Modelos matemáticos para la gestión de la ciberseguridad

David Ríos Insua

david.rios@icmat.es

Académico Numerario de la RAC

Instituto de Ciencias Matemáticas, ICMAT-CSIC, Madrid

José Manuel Camacho

josemanuel.camacho@icmat.es

Instituto de Ciencias Matemáticas, ICMAT-CSIC, Madrid

Palabras clave: ciberseguridad, ciberataques, gestión de riesgos, análisis de riesgo adversarios.

RESUMEN:

Algunas de las principales amenazas a las que debemos enfrentarnos en el mundo contemporáneo van asociadas a los ciberataques. A pesar de su potencial impacto negativo, los principales marcos de gestión de riesgos en ciberseguridad se basan en matrices de riesgos, con defectos que pueden conducir a asignaciones de recursos subóptimas. En el trabajo se presentan de forma divulgativa algunos modelos matemáticos que facilitan un mejor tratamiento de los riesgos en el dominio de la ciberseguridad.

ABSTRACT:

Some of the major global threats in modern times are related to cyber attacks. In spite of their potentially very negative impacts, the main cybersecurity risk management frameworks are based on relatively simplistic methods based on risk matrices with defects that may lead to suboptimal resource allocations. We present here several mathematical models that provide improved risk management in the cyber security domain.

1. INTRODUCCIÓN

Los ciberataques y, más concretamente, su potencial efecto, amenazan nuestras vidas y activos, no sólo los referidos a la información. Como ejemplo, el World Economic Forum señala consistentemente en los últimos años a los ciberataques, y otros fenómenos relativos a la ciberseguridad, como una de las principales amenazas en sus Global Risk Maps (World Economic Forum, 2021) anuales. En estas representaciones, se ilustran gráficamente riesgos globales para la sociedad de diversa índole a través de estimaciones de su verosimilitud de que se produzcan y de su impacto esperado. Por ejemplo, la representación del último año señala los fallos de ciberseguridad como una de las amenazas más probables y potencialmente más impactantes. Análogamente, la Estrategia Nacional de Seguridad incluye aspectos relacionados con la ciberseguridad entre las amenazas principales para nuestro país desde ya hace varios años (Ministerio de Defensa, 2017).

En los medios podemos encontrar también numerosos ejemplos de ciberataques notorios en los últimos años. Uno de los más relevantes es el reciente Sunburst (Tidy, 2020), que representa perfectamente la magnitud del desafío que supone la ciberseguridad. Sus autores lograron acceder a los sistemas de las principales agencias gubernamentales estadounidenses, y otras organizaciones en todo el mundo, a través de una actualización de uno de sus proveedores de software, logrando tener acceso a información confidencial sensible. Este caso pone de manifiesto que los ataques pueden realizarse a través de los proveedores de la organización objetivo final del ciberataque; la dimensión del problema se agrava si se tiene en cuenta que algunas empresas y administraciones pueden requerir los servicios de cientos, o incluso miles, de proveedores o terceros que los cibercriminales pueden emplear para infiltrarse.

Un aspecto que dificulta la evaluación y la gestión de los ciberriesgos es la aparición constante de nuevos actores y tipos de ciber riesgos, incluso recientemente utilizando métodos de la inteligencia artificial (IA) para implementar ataques aún más complejos y sofisticados. En esta línea, tuvo gran repercusión mediática un caso en el que unos ciberdelincuentes emplearon métodos de IA para simular la voz del CEO de una empresa y engañar a un empleado para que este les transfiriera dinero (Dawson, 2019).

En este contexto, han aparecido distintos estándares y marcos, como MAGERIT (Amutio et al., 2012) o ISO27001 (International Organization for Standardization, 2015), para gestionar la ciberseguridad. Tales marcos facilitan en gran manera tareas como catalogar amenazas, activos o atacantes. Sin embargo, en los aspectos referidos a análisis de riesgos se basan esencialmente en el empleo de matrices de riesgos con importantes y bien conocidos defectos, véase p. ej. (Cox, 2008). Por ello, describiremos de forma divulgativa en este trabajo diversos modelos matemáticos que nos permiten desarrollar aproximaciones más rigurosas en el tratamiento de la ciberseguridad.

2. CONTEXTO

En esta sección describimos diversas cuestiones contextuales en relación con el panorama actual de la ciberseguridad. También se describe el perfil cambiante de los individuos y organizaciones que implementan las ciberamenazas, lo que justifica que se llegue a considerar al ciberespacio como el quinto dominio de operaciones militares (Instituto Español de Estudios Estratégicos, 2012).

Aumento en el número e impacto de los ciberataques

Tal aumento, producido en los últimos años, ha hecho que la ciberseguridad se convierta en una de las principales preocupaciones a nivel global (World Economic Forum, 2020). Como ejemplo, la empresa de seguridad americana McAfee cataloga alrededor de 680 nuevas amenazas por minuto (Beek et al., 2021), siendo éstas cada vez más sofisticadas y difíciles de detectar. Además, el daño producido por los ciberataques es cada vez mayor, estimándose que, sólo en el año 2018, estos tuvieron un impacto de unos 600 mil millones de dólares (Lewis, 2018). Este fenómeno también se aprecia en los medios, donde son habituales términos como virus informático, troyano, spyware, ransomware o amenaza persistente avanzada, entre muchos otros.

