACIÓN

ÍNDICE

CAPÍTULO 4

INGENIERÍA DE LA DESINFORMACIÓN: INFRAESTRUCTURA TECNOLÓGICA DE LAS OPERACIONES DIGITALES EN CAMPAÑAS DE MANIPULACIÓN

Coordinadores:

Fran Casino

Ministerio del Interior – Oficina de Coordinación de Ciberseguridad (OCC)

Autores y colaboradores:

Marc Almeida Ros

David Arroyo Guardeño

Ivan Homoliak

Andrés Marín López

Rafael Mata Milla

Constantinos Patsakis

Oscar Walsch

INTRODUCCIÓN

El capítulo tendrá por objeto realizar un doble ejercicio de trasvase de conocimiento involucrando tres dominios que, si bien tienen características específicas, en la actualidad suelen aparecer combinados en el contexto concreto de las operaciones de influencia y en campañas de manipulación. En efecto, tal y como se pone de relieve en (Gioe y Smith, 2024, capítulo 8), elementos de desestabilización desde la esfera cibernética pueden aprovechar situaciones de conflicto o violencia para amplificar sus efectos, de forma que la interrelación entre ciberespionaje, cibercrimen y la ciberguerra es cada día más patente (Shapiro, 2023).

En este capítulo analizaremos las fases de la denominada *killchain* según define Mitre para el caso de las *Advanced Persistent Threats* (APT). Dicho análisis estar a acompañado de una evaluación a alto nivel de los diferentes estadios vinculados a *Advanced Persistent Manipulations* (APM). Se establecerá una relación entre el dominio cibernético de las APT y el informacional de las APM. Tal relación se llevará a término poniendo de relieve tanto elementos de infraestructura de redes como servicios de internet empleados para diseñar y ejecutar ataques de desinformación contra objetivos específicos. El enfoque adoptado en el capítulo pretende aplicar el conjunto de lecciones aprendidas en el caso de la ciberseguridad al dominio FIMI (*Foreign Information Manipulation and Interference*), evaluando a través de las diversas tecnologías y procedimientos cómo se diseñan estrategias, tácticas y operaciones contra objetivos concretos en base a su matriz de vulnerabilidades. Dichas vulnerabilidades engloban no sólo aspectos tecnológicos (Mirza et al., 2023), sino que también incluyen el conjunto de elementos psico-sociales que hacen más plausible emplear narrativas concretas contra una sociedad o segmento social específico.

A lo largo del capítulo la caracterización de herramientas, servicios y plataformas pondrá de relieve el carácter dual del ecosistema digital. Así, herramientas que en primera instancia son diseñadas para proteger la privacidad de los usuarios al navegar en Internet se constituyen en piezas clave de las estrategias de atacantes en operaciones de información de cara a bien desplegar ofensivas bajo falsa bandera, bien simplemente para dificultar por parte de los objetivos una posterior "ciberatribución". Ésta será dependiente de la capacidad de extraer indicadores y entidades a partir de trazas de sistemas, monitorización de redes y de información en foros en abierto, pero también en la *Deep Web* y la *Darknet*. La conexión entre actores, entre actores y medios, y entre actores e intenciones nos permitirá tanto comprender fenómenos FIMI como anticiparlos. De esta forma, en el capítulo se contextualizarán el conjunto de herramientas, servicios y estándares para relacionar conexiones casuales entre los diversos elementos del ecosistema FIMI y el uso de elementos y técnicas de infraestructura de red.

RELACIÓN ENTRE LOS ATAQUES INFORMACIONALES Y LAS ESTRATEGIAS AVANZADAS DE CIBERATAQUE

La denominada cadena de ataque o *killchain* describe las distintas fases de ataque de APT de acuerdo con el paradigma definido por MITRE ATT&ACK (Strom et al., 2018). Los ataques APT que cuentan con clara y definida esponsorización por parte de estados suelen estar combinados con estrategias híbridas de desestabilización. Así, las fases de exfiltración de información en la cadena de ataque de APT puede nutrir la configuración de las fases de selección de objetivos y de elección de medios y herramientas para desplegar operaciones de influencia y manipulación (Ahmad et al., 2019). En este sentido, cabe destacar el carácter crítico de toda brecha de datos, en especial aquellas que afectan a datos personales que habilitan el perfilado de ciudadanos (The UN Report on Disinformation: a role for privacy, 2021). Dicho perfilado es una información de gran relevancia en la configuración de ataques de ingeniería social y en la fabricación de contenidos viralizantes y que aprovechen sesgos, preferencias y vulnerabilidades psico-sociales de potenciales objetivos (Shapiro, 2023).

A continuación, analizaremos en detalle la relación entre los diversos componentes y elementos del ecosistema del *malware* de los cuales se nutren tantos las APT como las acciones de influencia en el ámbito FIMI. Para ello emplearemos como referencia el esquema de la Figura 1, detallando el nivel de criticidad del mercado de datos y de herramientas en la *Dark Web* (*Dark Web Marketplace*, DWM), el acceso y uso de software malicioso y de paquetes de utilidades para realizar phishing (*PhaaS*, *Phishing as a Service*) y ataques de ingeniería social. Además, se pondrá de relieve el carácter dual de tecnologías como la inteligencia artificial (*Malicious AI Generated Content*, MAIGC) y las fuentes abiertas de inteligencia (*Open source Intelligence*, OSINT).

Reconocimiento	OSINT
Ingeniería social	MAIGC
Despliegue inicial	PHAAS
Movimiento lateral	MAAS
Cifrado y exfiltración	RAAS
Recolección y secuestro	DWM

Figura 1: Estructura integral de las APT

MALWARE COMO SERVICIO (MaaS)

El uso instrumental de información y de software por parte de agentes extranjeros constituye un elemento cada vez más presente en el denominado fenómeno FIMI. El análisis en profundidad de incidentes como la inestabilidad provocada por los servicios de inteligencia rusos en 2019 en Georgia (Greenberg, 2020) pone de relieve estrategias complejas de ataque que combinan el uso de software malicioso y operaciones de influencia e información.

La configuración del dominio cruzado entre lo cibernético (los bits) y lo informacional (los bytes) queda claramente patente en la estrategia y operativa de ataque puesta en marcha desde 2006 por Rusia en Georgia (Beehner et al., 2018), siendo especialmente significativa toda la dinámica de generación y distribución de software malicioso o *malware* a partir del dominio StopGeorgia.ru.

"La configuración del dominio cruzado entre lo cibernético (los bits) y lo informacional (los bytes) queda claramente patente en la estrategia y operativa de ataque puesta en marcha desde 2006 por Rusia en Georgia"

El *malware* es una amenaza crítica para las organizaciones y las personas en todo el mundo. En las últimas décadas, hemos sido testigos de la evolución del *malware* desde simples programas diseñados para causar disrupciones menores, hasta convertirse en herramientas sofisticadas empleadas en operaciones de cibercrimen altamente organizadas. En este contexto, el *MaaS* ha emergido como un modelo de negocio lucrativo que facilita a los delincuentes cibernéticos el acceso a herramientas de ataque complejas sin necesidad de poseer habilidades técnicas avanzadas (Patsakis et al., 2024). En el contexto específico de FIMI, hay que tener en cuenta los servicios y proveedores del ecosistema de producción y distribución de *malware* como potenciales *proxies* para diseñar, desplegar y coordinar estrategias combinadas de desestabilización de un adversario (Borghard y Lonergan, 2016a). Es aquí donde el *MaaS* adquiere relevancia significativa e importancia creciente, siendo de especial importancia la dificultad creciente de realizar atribución y actividad FIMI y de evitar ser víctimas de ataques de falsa bandera (Skopik y Pahi, 2020).