Además, debe esperarse que la actual tendencia no decaiga, sino todo lo contrario, dado que muchos de los sistemas e infraestructuras relevantes para el funcionamiento de nuestra sociedad están cada vez más influidos por las tecnologías de la información y las comunicaciones (TIC). En efecto, las empresas, los gobiernos y los individuos somos cada vez más dependientes de Internet, haciendo que la superficie de ciberataques crezca y, con ella, el número e impacto de los mismos. Dispositivos como los sistemas de votación, los sistemas médicos, las infraestructuras críticas o los vehículos autónomos son susceptibles de ser dañados por un ciberataque, provocando inmensos daños, no solo económicos, sino también relacionados con la salud, el medio ambiente o la seguridad nacional, entre otros.

En el campo de la ciberseguridad, son cada vez más relevantes los *ataques dirigidos*, que tienen, típicamente, consecuencias mayores. Como ejemplo, en el ataque a Equifax en 2017 se robaron los datos de más de 140 millones de clientes (Federal Trade Commission, 2020). También son

notorios los ataques denominados como *ransomware,* un tipo de ciberataque que limita el acceso a determinadas partes o datos del sistema infectado y exige un rescate para liberar el mismo. Entre estos, destaca WannaCry que en el 2017 dejó inoperativo, entre otros, al servicio de sanidad británico, Telefónica y FedEx, provocando pérdidas estimadas en 4000 millones de dólares (Berr, 2016). Distintos gobiernos e instituciones públicas también se vieron afectados por ataques ransomware; en el año 2018, los servicios públicos de Atlanta fueron atacados, provocando daños de los que la ciudad tardó meses en recuperarse (Hatmaker, 2018).

Además, se espera que el problema se agrave con la creciente importancia del Internet de las cosas (IoT), que hace que cada vez haya más sistemas interconectados para diversos propósitos, estimándose que, para el 2030 haya más de 80000 millones de dispositivos conectados a Internet (Holst, 2019). De hecho, el crecimiento exponencial del IoT está provocando que los *ataques de denegación de servicio distribuido* sean cada vez más destructivos. Muchos dispositivos se fabrican rápidamente, de forma barata, no considerando su robustez ante ciberataques, lo que los hace vulnerables (NIST, 2021). En el 2016, la red de bots Mirai, compuesta por una serie de dispositivos conectados a Internet, aprovechó esta debilidad para atacar la compañía Dyn, que controla gran parte del sistema de nombres de dominios de Internet, provocando que importantes páginas web, como las de Twitter, Netflix, Spotify o el New York Times, dejaran de funcionar (Perlroth, 2016).

Mencionemos también que los impactos de los ciberataques no son siempre económicos, sino que, además, pueden directamente causar muertes, como ocurrió en 2020 en Düsseldorf (Eddy & Perlroth, 2020): los ciberdelincuentes atacaron su hospital universitario provocando que no se pudiera tratar a una paciente en estado grave, y tuviera que ser redirigida a un hospital a más de 30 kilómetros de distancia, provocando su muerte al no poder ser tratada a tiempo.

Por último, es importante destacar que, regularmente, surgen nuevos tipos de ataques. Por ejemplo, el aumento del valor de las criptomonedas ha provocado la aparición del *criptojacking*, un tipo de ataques que se apodera de ordenadores para minar bitcoins de manera secreta.

Los actores de las ciberamenazas.

Además del número creciente de individuos que realizan ciberataques, existe también una mayor diversidad de perfiles entre los actores que los llevan a cabo. En este crisol, podemos distinguir los denominados *hacktivistas*, estrechamente vinculados a movimientos políticos o sociales que pueden desde actuar en defensa de la libertad de expresión hasta estar estrechamente alineados con organizaciones terroristas. Otro grupo de actores que podemos diferenciar son los *insiders*, que constituyen la amenaza cibernética que provoca mayor número de incidentes (Cardenas et al., 2009) aunque, en contrapartida, éstos pueden ser los más fáciles de controlar mediante un programa adecuado de ciberseguridad.

En este contexto heterogéneo, cabe destacar que los ciberdelincuentes están mejorando también sus aptitudes y capacidades. Numerosos grupos de cibercriminales se han convertido en organizaciones profesionales maduras, algunas de las cuales emplean docenas de hackers y poseen grandes recursos financieros. Estas organizaciones vienen incentivadas por los mercados de la web oscura que demanda individuos altamente cualificados para robar datos o desarrollar herramientas automatizadas de ataque. De hecho, son estas herramientas las que han facilitado que personas sin conocimientos técnicos avanzados puedan realizar cibercrímenes.