El *MaaS* es un modelo de negocio en el que desarrolladores de *malware* ofrecen sus herramientas y servicios a otros delincuentes a cambio de un pago, generalmente en forma de criptomonedas para mantener el anonimato (Casino et al., 2021). Este modelo opera de manera similar al Software como Servicio (*SaaS*) en la industria legítima, donde los usuarios pue-den suscribirse o alquilar software sin necesidad de gestionarlo o desarrollarlo por sí mismos. En el caso de *MaaS*, los "clientes" son delincuentes que desean ejecutar ataques cibernéticos, pero carecen de las habilidades técnicas necesarias para desarrollar *malware* por su cuenta. Al ofrecer servicios como kits de *malware*, infraestructura de comando y control (C&C por sus siglas en inglés), servicios de ofuscación y evasión, y soporte técnico, el *MaaS* ha democratizado el acceso al cibercrimen, haciéndolo más accesible y rentable para una amplia gama de actores (Davidson, 2021), incluyendo aquellos que directamente o indirectamente participan en FIMI (Borghard y Lonergan, 2016a).

De cara a realizar la correcta conexión entre estrategias combinadas de acción informacional y cibernética en FIMI partiremos de marcos de referencia basados en estándares como DAD-CDM[1] y otras iniciativas encuadradas dentro de la plataforma OASIS[2]. La adecuada articulación de los

1 https://dad-cdm.org/
2 https://www.oasis-open.org/

estándares definidos en el marco de OASIS permite crear grafos de relaciones entre actores y entre actores y acciones, lo que habilita el análisis de causalidad y/o ciberatribución. A continuación, describimos en detalle los componentes, actores y funcionamiento de *MaaS*.

Componentes del MaaS

El ecosistema de *MaaS* está compuesto por varios componentes clave que facilitan la operación y éxito de este modelo de negocio:

- **Kits de *Malware***: Estos son paquetes de software que contienen todo lo necesario para ejecutar un ataque. Los kits de *malware* suelen incluir el código malicioso, instrucciones detalladas sobre cómo utilizarlo, y en algunos casos, herramientas adicionales como *exploit kits* o *loaders* que ayudan a distribuir el *malware*. Estos kits están diseñados para ser fáciles de usar, permitiendo que incluso aquellos con poca experiencia técnica puedan ejecutar ataques efectivos (Meland, et al., 2020).

- **Infraestructura de C&C**: La infraestructura de C&C es esencial para la mayoría de los tipos de *malware*, especialmente aquellos que requieren comunicación continua con el atacante. Los servidores de C&C permiten a los atacantes gestionar de manera remota las máquinas infectadas, enviar comandos, exfiltrar datos y actualizar el *malware* en respuesta a nuevas medidas de seguridad. Los proveedores de *MaaS* suelen ofrecer acceso a servidores de C&C como parte de su paquete, lo que facilita la gestión y control de las campañas de *malware* (Huang, et al., 2018).

- **Servicios de ofuscación y evasión**: Para que el *malware* sea efectivo, debe ser capaz de evadir la detección por parte de las soluciones de seguridad. Los proveedores de *MaaS* a menudo incluyen servicios de ofuscación, que alteran el código del *malware* para hacerlo menos detectable por los sistemas de antivirus. Además, pueden implementar técnicas de evasión, como el uso de empaquetadores personalizados o la inserción de código en procesos legítimos para evitar ser detectados (Patsakis et al., 2024).

- **Soporte técnico y actualizaciones**: En el modelo *MaaS*, los desarrolladores de *malware* no solo venden su software, sino que también ofrecen soporte técnico para garantizar que los clientes puedan utilizarlo de manera efectiva. Esto puede incluir ayuda con la configuración del *malware*, la resolución de problemas, y la provisión de actualizaciones para hacer frente a nuevas medidas de seguridad. Al igual que en el SaaS, los proveedores de *MaaS* buscan mantener satisfechos a sus clientes para asegurar una fuente constante de ingresos (Huang et al., 2018).

Actores Clave en el Ecosistema de MaaS

El ecosistema de *MaaS* se nutre de una variedad de actores, cada uno desempeñando un papel específico en la cadena de valor del cibercrimen. A continuación, se describen los principales actores involucrados:

- **Desarrolladores de *Malware*:** Estos son los actores más técnicos en el ecosistema de *MaaS*. Son responsables de crear y mantener el software malicioso que se distribuye a través de los mercados de *MaaS*. Los desarrolladores de *malware* suelen ser programadores altamente cualificados con un profundo conocimiento de las vulnerabilidades del software, las técnicas de evasión de seguridad y las estrategias de ataque cibernético. Estos individuos o grupos a menudo operan en la clandestinidad, aprovechando foros de la *Darknet* y otros canales privados para vender sus productos (U.S. Department of Justice, 2022).

- **Operadores de *MaaS*:** Los operadores de *MaaS* son responsables de gestionar la infraestructura que sustenta estos servicios. Esto incluye no solo la distribución de kits de *malware*, sino también el mantenimiento de servidores y la provisión de servicios de soporte y actualización. Los operadores de *MaaS* actúan como intermediarios entre los desarrolladores de *malware* y los clientes, asegurando que los primeros reciban su compensación mientras los segundos obtienen las herramientas necesarias para llevar a cabo sus ataques (Europol, 2021).

- **Afiliados:** Los afiliados son un grupo clave dentro del ecosistema *MaaS*. A menudo carecen de las habilidades técnicas para desarrollar su propio *malware*, pero están dispuestos a distribuir el *malware* proporcionado por los operadores de *MaaS*. A cambio, los afiliados reciben una comisión ba-sada en los ingresos generados por las campañas de *malware* que llevan a cabo. Un ejemplo paradigmático de este modelo es el "*Ransomware*-as-a-Service" (RaaS), el cual permite a los operadores de *ransomware* maximizar su alcance y ganancias sin necesidad de involucrarse directamente en la distribución del *malware* (Europol, 2023a).

- **Clientes/Criminales:** Los clientes de *MaaS* son los actores que adquieren servicios de *malware* para llevar a cabo sus propias operaciones delicti-vas. Estos pueden ser individuos o grupos organizados, y sus motivaciones varían desde el lucro financiero hasta la venganza personal o el espionaje. Los clientes de *MaaS* pueden carecer de habilidades técnicas avanzadas, pero gracias a los servicios proporcionados por los operadores de *MaaS*, pueden lanzar ataques efectivos con una inversión mínima en términos de tiempo y recursos (Cable, s.f.).

- **Intermediarios y revendedores:** Además de los actores directamente involucrados en la creación y distribución de *malware*, existen intermediarios y revendedores que facilitan las transacciones entre los diferentes actores del ecosistema. Estos intermediarios pueden ofrecer servicios como el cambio de criptomonedas, el acceso a servidores comprometidos, o la reventa de kits de *malware* a nuevos clientes (Europol, 2023b).

Funcionamiento de los Mercados de MaaS

Los mercados de *MaaS* han adoptado muchas de las características de los negocios virtuales legítimos, lo que facilita a los delincuentes la compra de servicios de *malware*. Estos mercados suelen operar en la *Darknet*, donde los proveedores de *MaaS* publican anuncios detallados de sus productos y servicios, que incluyen descripciones de las funcionalidades del *malware*, precios, y en algunos casos, reseñas de otros usuarios.

- ***Ransomware*-as-a-Service (RaaS):** El RaaS es uno de los modelos más prevalentes en los mercados de *MaaS*. En este modelo, los operadores de *ransomware* proporcionan el software necesario para cifrar los datos de las víctimas y exigir un rescate a cambio de la clave de descifrado. Los afiliados se encargan de distribuir el *ransomware*, a menudo a través de correos electrónicos de *phishing* o mediante la explotación de vulnerabilidades en sitios web o aplicaciones. A cambio, los afiliados reciben un porcentaje de los pagos de rescate que logran recolectar. Este modelo ha demostrado ser extremadamente rentable, tanto para los operadores de RaaS como para sus afiliados (Meland et al., 2020).

- **Exploit Kits y *Phishing*-as-a-Service:** Además del *ransomware*, los mercados de *MaaS* también ofrecen kits de exploits y servicios de *phishing*. Los *exploit kits* son paquetes de software que explotan vulnerabilidades en software popular (como navegadores web o sistemas operativos) para instalar *malware* en las máquinas de las víctimas. El *Phishing-as-a-Service*, por otro lado, proporciona plantillas de correos electrónicos y páginas web fraudulentas que los delincuentes pueden utilizar para robar credenciales de inicio de sesión u otra información sensible. Estos servicios son populares porque permiten a los delincuentes lanzar ataques dirigidos sin necesidad de desarrollar sus propias herramientas (Meland et al., 2020).

- **Servicios de *botnets*:** Las *botnets* son redes de ordenadores comprometidos que los delincuentes pueden controlar de manera remota. En los mercados de *MaaS*, es posible alquilar acceso a *botnets* para realizar una variedad de ataques, incluidos DDoS, el envío masivo de spam, o el minado de criptomonedas. Los operadores de *botnets* mantienen y actualizan las redes comprometidas, asegurando que sigan siendo efectivas y difíciles de detectar por las soluciones de seguridad. Este servicio permite a los delincuentes escalar sus operaciones rápidamente y sin necesidad de infraestructura propia (Huang et al., 2018).

Desafíos y Amenazas para la Seguridad Cibernética

El crecimiento de *MaaS* plantea numerosos desafíos para la seguridad cibernética. La naturaleza descentralizada y anónima de estos servicios dificulta la identificación y persecución de los responsables. Además, la facilidad de acceso a herramientas de ataque sofisticadas ha reducido la barrera de entrada al cibercrimen, permitiendo que más actores participen en actividades delictivas sin necesidad de un conocimiento técnico profundo.

- **Evolución Rápida de las Amenazas:** Uno de los desafíos más significativos que presenta el *MaaS* es la rápida evolución de las amenazas. Los desarrolladores

de *malware* y los operadores de *MaaS* están en constante competencia con las soluciones de seguridad, actualizando y mejorando sus productos para evitar la detección. Esto significa que las organizaciones deben estar en alerta constante y actualizar regularmente sus sistemas para protegerse contra las nuevas variantes de *malware* (Casino et al., 2022).

- **Impacto Económico y Social**: El cibercrimen facilitado por *MaaS* tiene un impacto devastador en la economía global. Además de los costos directos asociados con la recuperación de ataques de *malware*, las organizaciones también enfrentan pérdidas de reputación, pérdida de confianza de los clientes, y posibles sanciones legales y regulatorias. En el ámbito social, los ataques facilitados por *MaaS* también pueden poner en peligro la infraestructura crítica, como los servicios de salud, energía y transporte, con consecuencias potencialmente catastróficas (Freeze, 2022).

- **Colaboración Internacional y Medidas de Mitigación**: Para enfrentar la amenaza del *MaaS*, es crucial una colaboración internacional. Los cibercriminales operan a menudo desde múltiples jurisdicciones, lo que dificulta su persecución por parte de las autoridades nacionales. Las agencias de seguridad deben trabajar juntas para compartir información y coordinar esfuerzos para desmantelar estas redes. Además, las organizaciones deben adoptar un enfoque proactivo en su ciberseguridad, invirtiendo en tecnologías avanzadas de detección y respuesta, y educando a sus emplea-dos sobre las amenazas más recientes (Europol, 2021; Casino et al., 2022).

INFRAESTRUCTURA Y CAMPAÑAS DE DESINFORMACIÓN

Los actores de amenaza que realizan campañas de desinformación necesitan usar estructuras y medios técnicos desde los que apoyarse para lograr sus objetivos, siendo especialmente relevante todo el conjunto de herramientas, servicios y plataformas que permiten diferir la causa de una estrategia FIMI de su consecuencia. A lo largo de esta sección haremos énfasis en los recursos tecnológicos que permiten ofuscar la actividad FIMI, impidiendo la identificación de actores e intereses.

Tal y como aparece recogidos en trabajos como (Huang et al., 2018), en el contexto actual existe una paulatina división de tareas y de especialización de las herramientas y servicios para cibercrimen, en sentido global, pero también en el caso particular de actividades ilícitas en la fabricación, instrumentalización y explotación de contenido fabricado o descontextualizado. Dicho de otra forma, en los últimos 10 años se ha producido una transformación *fordiana* del ecosistema del cibercrimen y de la guerra cognitiva y ataques reputacionales.

"En los últimos 10 años se ha producido una transformación fordiana del ecosistema del cibercrimen y de la guerra cognitiva y ataques reputacionales."

Estructura y elementos

Algunos ejemplos de las estructuras y medios técnicos usados son los siguientes:

- **VPN y *proxies* residenciales.** Una VPN, del inglés Virtual Private Network, permite navegar a través de Internet impidiendo la interceptación de la información que intercambiamos. Si bien el uso de servicios de VPN no garantiza el anonimato, pueden ser usados como una capa más de seguridad que proteja la identidad final del actor de amenaza, aumentando la dificultad de su atribución. El esfuerzo por parte de fuerzas y cuerpos de seguridad del estado a la hora de llevar a cabo investigaciones sobre direcciones IP que procedan de servicios de VPN es costoso en tiempo y en coordinación, especialmente si los propios servicios de VPN no son cooperativos y siguen políticas de no almacenamiento de registros de actividad de sus usuarios o clientes. Como se destacará más adelante a destacar más adelante, aquí existe la circunstancia dual de que proteger la privacidad puede ayudar a dificultar labores de atribución en investigaciones de ilícitos en el dominio ciber y, concretamente, en el ecosistema FIMI. Es más, si la infraestructura de red usada está distribuida de forma internacional, entonces la labor del investigador requiere la coordinación entre distintos marcos jurisdiccionales que no son siempre compatibles y que dificultan la necesaria colaboración.

  En el caso de los servidores *proxy*, un usuario se conecta a este servicio para no acceder de forma directa a un servicio o plataforma en Internet. El servidor *proxy* actúa de intermediario entre un cliente y un servidor final, de forma que se añade una capa extra entre ambos. Existen diversos tipos de servidores *proxy*, siendo especialmente relevantes los servidores *proxy* anónimos y los residenciales. En los servidores *proxy* anónimos la dirección IP del cliente es modificada, de forma que la plataforma o servicio al que accede un cliente no conocerá la dirección de origen. Esto permite enmascarar la acción de ataque de un cibercriminal o de un actor en campañas de desinformación, pero el grado de protección dependerá del tipo de servidor. En el caso de los *proxies* residenciales el servidor está ubicado en emplazamientos concretos no dependientes de ningún centro de datos o proveedor de servicios, de forma que una autoridad judicial tendría más dificultades de cara a conseguir la colaboración de este proveedor de servicio *proxy*.