Sin embargo, tal vez las amenazas más formidables en la actualidad sean las provenientes de los estados. Aunque, teóricamente, están limitados por las posibles repercusiones militares, económicas y políticas asociadas a lanzar ciberataques, los actores estatales están desarrollando cada vez más programas ofensivos y almacenando ciberarmas que podrían ser liberadas accidental o intencionadamente. De hecho, ha habido ya numerosos ataques perpetrados por estados. Se cree, por ejemplo, que el ataque Stuxnet contra el sistema nuclear iraní fue implementado por Estados Unidos e Israel (Nakashima, 2012). En el año 2007 se produjo uno de los primeros cibearataques rusos, con el objetivo de paralizar Estonia (McGuinness, 2017) como reacción política. En este contexto, también destaca un ataque en el 2015, posiblemente de origen ruso, dirigido contra la red eléctrica ucraniana que dejó a más de 225000 personas sin electricidad durante seis horas (Polityuk et al., 2017); esta es la primera ofensiva cibernética conocida contra una red eléctrica e ilustra el potencial de los ataques contra sistemas ciberfísicos con consecuencias en el mundo real. Muchos expertos también atribuyen el arma cibernética NotPetya a Rusia; aunque pretendía ser un rasomware, e inicialmente estaba dirigido contra Ucrania, alcanzó inadvertidamente una serie de objetivos no relacionados, afectando a miles de empresas, entre ellas Maersk, DHL, y Saint-Gobain, causando unos daños estimados en 10.000 millones de dólares (Greenberg, 2018).

3. MARCOS CONCEPTUALES ACTUALES

Para afrontar los desafíos que plantean las ciberamenazas, es esencial el uso de una sólida metodología de gestión de riesgos de ciberseguridad (CSRM). Las técnicas de CSRM

se basan en gran medida en el análisis de riesgos (Cooke & Bedford, 2001), que permite a las organizaciones evaluar los riesgos sobre sus activos, así como determinar las medidas que deben aplicarse para reducir la probabilidad de que se produzcan las amenazas y su impacto, en caso de que tengan lugar. Se han desarrollado numerosos marcos para apoyar la gestión de riesgos de ciberseguridad, entre los que se encuentran la norma internacional ISO 27005 (International Organization for Standardization, 2011); CRAMM (CCTA, 20003); (Amutio et al., 2012); EBIO (Agence Nationale de la Securité des Systèmes d'Informatio, 2010) o el marco de gestión de riesgos del NIST en Estados Unidos [NIST12]. Del mismo modo, una serie de marcos de evaluación del cumplimiento y control ofrecen orientación sobre la aplicación de las mejores prácticas de ciberseguridad. Entre ellos, destacan ISO 27001 (International Organization for Standardization, 2013), y la Cloud Controls Matrix (Cloud Security Alliance, 2016). Las metodologías y los marcos mencionados proporcionan un amplio catálogo de amenazas, activos y controles, así como directrices detalladas para la aplicación de medidas destinadas a proteger los activos digitales.

Un escenario genérico con todos los factores que intervienen en la CSRM puede describirse como sigue: una *organización* se enfrenta a posibles amenazas cibernéticas que pueden tener un impacto significativo sobre la misma; para enfrentarse a ellas, la organización debe elegir una *cartera de ciberseguridad*, que consiste en una combinación de productos (agrupados en controles de seguridad y de recuperación) y productos de ciberseguro que le permita gestionar su riesgo de ciberseguridad de la mejor manera posible. La Figura 1 profundiza sobre el escenario genérico expuesto, a partir de un esquema dividido en cuatro elementos principales: organización, amenazas, impactos y cartera de ciberseguridad. Más concretamente, muestra que:

- La *organización* se describe en términos de su *perfil* y sus *activos*, junto con lo que denominaremos *otros rasgos de la organización*.

- Entre las *amenazas*, se puede discernir entre *ciberamenazas no dirigidas*; *amenazas accidentales*, y *amenazas ambientales*. También se pueden producir *ciberamenzazas dirigidas* en las que los actores de la amenaza, o atacantes, dirigen sus ataques específicamente contra una organización de interés.

- Los *impactos* de un ataque incluyen *impactos asegurables*, parcialmente cubiertos por productos de seguro, y los *no asegurables*, que no están cubiertos por un seguro.

- La *cartera de ciberseguridad* incluye los *controles de seguridad*, que se establecen por la organización e incluyen medidas para prevenir, proteger y contrarrestar los ciberataques, incluyendo la detección y respuesta frente a amenazas, lo que ayuda a reducir la probabilidad de que se produzcan éstas y a mitigar su impacto. Podemos dividir aún más los controles de seguridad entre *controles de procedimiento*, *técnicos*, y *físicos*. Además, la cartera de ciberseguridad también puede incluir controles de recuperación, normalmente implementados para responder y recuperarse frente a los ciberataques, reduciendo su impacto. También incluye *contratos de seguros* que permiten transferir el riesgo, ayudando a reducir las eventuales consecuencias financieras de un ataque.