  Al igual que en el caso de las VPN, el uso de *proxies* residenciales puede favorecer el acceso a información restringida de forma geográfica. De igual forma, su uso por actores de amenaza puede intentar engañar en una investigación sobre el origen real de la amenaza, al situarse en terceros países que también pueden tener interés en la difusión de desinformación. El uso de *proxies* residenciales es parte del arsenal desplegados, por ejemplo, por APT29, grupo que ha estado especialmente activo en Ucrania antes de la crisis de 2014, y volvió a realizar campañas activas de *phishing* a partir de finales de 2018. En el marco de FIMI, es especialmente importante la conexión entre APT29 y acciones diplomáticas por parte de Rusia (Cunningham, 2020).

- ***SIM Swapping* y SMS *Phishing*.** El intercambio de SIM se ha asociado de forma habitual a estafas relacionadas con las finanzas, sin embargo, adquiere un nuevo papel en las campañas de desinformación. Además de poder conceder acceso total o parcial a un dispositivo o servicio, y por tanto ser usado con fines de ciberespionaje o para acceder a otros elementos relacionados con la fuente que contengan

información de interés, esta técnica podría permitir la adquisición de medios de difusión legítimos, como cuentas de usuario en redes sociales que sean reputadas y cuenten con una gran audiencia de su interés.

Está técnica tiene otras implicaciones al permitir tanto la difusión de desinformación como la desacreditación de la víctima suplantada. De la misma forma que el *SIM Swapping*, el *phishing* puede ser usado para adquirir acceso a una cuenta legítima, permitiendo la difusión de información a una mayor audiencia. También permite la adquisición de información real que puede ser manipulada para realizar desinformación, como en el caso de David Satter, cuya cuenta de correo fue accedida de forma ilegal, lo que permitió que se modificaran sus correos electrónicos con vistas a publicar información falsa en relación con una supuesta operación financiada por EE. UU. para desestabilizar a Rusia (Hulcoop et al., 2017).

El ecosistema del *phishing* mediante SMS ha adquirido también una especial relevancia en los últimos tiempos. El conjunto de sistemas de alojamiento de servidores para campañas de *phishing* y de ingeniería social da una idea del grado de sofisticación de este tipo de actividad (Nahapetyan et al., 2024), así como del conjunto de retos en lo relativo a la supervisión de proveedores de servicios y de plataformas según lo establecido por el acta europea de servicios digitales o Digital Services Act (DSA). La recolección de servicios de hosting, patrones de generación de certificados digitales mediante el análisis de los logs de *Certificate Transparency*, o la identificación y trazabilidad de kits de desarrollo de campañas de *phishing* son elementos cruciales de cara al desarrollo de estrategias de contención frente a FIMI.

- **Inteligencia Artificial Generativa y suplantación (*Human Spoofing*).** El uso de Inteligencia Artificial permite generar desinformación tanto basada en texto como multimedia. Los *Large Language Models*, o Modelos de Lenguaje Extensos, pueden generar información verosímil y convincente sobre el asunto a desinformar. Permiten además la generación masiva de desinformación, llevando entonces a la infoxicación. La IA puede ser también usada para el refinamiento de la desinformación, intentando que la información sea más persuasiva, mejorando su calidad, o adaptándola a la audiencia objetivo. Aunque la mayoría de los modelos de lenguaje estén censurados o prevengan este tipo de conductas, son fácilmente eludibles por medio de distintas técnicas, cadenas de pensamiento, o el uso de modelos no censurados o protegidos, entre otros (Barman et al., 2024).

  En el caso de la generación de imágenes, si antes generalmente se necesitaba una imagen que poder descontextualizar, ahora directamente se puede fabricar. Esto es especialmente interesante con modelos como Flux, que no contemplan por ejemplo la censura de personajes reconocidos, y donde la dificultad para diferenciar las imágenes generadas de las reales es cada vez mayor. El acceso a estas tecnologías es fácilmente asequible, ya que no requiere conocimientos técnicos, y las especificaciones requeridas para ejecutar este software no son tan demandantes como se puede pensar. El *Human Spoofing*, mediante el uso de IAs generativas como las reseñadas anteriormente, pone a disposición de actores de desinformación la posibilidad de suplantar de forma muy verosímil a personajes públicos, ya no solo para la difusión del contenido creado, sino para el

"El Human Spoofing, mediante el uso de IAs generativas como las reseñadas anteriormente, pone a disposición de actores de desinformación la posibilidad de suplantar de forma muy verosímil a personajes públicos"

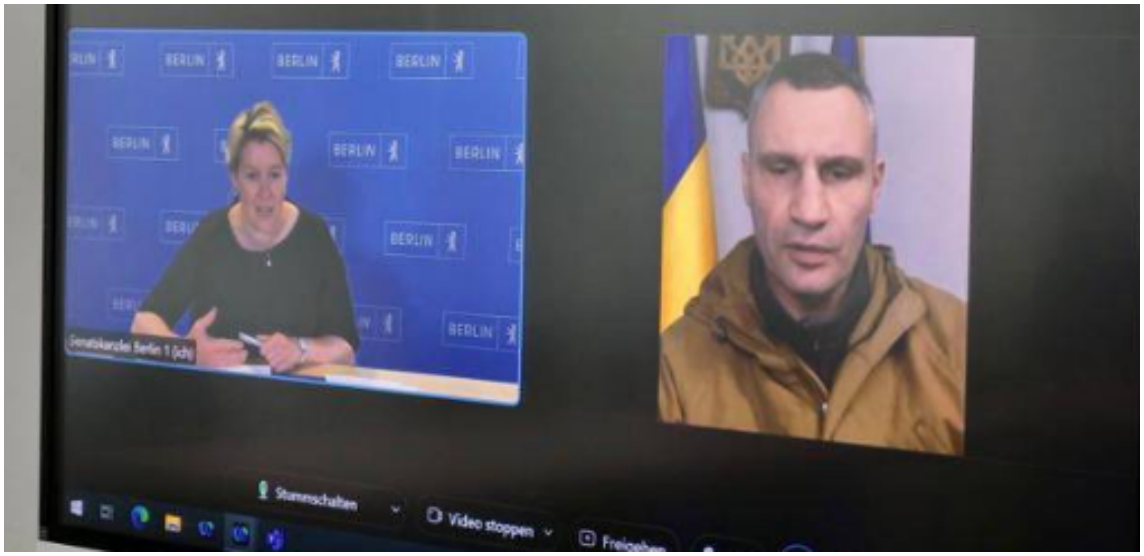

Figura 2. Alcaldesa de Berlín siendo manipulada mediante la suplantación de identidad del alcalde de Kiev Vitali Klitschko. Fuente: Fishcer (2022).

engaño y el uso de ingeniería social para con otros. Un ejemplo de esto es la falsa videollamada mantenida entre el alcalde de Madrid y su supuesto homólogo en Kiev, un acto que podría ser catalogable entre la broma y la guerra híbrida (Twomey et al., 2023).

- **Redes sociales, influencia y búsqueda de talento.** El uso combinado de redes sociales, las dinámicas multiplataforma y el despliegue de acciones coordinadas. La democratización en el acceso a canales de generación y distribución de contenido facilita el despliegue de operaciones para influir en la toma de decisión. Desde el punto de vista de su investigación, los servicios de mensajería instantánea con cifrado de extremo a extremo suponen en este sentido un gran desafío, en la medida que impiden una interceptación directa de tráfico e impulsan coordinación entre fuentes y medios de manipulación (Hoseini et al., 2024; Hanley y Durumeric, 2024). Entre las aplicaciones de redes sociales, plataformas e incluso foros en la *Deep Web*, cabe destacar el despliegue de infraestructura para la formación y reclutamiento de talento en la esfera del cibercrimen (Wang et al., 2023; Pandey, 2022).

Tecnologías de uso dual

Definimos tecnologías de uso dual como aquellas tecnologías que pueden ser usadas tanto para fines civiles como militares. En este caso, hacemos uso del término para referirnos a aquellas tecnologías que, además de su uso civil, permiten ser usadas por actores de amenaza para realizar operaciones de desinformación. Las tecnologías de uso dual por excelencia son las redes sociales. Sin embargo, en este apartado nuestra atención se dirige hacia la infraestructura necesaria para la realización de operaciones de desinformación.

- **Marketing digital.** El marketing digital se ha consolidado como una herramienta esencial para las empresas y organizaciones que buscan promocionar sus productos y servicios. A través de estrategias como el posicionamiento en buscadores (SEO) y la publicidad en redes sociales, las marcas pueden llegar a audiencias específicas de manera eficiente y efectiva. Sin embargo, las mismas técnicas y plataformas utilizadas para fines comerciales pueden ser empleadas para llevar a cabo operaciones de influencia y desinformación. Actores de amenaza pueden aprovechar las herramientas de segmentación y análisis de datos para identificar y dirigirse a audiencias vulnerables, difundiendo información falsa o manipulada para influir en opiniones, comportamientos o decisiones políticas (Domenico et al., 2021).

  Por ejemplo, durante procesos electorales, actores estatales o no estatales han utilizado campañas de desinformación en redes sociales para sembrar discordia, polarizar a la sociedad o desacreditar a candidatos. Mediante la creación de contenido engañoso y el uso de *bots* y perfiles falsos, pueden amplificar mensajes y hacerlos parecer más legítimos o populares de lo que realmente son. Así, existen grupos organizados que están financiados por estados o prestan sus servicios para efectuar campañas de manipulación en redes sociales. Este es el caso de las denominadas granjas de *trolls* y/o *bots* (Hughes y Waismel-Manor, 2021), o del uso de infraestructura militar en desuso para relacionar actividades de soporte para el cibercrimen y la ciberguerra (Caesar, 2020). Además, técnicas de microsegmentación o los sistemas de recomendación (Deldjoo et al., 2024) permiten adaptar mensajes específicos a grupos particulares, aumentando la efectividad de la manipulación (O Fathaigh et al., 2021).

- **Wikipedia.** La enciclopedia cooperativa es uno de los sitios web más visitados globalmente, sirviendo como fuente de información accesible y gratuita para millones de personas. La participación de cualquier persona en su edición es a su vez su principal ventaja y desventaja. Desde la óptica de las campañas de desinformación e influencia, esta capacidad es susceptible de ser explotada por actores de amenaza mediante la edición maliciosa y el sesgo de contenido, los ataques de edición coordinados, la manipulación de fuentes y referencias, o la divulgación de fuentes y referencias aparentemente legitimas pero que constituyen parte de la campaña de desinformación.

  Aunque la mayoría de los engaños en Wikipedia se detectan rápidamente y tienen poco impacto, un pequeño número de ellos sobreviven mucho tiempo y reciben muchas visitas (Kumar, West, y Leskovec, 2016). Existen formas de clasificar automáticamente si un artículo dado es engañoso, consiguiendo una precisión mayor al de la revisión manual realizada por moderadores humanos. Según el

artículo anteriormente citado, el lector humano tiende a considerar los artículos cortos como contenido engañoso, mientras que en realidad es en grandes artículos donde es más propenso el engaño, así como se demuestra que la capacidad para evadir la moderación de contenido malicioso es baja sin herramientas automáticas especializa-das. Esto hace de la Wikipedia un escenario idóneo para la divulgación de desinformación, dadas sus características de edición y acceso abierto, y el masivo tráfico que recibe.

- **Open Source Intelligence.** La inteligencia de fuentes abiertas, OSINT por sus siglas en inglés, ha constituido desde hace años una gran fuente de información gracias a su fácil accesibilidad, gran disponibilidad de información y bajo coste económico en comparación con otro tipo de fuentes. La proliferación de la tecnología y el acceso masivo a internet han transformado la manera en que se recopila y analiza la información. La inteligencia de fuentes abiertas se ha convertido en una herramienta esencial para extraer datos relevantes de un océano de información disponible públicamente. Su facilidad para ser obtenida ha permitido que sea activamente explotada tanto por servicios de inteligencia como por organizaciones o particulares, consiguiendo ventajas estratégicas y una mejora en la toma de decisiones en diversos sectores.

  Aunque su uso está ampliamente extendido y aceptado, es necesario reconocer la importancia de la desinformación y manipulación en este campo. Un actor de amenaza podría realizar infoxicación mediante la generación de contenido artificial erróneo o ligeramente falso, así como fabricación de información sesgada que pueda ser malinterpretada por actores enemigos o competidores (Flamer, 2023).

- **Internet Archive.** Los servicios de archivado web, como *Wayback Machine* (del Internet Archive) o Archive.is, permiten almacenar páginas web de manera automatizada o a demanda, posibilitando que cualquier usuario acceda a la información incluso si ha sido eliminada o modificada, o simplemente sin visitar el sitio original. Estos servicios desempeñan un papel crucial en la preservación de la historia digital, facilitando la investigación académica y el acceso a información que, de otro modo, podría perderse.

  Sin embargo, aunque su propósito legítimo es evidente, se ha observado que algunos actores malintencionados utilizan esta tecnología con diversos fines, como la propagación de información errónea, retractada o desinformación; el acceso a noticias de medios contrarios a sus ideales con el fin de reducir los ingresos publicitarios de dichos medios; la evasión de medidas de censura durante la difusión de contenido desinformativo en redes sociales; y la captura de publicaciones en redes sociales y noticias que podrían ser eliminadas debido a controversias (Zannettou et al., 2018; Acker y Chaiet, 2020) .

- **Uso de servicios cloud legítimos como estrategia de evasión.** El uso de servicios legítimos para llevar a término ciberataques es algo que se ha venido observando de forma habitual en el caso del despliegue de *botnets* y sistemas C&C (Al lelah et al., 2023). Dado el carácter excesivamente central de algunas plataformas, ca-be estimar como de alta criticidad todo uso instrumental de los servicios de dicha plataforma para eludir los sistemas de filtrado y de control de seguridad de instituciones, organizaciones y empresas (Alcantara, 2024).

DISCUSIÓN, RETOS Y DESAFÍOS

Esta sección resume las ideas obtenidas del análisis anterior, destacando las implicaciones de las campañas de desinformación y ofreciendo una perspectiva sobre los pasos futuros. Los puntos principales de discusión se centran en los avances tecnológicos necesarios para mitigar la amenaza de la desinformación.

Desarrollo de estándares y procedimientos para la identificación de actores, herramientas y terceros interpuestos en el dominio FIMI. Desde el punto de vista de la atribución de campañas y acciones de injerencia externa, la asociación entre estados y actores del ecosistema MaaS constituye un reto tecnológico y metodológico. La extensión de los tradicionales proxies en el dominio del enfrentamiento tácito y militar se amplifica gracias a los medios cibernéticos. La sofisticación y especialización en lo relativo a servicios, productos y plataformas para creación y distribución de contenido fabricado y malware contribuye a incrementar el carácter poliédrico del concepto de proxy o, mejor dicho, de ciber-proxy (Borghard y Lonergan, 2016b). El análisis de riesgos y amenazas asociados a los diversos tipos de proxies en el ámbito cibernético demanda el desarrollo de procedimientos de identificación, anotación y distribución de evidencias e inteligencia a nivel nacional y transnacional. En este sentido, a nivel europeo sería interesante incorporar a la red de ISAC soluciones tecnológicas derivadas de marcos teóricos para el análisis de riesgo y la atribución de acciones FIMI.