Sin embargo, las metodologías y marcos de CSRM que hemos descrito tienen mucho margen de mejora desde el punto de vista metodológico, en lo que se refiere a análisis y gestión de riesgos. En efecto, en general hacen uso de matrices de riesgos, no carentes de conocidas limitaciones (Cox, 2008). Uno de estos defectos, en comparación con métodos más rigurosos, se debe a que las calificaciones ordinales de probabilidad y gravedad empleadas en tales matrices son propensas a promover la ambigüedad y la interpretación subjetiva de los riesgos. Además, asignan sistemáticamente la misma calificación a amenazas significativamente diferentes desde el punto de vista cualitativo, lo que puede conducir potencialmente a asignaciones subóptimas de los recursos de ciberseguridad. El problema puede ser aún más importante si tenemos en cuenta la creciente variedad de amenazas de ciberseguridad, así como la creciente complejidad de los controles de seguridad empleados en la gestión de riesgos en este dominio.

Otro inconveniente de los métodos actuales es que no suelen tener en cuenta explícitamente la intencionalidad de ciertas amenazas, con algunas excepciones como la metodología IS1 del Reino Unido (HM Government, 2012). La gran mayoría de empresas de seguridad y organismos del sector destacan la importancia de la defensa frente a adversarios, y no solo frente a amenazas accidentales o ambientales (ISF, 2017). En consecuencia, los enfoques actuales de CSRM pueden llevar a las empresas a priorizar incorrectamente riesgos cibernéticos y las medidas que deben aplicar para defenderse de ellos.

Las mencionadas deficiencias, junto al crecientemente amenazante panorama de las ciberamenzas, pone de manifiesto la necesidad de nuevos enfoques en la gestión de riesgos cibernéticos como los descritos en la Sección 4, basados todos ellos en sólidos modelos matemáticos.

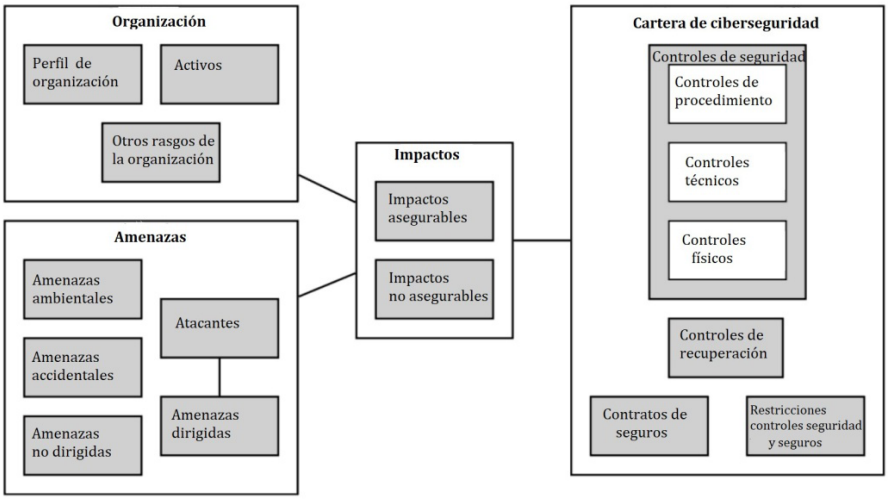

Figura 1. Esquema detallado sobre los factores involucrados en CSRM.

4. MODELOS PARA LA CIBERSEGURIDAD

En la sección 3 hemos esbozado los marcos principales para la gestión de riesgos en ciberseguridad e identificado algunos problemas que les afectan. Como consecuencia, tales marcos pueden conducir a decisiones subóptimas en la gestión de recursos de ciberseguridad. En esta sección presentamos modelos matemáticos que pueden ayudar a una mejor gestión de distintos aspectos de la ciberseguridad.

4.1 Objetivos en la gestión de ciberriesgos

Como indicamos, los ciberataques pueden tener consecuencias muy negativas en términos de costes, pérdida de reputación o, incluso, víctimas. Una primera cuestión importante es determinar los objetivos o medidas de desempeño que debemos optimizar en la gestión de la ciberseguridad.

El enfoque dominante para describir tales objetivos es en términos de la popular tríada CID (Confidencialidad, Integridad y Disponibilidad) de seguridad de la información (Mowbray, 2013). Sin embargo, la creciente variedad de ataques e intereses de los atacantes y la necesidad de integrar mejor la ciberseguridad en términos operativos hace que estos objetivos estén sólo orientados a las TIC y estén ya algo obsoletos: aunque estos son útiles para expresar la seguridad desde una *perspectiva técnica*, no contempla una *perspectiva organizativa* en la que tienen más relevancia los activos y las actividades de la organización.

Además, las normas actuales para describir los impactos de un ciberataque son limitadas. Aunque estas incluyen varias categorías generales de impacto (legal y regulatorio, productiva, reputación financiera y pérdida de clientes), con algunos de ejemplos de subcategorías, estas no cumplen los requisitos del análisis de decisiones (French & Ríos, 2000): la mayoría de ellas ofrecen una lista de impactos empresariales recurrentes o importantes en lugar de un compendio exhaustivo que abarque impactos menos típicos (por ejemplo, los físicos); igualmente, ofrecen objetivos o impactos que se solapan de alguna manera.