Firma Digital y Protocolos Seguros para la Verificación de Contenidos. Las campañas de desinformación pueden combatirse eficazmente adoptando protocolos seguros y firmas digitales que autentiquen el origen y la integridad del contenido. Técnicas como los Entornos de Ejecución Confiable (TEE) y las Pruebas de Conocimiento Cero (ZK-SNARKs) están emergiendo como opciones viables para verificar la autenticidad de la información. Estas técnicas criptográficas garantizan que se conserve la autoría original del contenido digital, reduciendo el potencial de que se difunda contenido manipulado o fabricado sin control. Implementar tales protocolos en entornos donde la IA Generativa puede fabricar desinformación será crucial en el futuro.

"Las campañas de desinformación pueden combatirse eficazmente adoptando protocolos seguros y firmas digitales que autentiquen el origen y la integridad del contenido"

Un caso de estudio bien conocido de lucha contra la desinformación utilizando zk-SNARKs es la protección de imágenes digitales. Las cámaras necesitan un elemento seguro para producir firmas en imágenes o videos capturados. Este elemento seguro actúa de manera similar a un TEE, aunque se describe con mayor precisión como un sistema de firma a prueba de manipulaciones o elemento seguro. Algunas compañías ya han contribuido a un estándar conocido como firmas C2PA (Coalition for Content Provenance and Authenticity) (C2PA, 2024), que no se centra necesariamente en hardware seguro, sino en vincular contenido multimedia con metadatos confiables, como la geolocalización. C2PA permite a los usuarios equilibrar privacidad y autenticidad, por ejemplo, adjuntando la geolocalización a una imagen mientras revelan selectivamente la información que eligen compartir. De esta manera, cualquiera con acceso a una imagen de alta resolución, como una foto de 30MP y su firma correspondiente, puede demostrar que una cámara autenticada capturó la imagen en un lugar específico. Este mecanismo ayuda a combatir la desinformación, como la difusión de fotos falsas de zonas de conflicto. Posibles ataques, como tomar una foto

de una imagen impresa, siguen siendo un vector de ataque separado que debe abordarse independientemente.

Preservación de la Integridad durante modificaciones de contenido. Las agencias de noticias y redes de televisión a menudo modifican fotos y videos originales antes de publicarlos. Modificaciones como recortar, cambiar el tamaño o convertir a escala de grises resultan en la pérdida de la vinculación original entre el medio firmado y sus metadatos. Esto presenta un desafío significativo para mantener la integridad del contenido durante el proceso de modificación. Para abordar este problema, zk-SNARKs pueden emplearse para preservar la vinculación entre la imagen original y las modificaciones. Definiendo un circuito que represente las transformaciones aplicadas al medio original, es posible retener la firma y la vinculación, incluso después de que el contenido haya sido alterado. Varios autores han discutido este enfoque (Datta et al., 2024), aunque uno de los mayores desafíos identificados fue la sobrecarga computacional significativa asociada con la generación de pruebas zk-SNARK. Su método requería grandes cantidades de memoria, hasta 64GB, para generar una sola prueba, lo que lo hacía poco práctico para una adopción generalizada. Sin embargo, un artículo más reciente (Della Monica et al., 2024) optimizó este enfoque aplicando un principio de dividir y con-quistar. Propusieron dividir una imagen en mosaicos y modificar el protocolo C2PA para firmar mosaicos agregados utilizando una estructura de árbol de Merkle. Este método permite la generación de pruebas zk-SNARK para mosaicos individuales, reduciendo significativamente la carga computacional. Como resultado, este enfoque optimizado permite la generación de pruebas zk-SNARK incluso en hardware común, como solo 4GB de RAM.

Si bien los enfoques para imágenes estáticas han mostrado resultados prometedores, el contenido de video presenta un desafío computacional mucho mayor ya que hace que la generación de pruebas zk-SNARK sea significativamente más lenta y demande más recursos. Se están llevando a cabo investigaciones para paralelizar el proceso de generación de pruebas zk-SNARK utilizando hardware de GPU, lo que ha mostrado mejoras de velocidad de hasta 4 veces. Sin embargo, estos esfuerzos aún no logran la eficiencia computacional requerida para el procesamiento en tiempo real o a gran escala de videos.

Detección automatizada de desinformación usando IA y Big Data. Las LLMs y su capacidad para generar información falsa convincente plantea un desafío. La sobrecarga de información ya es un problema prevalente, y la capacidad de la IA para producir grandes cantidades de información falsa agravará este problema (Xu et al., 2023). Abordar la sobrecarga de información requerirá tanto soluciones técnicas (p. ej., mejores algoritmos de filtrado y sistemas automatizados de verificación) así como esfuerzos educativos para mejorar la alfabetización mediática de la sociedad. En este contexto, la automatización de la detección de desinformación mediante sistemas basados en IA que aprovechen correlaciones semánticas y técnicas de procesamiento de lenguaje natural (NLP) es fundamental. Sin embargo, aún quedan desafíos por abordar en cuanto a la escalabilidad y precisión de dichos sistemas. Además, el mantenimiento de bases de datos continuamente actualizadas para rastrear fuentes conocidas de desinformación jugar´ un papel vital en la efectividad de estos sistemas de IA (Mansurova et al., 2024).

Expertos en la verificación de noticias. Si bien los sistemas automatizados juegan un papel importante, no se puede ignorar la participación de expertos en la verificación de la exactitud y legitimidad de contenido. El rol de verificadores, periodistas y expertos de diferentes campos debe fortalecerse para contrarrestar la creciente influencia de la desinformación generada por IA. Integrar la retroalimentación de expertos en los sistemas de IA para señalar con-tenido falso podría crear un sistema híbrido robusto (Mahmud et al., 2023). Además, el concepto de “experto” en sí

requiere definiciones que puedan usarse para identificar al personal adecuado y sus habilidades y comportamientos, entre otras características, en diferentes contextos de información. Esto constituye una línea de investigación en sí misma.

Tecnologías emergentes en la protección de contenido verificado: Blockchain y Registros Inmutables. La correcta curación de noticias a través de expertos, su eficiente anotación y distribución requiere de la existencia de protocolos y procedimientos que garanticen su custodia e integridad. La tecnología Blockchain ofrece otra solución potencial para contrarrestar la desinformación mediante la creación de registros digitales inmutables de con-tenido. Esto permitiría rastrear la información a medida que se propaga en diferentes plataformas, garantizando que las alteraciones o manipulaciones del contenido original sean visibles, rastreables y verificables. Este enfoque podría ser particularmente efectivo en plataformas de noticias y redes sociales, donde la información falsa puede difundirse rápidamente (Fraga-Lamas y Fernandez-Carames, 2020).

Colaboración multidisciplinaria en la lucha contra la desinformación. El enfoque de modelado de amenazas en ciberseguridad puede servir para caracterizar mejor el perfil de atacantes en el arco de la desinformación, sus patrones de ataque, sus objetivos preferenciales y las técnicas que usan de forma más habitual (Mirza et al., 2023). Ahora bien, para ello se requiere un enfoque colaborativo que abarque múltiples disciplinas. Expertos en ciberseguridad, fuerzas del orden, análisis de datos, psicología y análisis semántico deben trabajar juntos para desarrollar contramedidas efectivas. Esta colaboración interdisciplinaria también debe extenderse a los legisladores, asegurando que el marco legal evolucione junto con los avances tecnológicos (Casino, et al., 2022). Los protocolos diseñados para detectar y mitigar la desinformación deben estar fundamentados en estándares legales, proporcionando tanto privacidad como seguridad (Ramasauskaite, 2023).