A lo anterior hay que añadir la complejidad asociada a entornos hiperconectados, que demanda objetivos que describan de una forma más apropiada el alcance de los impactos de las ciberamenazas. Los avances hasta ahora en este sentido se han centrado en propuestas técnicas adicionales en línea con objetivos relacionados con la *autentificación*, la *autoría* o la *auditabilidad* (Krutz & Vines, 2004), que carecen de una perspectiva empresarial. De hecho, algunos ciberataques recientes ilustran la necesidad de utilizar nuevos objetivos para describir el potencial impacto de estas amenazas. Algunos de los ataques arriba mencionados ejemplifican la importancia de considerar el *impacto sobre terceros* como objetivo a optimizar.

Otros factores que se pueden tener en cuenta como objetivos son los *impactos físicos* y *ambientales*. Aunque los ataques cibernéticos con *impacto físico* son inusuales por el momento, la aparición de sistemas ciberfísicos e infraestructuras inteligentes ponen los riesgos de estos ataques en un primer plano, siendo un ejemplo principal Stuxnet (Nakashima, 2012). De manera similar, en relación con el *impacto ambiental*, (Sayfayn & Madnick, 2017) relatan cómo un pirata informático provocó que 800.000 litros de aguas residuales sin tratar inundaran las vías fluviales de una zona.

Dadas tales limitaciones, esbozamos aquí un conjunto

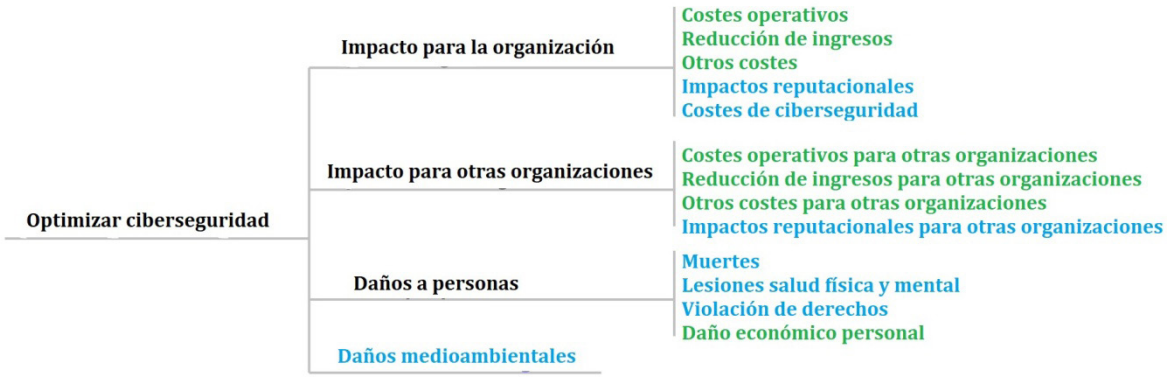

Figura 2. Objetivos de ciberseguridad. Verde, evaluado en términos monetarios; azul, no medible directamente en términos monetarios; negro, con ambos tipos de subobjetivos.

de atributos que representa de manera más amplia los impactos de las amenazas cibernéticas y son más susceptibles de interpretación desde una perspectiva organizativa, además de satisfacer las propiedades típicas de un árbol de objetivos en un análisis de decisiones (French & Ríos Insua, 2000): se han obtenido enumerando los objetivos e impactos en ciberseguridad y ordenándolos jerárquicamente de forma exhaustiva, medible, sin solapamientos, pertinente, inequívoca y comprensible.

Para llevar a cabo el procedimiento anterior, aparte de revisar las listas existentes de impactos de ciberseguridad, se ha recurrido a los campos de la gestión de activos y del derecho. Esto facilita la distinción entre objetivos que pueden evaluarse en términos monetarios (directamente o a través de una estimación) y otros que no son monetarios y, por tanto, necesitan una consideración especial a la hora de su evaluación. También ayuda a distinguir entre los propietarios de los objetivos como, por ejemplo, los daños medioambientales sufridos por terceros además de los monetarios, o las consecuencias reputacionales que los ataques podrían causar a la organización.

El árbol genérico de objetivos de ciberseguridad resumido en la Figura 2, cumple con las propiedades mencionadas anteriormente, es el resultado de una profunda tarea de validación usando el proceso descrito. Todos los objetivos se refieren a minimización. Como ejemplo, los impactos monetarios podrían tratarse mediante valores actuales netos (VAN) (French & Ríos Insua, 2000).

Según el color de los objetivos en la Figura 2, estos pueden expresarse en términos monetarios o no. Estos últimos deben cuantificarse para su uso en los modelos de gestión de riesgos. Para ello, primero identificamos los principales escenarios que podrían provocar los diversos incidentes cibernéticos. Estos escenarios hipotéticos deberían ser exhaustivos en el sentido de cubrir todos los tipos factibles de impactos relacionados con el objetivo que las partes interesadas, los activos y las actividades de la organización pueden sufrir si son atacados. Una vez identificados, se deben cuantificar. Para ello se intenta utilizar un atributo natural, si está disponible y es lo suficientemente sencillo de evaluar; cuando esto no es posible, podemos buscar un atributo construido o un atributo indirecto o proxy (Keeney, 1992). Pueden verse los atributos adoptados y ejemplos de aplicación en (Couce-Vieira et al., 2020).

4.2 Monitorización de redes

Los impactos expuestos en la sección 4.1 ponen de manifiesto la necesidad de desarrollar marcos sólidos de gestión de riesgos de ciberseguridad y el papel central que estos desempeñarán en la seguridad empresarial e industrial en un futuro próximo. Además, la creciente importancia y disponibilidad de datos en todas las actividades relacionadas con la empresa, acentúa la necesidad de tales marcos para mejorar la toma de decisiones en materia de ciberseguridad. Como ejemplo, (Sedgewick, 2014) proporciona un conjunto de normas y directrices que apoyan dichos marcos, cubriendo las actividades clave de ciberseguridad.

Dos de las actividades ahí recogidas se refieren a la supervisión continua de la seguridad de los dispositivos conectados a Internet (DCI) y a la detección de anomalías. Esto ha llevado al desarrollo de herramientas que recogen periódicamente información de alta frecuencia de los DCI de una organización para apoyar su monitorización. De hecho, podemos encontrar organizaciones con varios cientos de miles de estos dispositivos de los que obtenemos decenas de variables cada pocos minutos. Esto plantea

tremendos retos a la hora de procesar esas enormes cantidades de datos, y de realizar las predicciones y tomar las decisiones pertinentes en tiempo real para mitigar, con suficiente antelación, los problemas de seguridad y protección potenciales o reales en una red. En particular, es prácticamente imposible analizar las series temporales de cada dispositivo individual a través de personas, lo que crea la necesidad de un marco automatizable y un sistema que lo implemente. Además, en este contexto, muchos modelos estándar del análisis de series temporales resultan inútiles debido a la variedad de las series que hay que afrontar, bien porque no pueden enfrentarse a una cantidad enorme de datos de alta frecuencia, bien porque no pueden emplearse en modo automatizado. En consecuencia, en (Naveiro et al., 2019b) se introduce un marco para la supervisión de series temporales y la detección de anomalías sobre el que se construye un sistema de supervisión de seguridad en redes a gran escala. Desde un punto de vista funcional, necesitamos que el sistema sea automático, versátil, escalable y preciso. Los trabajos anteriores sobre supervisión de redes como (Brutlag, 2000), (Taylor & Letham, 2018), (Bianco et al., 2001) o (Malhotra et al., 2015) no cubren totalmente tales requisitos.

El marco propuesto en (Naveiro et al., 2019b) representa un avance considerable en la consecución de un sistema completamente automatizado, escalable y versátil para la monitorización de series temporales a gran escala y la detección de anomalías en el ámbito específico de la monitorización de la seguridad de las redes. En el mismo, la identificación de comportamiento aberrante suele basarse en métodos heurísticos desarrollados por los analistas, careciendo de capacidad predictiva. En cambio,

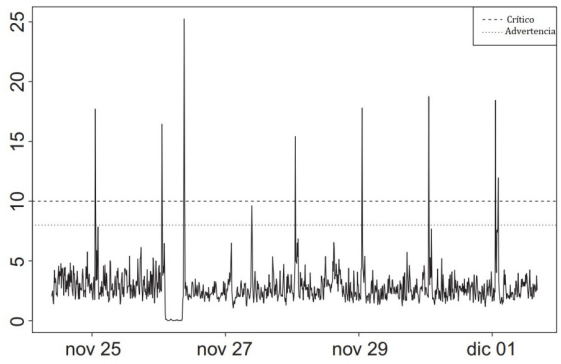

Figura 3. Series temporales continuas con términos lineales y de estallido.

este marco proporciona dicha capacidad de forma automatizada. Para ello, se utiliza un enfoque bayesiano. Para las series continuas, se emplea una versión modificada de los modelos lineales dinámicos (DLM) (West & Harrison, 1999) que incorpora la eventual existencia de estallidos regulares originados por procesos físicos como copias de respaldo, especialmente relevantes en este dominio de aplicación. Una de las principales ventajas de los DLMs es que pueden construirse utilizando diferentes bloques que capturen cada una de las características específicas de las series temporales típicas. También se introduce una forma eficaz de identificación automática de los correspondientes modelos, que resumen los aspectos relevantes de las series en unos pocos parámetros, facilitando su escalabilidad espacial. Por otra parte, el cálculo de las distribuciones predictivas es relativamente rápido, lo que hace que el enfoque sea escalable en el tiempo tras un ajuste adecuado.