Desafíos en la atribución cibernética. Uno de los desafíos persistentes en la lucha contra la desinformación es el tema de la atribución cibernética. Identificar la fuente de la desinformación, especialmente en casos que involucran campañas patrocinadas, requiere herramientas sofisticadas y cooperación internacional. Si bien se han logrado avances, particularmente con el uso de IA y OSINT, las tácticas en evolución de los actores maliciosos hacen que la atribución sea cada vez más difícil. Las operaciones de falsa bandera, infraestructuras de Internet anonimizadas (p. ej., VPN, proxies) y el uso de canales de comunicación encriptados dificultan la labor de identificar a los actores de campañas de desinformación. Los esfuerzos futuros deben centrarse en mejorar los marcos de atribución cibernética para responsabilizar a los responsables a nivel internacional (Maesschalck, 2024).

CONCLUSIONES

Este capítulo ha explorado la compleja interacción entre las campañas de desinformación, los avances tecnológicos y los impactos sociopolíticos resultan-tes. A lo largo del capítulo, hemos visto cómo la desinformación y las campañas cibernéticas sofisticadas se nutren a menudo de infraestructuras organizadas y plataformas que aprovechan el malware y las tecnologías de IA. Las APM reflejan las etapas de las ATP, donde la guerra de información opera junto con los ciberataques, difundiendo contenido manipulado que puede potencialmente desestabilizar a la sociedad.

El MaaS ejemplifica la mercantilización de herramientas de cibercrimen, que permiten incluso a actores no técnicos orquestar campañas de desinformación. Este proceso se refuerza a través de la evolución de botnets, ransomware y modelos de phishing como servicio. En paralelo, mientras que las herramientas de IA permiten la generación de contenido falso o modificado a gran escala, también allanan el camino para su detección. En resumen, identificamos varias tecnologías y estrategias clave que deben ser exploradas, como el uso de blockchain para la integridad del contenido, la mejora de la eficiencia de zk-SNARK, la necesidad colaboración interdisciplinaria y, finalmente, abordar los desafíos de la atribución cibernética.

REFERENCIAS BIBLIOGRÁFICAS

Acker, A., y Chaiet, M. (2020, 28 de septiembre). The weaponization of web archives: Data craft and COVID-19 publics. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-41

Ahmad, A., Webb, J., Desouza, K. C., y Boorman, J. (2019). Strategically-motivated advanced persistent threat: Definition, process, tactics and a disinformation model of counterattack. *Computers & Security, 86*, 402– 418.

Alcantara, J. M. (2024, 23 de marzo). Phishing with cloudflare workers: Transparent phishing and html smuggling. *Netskope*. https://www.netskope.com/blog/phishing-with-cloudflare-workers-transparent-phishing-and-html-smuggling

Al lelah, T., Theodorakopoulos, G., Reinecke, P., Javed, A., y Anthi, E. (2023). Abuse of cloud-based and public legitimate services as command-and-control (c&c) infrastructure: a systematic literature review. *Journal of Cybersecurity and Privacy, 3 (3)*, 558–590.

Barman, D., Guo, Z., y Conlan, O. (2024). The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination. *Machine Learning with Applications, 16*, *100545*. https://doi.org/10.1016/j.mlwa.2024.100545

Beehner, L., Collins, L., Ferenzi, S., Person, R., y Brantly, A. F. (2018). Analyzing the russian way of war: Evidence from the 2008 conflict with Georgia. *Modern War Institute*. Disponible en: https://mwi.westpoint.edu/wp-content/uploads/2018/03/Analyzing-the-Russian-Way-of-War.pdf

Borghard, E. D., y Lonergan, S. W. (2016a). Can states calculate the risks of using cyber proxies? *Orbis, 60 (3)*, 395-416. https://doi.org/10.1016/j.orbis.2016.05.009

Borghard, E. D., y Lonergan, S. W. (2016b). Can states calculate the risks of using cyber proxies? *Orbis, 60 (3)*, 395–416.

C2PA. (2024). Coalition for content provenance and authenticity specification. *c2pa.org*. https://c2pa.org/specifications/specifications/2.0/specs/C2PA_Specification.html

Cable, J. (s.f.). Ransomwhere — ransomwhe.re. https://ransomwhe.re/.

Caesar, E. (2020, 27 de julio). The cold war bunker that became home to a dark-web empire. *The New Yorker.* https://www.newyorker.com/magazine/2020/08/03/the-cold-war-bunker-that-became-home-to-a-dark-web-empire

Casino, F., Dasaklis, T. K., Spathoulas, G. P., Anagnostopoulos, M., Ghosal, A., Borocz, I., Solanas, A., Conti, M., Patsakis, C. (2022). Research trends, challenges, and emerging topics in digital forensics: A review of reviews. *IEEE Access, 10*, 25464-25493. https://ieeexplore.ieee.org/document/9720948

Casino, F., Lykousas, N., Katos, V., y Patsakis, C. (2021). Unearthing mali-cious campaigns and actors from the blockchain dns ecosystem. *Computer Communications, 179*, 217–230.

Casino, F., Pina, C., López-Aguilar, P., Batista, E., Solanas, A., y Patsakis, C. (2022). Sok: Cross-border criminal investigations and digital evidence. *Journal of Cybersecurity, 8 (1)*. https://doi.org/10.1093/cybsec/tyac014

Cunningham, C. (2020). *A Russian Federation Information Warfare Primer*. The Henry M. Jackson School of international Studies. University of Washington. https://jsis.washington.edu/news/a-russian-federation-information-warfare-primer/

Datta, T., Chen, B., y Boneh, D. (2024). VerITAS: Verifying image transformations at scale. *Cryptology ePrint Archive*. https://eprint.iacr.org/2024/1066

Davidson, R. (2021). The fight against *malware* as a service. *Network Security, 2021 (8)*, 7–11. https://doi.org/10.1016/S1353-4858(21)00088-X

Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., y Zanzonelli, D. (2024). Fair-ness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction, 34 (1)*, 59–108. https://doi.org/10.48550/arXiv.2205.11127

Della Monica, P., Visconti, I., Vitaletti, A., y Zecchini, M. (2024). Trust nobody: Privacy-preserving proofs for edited photos with your laptop. *Cryptology ePrint Archive*. https://eprint.iacr.org/2024/1074

Domenico, G. D., Sit, J., Ishizaka, A., y Nunan, D. (2021, enero). Fake news, social media and marketing: A systematic review. Journal of Business Research, 124 , 329–341. https://doi.org/10.1016/j.jbusres.2020.11.037

Europol. (2021). DarkMarket: world's largest illegal dark web marketplace ta-ken down. https://www.europol.europa.eu/media-press/newsroom/news/darkmarket-worlds-largest-illegal-dark-web-marketplace-taken-down

Europol. (2023a). 288 dark web vendors arrested in major marketplace seizure. https://www.europol.europa.eu/media-press/newsroom/news/288-dark-web-vendors-arrested-in-major-marketplace-seizure

Europol.(2023b).Takedown of notorious hacker marketplace selling your identity to criminals. https://www.europol.europa.eu/media-press/newsroom/news/takedown-of-notorious-hacker-marketplace-selling-your-identity-to-criminals

Fishcer, D. (2022, 1 de julio). Fake Video Calls Aim to Harm Ukraine. *Refugees Human Rights Watch*. https://www.hrw.org/news/2022/07/01/fake-video-calls-aim-harm-ukraine-refugees

Flamer, N. (2023). 'The enemy teaches us how to operate': Palestinian hamas use of open source intelligence (osint) in its intelligence warfare against Israel (1987-2012). *Intelligence and National Security, 38 (7)*, 1171–1188. https://dx.doi.org/10.1080/02684527.2023.2212556

Fraga-Lamas, P., y Fernandez-Carames, T. M. (2020). Fake news, disinformation, and deepfakes: Leveraging distributed ledger technologies and block-chain to combat digital deception and counterfeit reality. *IT professional, 22 (2)*, 53–59.

Freeze, D. (2022). Cybercrime to cost the world 8 trillion annually in 2023. *Cybercrime Magazine*. https://cybersecurityventures.com/cybercrime-to-cost-the-world-8-trillion-annually-in-2023/

Gioe, D. V., y Smith, M. W. (2024). *Great power cyber competition: Competing and winning in the information environment*. Routledge. Taylor & Francis.

Greenberg, A. (2020, 20 de febrero). The US Blames Russia's GRU for Sweeping Cyberattacks in Georgia. *Wired*. https://www.wired.com/story/us-blames-russia-gru-sweeping-cyberattacks-georgia/

Hanley, H. W., y Durumeric, Z. (2024). Partial Mobilization: Tracking Multilingual Information Flows amongst Russian Media Outlets and Telegram. *Proceedings of the International AAAI Conference on Web and Social Media, 18*, 528–541. https://doi.org/10.48550/arXiv.2301.10856

Hoseini, M., de Freitas Melo, P., Benevenuto, F., Feldmann, A., y Zannettou, S. (2024). Characterizing Information Propagation in Fringe Communities on Telegram. *Proceedings of the International AAAI Conference on Web and Social Media, 18*, 583–595. https://doi.org/10.1609/icwsm.v18i1.31336

Huang, K., Siegel, M., y Madnick, S. (2018). Systematically understanding the cyber attack business: A survey. *ACM Computing Surveys (CSUR), 51(4).* https://doi.org/10.1145/3199674

Hughes, H. C., y Waismel-Manor, I. (2021). The Macedonian Fake News Industry and the 2016 US Election. *Political Science & Politics, 54 (1)*, 19-23. https://doi.org/10.1017/S1049096520000992

Hulcoop, A., Scott-Railton, J., Tanchak, P., Brooks, M., y Deibert, R. (2017). *Tainted Leaks: Disinformation and Phishing with a Russian Nexus*. Citizen Lab Research Report, University of Toronto. https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/

Kumar, S., West, R., y Leskovec, J. (2016). Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. *Proceedings of the 25th International Conference on World Wide Web,* 591-602. https://doi.org/10.1145/2872427.2883085

Maesschalck, S. (2024). Gentlemen, you can't fight in here. Or can you?: How cyberspace operations impact international security. *World Affairs, 187(1)*, 24-36. https://doi.org/10.1002/waf2.12004

Mahmud, M. A. I., Talukder, A. T., Sultana, A., Bhuiyan, K. I. A., Rahman, M. S., Pranto, T. H., y Rahman, R. M. (2023). Toward News Authenticity: Synthesizing Natural Language Processing and Human Expert Opinion to Evaluate News. *IEEE Access, 11*, 11405-11421. https://doi.org/10.1109/ACCESS.2023.3241483

Mansurova, A., Mansurova, A., y Nugumanova, A. (2024). QA-RAG: Exploring LLM Reliance on External Knowledge. *Big Data and Cognitive Computing, 8(9)*, 115. https://doi.org/10.3390/bdcc8090115

Meland, P. H., Bayoumy, Y. F. F., y Sindre, G. (2020). The Ransomware-as-a-Service economy within the darknet. *Computers & Security, 92*, 101762. https://doi.org/10.1016/j.cose.2020.101762

Mirza, S., Begum, L., Niu, L., Pardo, S., Abouzied, A., Papotti, P., y Pöpper, C. (2023). Tactics, threats & targets: Modeling disinformation and its mitigation. *Network And Distributed System Security (NDSS) Symposium*. https://doi.org/10.14722/ndss.2023.23657

Nahapetyan, A., Prasad, S., Childs, K., Oest, A., Ladwig, Y., Kapravelos, A., y Reaves, B. (2024). On SMS *phishing* tactics and infrastructure. *2024 IEEE Symposium on Security and Privacy (SP)*, 1-16. https://doi.org/10.1109/SP54263.2024.00169.

O Fathaigh, R., Dobber, T., Zuiderveen Borgesius, F., y Shires, J. (2021). Micro-targeted propaganda by foreign actors: An interdisciplinary exploration. *Maastricht Journal of European and Comparative Law, 28 (6)*, 856–877.

Pandey, R. (2022). Exploring HackTown: A College for Cybercriminals. *Isaca.*org. https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2022/exploring-hacktown-a-college-for-cybercriminals

Patsakis, C., Arroyo, D., y Casino, F. (2024). The malware as a service ecosystem. *arXiv preprint*. https://doi.org/10.48550/arXiv.2405.04109

Patsakis, C., Casino, F., y Lykousas, N. (2024). Assessing LLMs in Malicious Code Deobfuscation of Real-world Malware Campaigns. *Expert Systems with Applications, 256 (6)*, 124912. http://dx.doi.org/10.1016/j.eswa.2024.124912

Ramasauskaite, O. (2023). *The role of collaborative networks in combating digital disinformation*. 2. International Conference on Economics "Regional Development - Digital Economy" : proceedings book. December 21-23, 2023 / Baku, Azerbaijan / The Scientific-Research Institute of Economic Studies under the Azerbaijan State University of Economics (UNEC), 432-437. https://vb.mruni.eu/object/elaba:184652082/

Shapiro, S. J. (2023). *Fancy Bear Goes Phishing: The Dark History of the Information Age, in Five Extraordinary Hacks*. Random House.

Skopik, F., y Pahi, T. (2020). Under false flag: using technical artifacts for Ccyber attack attribution. *Cybersecurity, 3*, 8. https://doi.org/10.1186/s42400-020-00048-4

Strom, B. E., Applebaum, A., Miller, D. P., Nickels, K. C., Pennington, A. G., y Thomas, C. B. (2018). Mitre att&ck: Design and philosophy. En Technical report. The MITRE Corporation.

Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., y Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLOS ONE, 18 (10)*, e0291668. https://doi.org/10.1371/journal.pone.0291668

The UN report on disinformation: a role for privacy. (2021, 17 de mayo). Privacyinternational.org. https://privacyinternational.org/news-analysis/4515/un-report-disinformation-role-privacy

U.S. Department of Justice. (2022). *Justice Department Investigation Leads to Shutdown of Largest Online Darknet Marketplace*. https://www.justice.gov/opa/pr/justice-department-investigation-leads-shutdown-largest-online-darknet-marketplace

Wang, Y., Roscoe, S., Arief, B., Connolly, L., Borrion, H., y Kaddoura, S. (2023). The Social and Technological Incentives for Cybercriminals to Engage in Ransomware Activities. En Arief, B., Monreale, A., Sirivianos, M., Li, S. (Eds.), *Security and Privacy in Social Networks and Big Data. SocialSec 2023. Lecture Notes in Computer Science*, 14097. Springer, Singapore. https://doi.org/10.1007/978-981-99-5177-2_9

Xu, D., Fan, S., y Kankanhalli, M. (2023). Combating misinformation in the era of generative AI models. En Proceedings of the 31st acm international conference on multimedia. *Association for Computing Machinery,* pp. 9291-9298. https://doi.org/10.1145/3581783.3612704

Zannettou, S., Blackburn, J., De Cristofaro, E., Sirivianos, M., y Stringhini, G. (2018, junio). Understanding Web Archiving Services and Their (Mis)Use on Social Media. *Proceedings of the International AAAI Conference on Web and Social Media, 12 (1)*. https://doi.org/10.1609/icwsm.v12i1.15018