En este marco, los siguientes tipos de series temporales continuas son relevantes en la supervisión del tráfico de red de alta frecuencia: *serie con un nivel o tendencia lineal*, posiblemente variable en el tiempo; *serie con un nivel o tendencia lineal junto con uno (o más) términos estacionales*, que suelen describir patrones diarios y/o semanales; *series con un nivel o tendencia lineal*, junto con varios términos de estallido, típicamente asociados con procesos de compresión o de respaldo; *series con un nivel o tendencia lineal junto con uno (o más) bloques estacionales* así como varios términos de estallido. En consecuencia, la expresión general adoptada para estos modelos será

$$Z_n = Y_n + S_n + B_n + \epsilon_n,$$

donde Y_n designa un bloque de tendencia lineal; S_n designa el(los) bloque(s) estacional(es); B_n designa el(los) bloque(s) de estallido; y, finalmente, ϵ_n designa un término de ruido. Obviamente, no será necesario incluir los tres bloques en todos los casos. En el caso de que se detecten algunos procesos de estallido, el modelo DLM se *desactiva* en los momentos correspondientes y el análisis de estos puntos se realiza por separado. Para ello, se emplea un modelo normal para cada proceso de estallido particular.[1]

En el marco descrito, asociados a la serie correspondiente se introducen dos valores de referencia, designados W (*advertencia*) y C (*crítico*), vinculados a niveles de observación tales que, si se superan, deben llevar a emitir señales de advertencia y de criticidad, respectivamente. Por ejemplo, en un disco de almacenamiento podríamos estar monitorizando su uso y establecer $C = 0{,}95$, lo que significa que cuando alcancemos una saturación del 95%, el rendimiento del disco podría degradarse e incluso colapsarse, induciendo una pérdida. Estos umbrales dependerán del dispositivo y de su relevancia global. Obsérvese que este sistema de alarma incluye dos niveles para tratar de proporcionar una información más rica sobre los posibles fallos del dispositivo monitorizado.

Varias actividades clave de ciberseguridad en la gestión de redes están relacionadas con la supervisión de la

1 Para las series discretas, se utilizan cadenas de Markov de tiempo discreto (Ríos Insua et al., 2012) que también cumplen con las propiedades deseadas anteriores.

seguridad y la detección de anomalías dentro de las series temporales, en las que, a su vez, se basan para:

i. Hacer previsiones a corto plazo, permitiendo:

   - Identificar el comportamiento anómalo de las series cuando los valores observados no se encuentran dentro de los intervalos predictivos.

   - Identificar el comportamiento no seguro cuando los intervalos predictivos de los modelos cubren realmente los umbrales *W* y/o *C*.

ii. Hacer previsiones a largo plazo. Permiten prever problemas críticos con suficiente antelación cuando los umbrales *W* y/o *C* aparecen en los intervalos de predicción a largo plazo: los valores críticos que se encuentran dentro de ellos señalan un comportamiento potencialmente problemático en un futuro lejano.

Para llevar a cabo estas tareas, se necesitan procedimientos para realizar predicciones puntuales y por intervalos de las series temporales implicadas. El ancho de los intervalos debe ser adaptable para controlar los falsos positivos, así como para tener en cuenta el valor del activo correspondiente. Siempre que las predicciones alcancen el nivel *W* o el nivel *C*, debe emitirse una alarma, siendo más prioritario este último caso.

El marco descrito está implementado y funciona como parte de un sistema que controla una red con más de trescientos mil dispositivos, proporcionando cada uno de ellos varias series temporales de muy alta frecuencia.

4.3 Gestión de ciberriesgos en la cadena de suministros

Los terremotos, las crisis económicas, las huelgas, los atentados terroristas y otros acontecimientos pueden interrumpir las operaciones de la cadena de suministro con un impacto significativo sobre el rendimiento de las organizaciones. Por ejemplo, Ericsson perdió 400 millones de euros después de que la planta de semiconductores de su proveedor se incendiara en 2000 (Latour, 2001). La actual situación post-pandemia de falta de suministros se asocia también a una rotura masiva de la cadena de suministros.

La *Gestión de CiberRiesgos en la Cadena de Suministros* (GCRCS) (Torres et al., 2020) ha surgido como disciplina para aplicar estrategias de gestión de riesgos en una cadena de suministro con el objetivo de reducir sus vulnerabilidades y evitar las interrupciones en la disponibilidad de servicios y productos. Este campo ha adquirido recientemente interés debido a la proliferación de ciberataques y la creciente interconexión de las organizaciones, como muestra el ejemplo de Target. En este caso los hackers accedieron al sistema informático de la empresa a través de un proveedor de aire acondicionado, robando información personal de 70 millones de clientes (Markowsky & Markowsky, 2014).

Por su importancia, esbozamos un problema estándar de GCRCS. Consideremos una empresa interconectada con sus proveedores. Tanto la empresa como los proveedores están sujetos a varios tipos de ataques. Algunos ejemplos de estos ataques sería a través de botnets o de información de acceso robada. Los ataques a los proveedores podrían trasladarse a la empresa. Por ejemplo, imaginemos un caso en el que uno de los proveedores se infecta a través de un malware; el atacante podría entonces escanear la red del proveedor y enviar correos electrónicos infectados a la empresa, que tendría más probabilidades de infectarse porque el software recibido procede de una fuente aparentemente legítima.

Nos basamos en un sistema de inteligencia de amenazas (TIS) (Tittel, 2017) que recoge datos sobre los vectores de ciberataques recibidos por la empresa y sus proveedores. Los vectores podrían incluir, por ejemplo, información como las IP de los dispositivos infectados por redes de bots o el número de infecciones de malware encontradas. El TIS también puede proporcionar datos sobre el entorno de seguridad incluyendo, por ejemplo, el número de menciones negativas en blogs de hacktivistas, y la postura de seguridad, cubriendo la cadencia de parcheo o el número de vulnerabilidades. Toda la información anterior estaría disponible, tanto para la empresa como para sus proveedores. A partir de estos datos, y de otra información disponible, se pueden evaluar:

- las probabilidades de que los proveedores de la empresa sean atacados;

- la probabilidad de que la empresa sea atacada, bien directamente, bien a través de sus proveedores;

- los impactos que dichos ataques podrían inducir sobre la empresa,

y agregar dicha información para facilitar la evaluación del ciberriesgo de la empresa en relación con sus proveedores para apoyar las decisiones de gestión de tal riesgo.

En esta situación, las empresas son reticentes a proporcionar datos sobre ataques *suficientemente dañinos* por razones obvias de reputación. En este contexto, tal término indica que el ataque es relevante y ha causado un daño significativo a la empresa o sus proveedores. En este contexto, para paliar la escasez de datos, y así poder

implementar los modelos requeridos, se podría extraer información de especialistas en ciberseguridad mediante técnicas estructuradas de juicio de expertos (Cooke, 1991), (O'Hagan et al., 2006).

El enfoque anterior se utiliza periódicamente basándonos en la recopilación de datos a través del TIS y la agregación de resultados para evaluar los ciberriesgos en la cadena de suministro. Esto permite emplear modelos de predicción sobre los índices de riesgos, lo que tiene varias aplicaciones como emitir alarmas, clasificar a los proveedores o facilitar la negociación de acuerdos de nivel de servicio entre la empresa y sus proveedores (Redondo et al., 2019).

4.4 Un marco general para la gestión de riesgos en ciberseguridad

En esta sección, describimos un modelo CSRM alternativo a los empleados en los estándares aludidos en la Sección 3. Su objetivo es apoyar a las organizaciones en sus decisiones de gestión de riesgos de ciberseguridad, específicamente para determinar su asignación óptima de recursos. El modelo descrito se basa en el Análisis de Riesgos Adversarios (ARA) apoyándose en el empleo de Diagramas de Influencia Multi-Agente. Comenzamos pues con una visión de tales elementos.

Análisis de Riesgos Adversarios.

El ARA es un subcampo relativamente reciente del análisis de decisiones que combina conceptos del análisis de riesgos y la teoría de los juegos. Permite al analista de riesgos asignar mejor sus recursos para protegerse frente a adversarios, teniendo en cuenta sus estrategias y sus incertidumbres (Banks et al., 2015). En ARA, se busca maximizar la utilidad esperada de un decisor, que denominamos *defensor*. Para ello, se tienen en cuenta sus utilidades esperadas dadas las probabilidades de que se produzcan las distintas amenazas y sus impactos. Esto incluye la estimación de las probabilidades de que el defensor sea atacado por otros decisores a los que nos referimos como *atacantes*. Para ello, se determinan las utilidades esperadas de éstos a la hora de atacar al defensor mediante simulación, lo que conlleva considerar sus alternativas de decisión; como se dispone información limitada sobre los mismos, también se incorpora la incertidumbre sobre sus estrategias, creencias y preferencias empleando distribuciones de probabilidad subjetivas. Posteriormente, se emplea optimización para determinar la estrategia de decisión del defensor que maximiza su utilidad esperada. Este tipo de problema es de simulación-optimización dado que se emplea simulación, para predecir el comportamiento de los atacantes, y optimización, para determinar la mejor estrategia para el defensor dadas las circunstancias.

El ARA se diferencia del análisis de riesgos estándar, pues supone que los adversarios son inteligentes y piensan estratégicamente; y de la teoría de juegos estándar, debido a que no supone que los decisores dispongan del conocimiento de las estrategias de los adversarios, así como de sus preferencias y creencias. Tampoco busca un equilibrio entre los decisores, ya que el problema se resuelve desde la perspectiva del defensor en lugar de resolver los problemas de todos los decisores conjuntamente.

Diagramas de Influencia Multiagente.

Aportan una representación de los problemas de decisión con más de un agente, en particular en gestión de ciberriesgos, y se consideran una extensión de los diagramas de influencia (French & Ríos Insua, 2000). Empleando la terminología en (Banks et al., 2015), incluimos cuatro tipos de nodos:

- *Nodos de incertidumbre*: corresponden a cada una de las incertidumbres que deben modelizarse. Se representan como un *oval*.
- *Nodos determinísticos*: son un subtipo de nodos de incertidumbre en el que se conocen con seguridad los resultados del nodo. Se representan mediante un *doble oval*.
- *Nodos de decisión*: corresponden a decisiones que el agente tiene que tomar entre un conjunto de alternativas. Se representan mediante un *rectángulo*.
- *Nodos de valor*: modelizan las preferencias del agente, determinadas por su función de utilidad. Se representan mediante un *hexágono*.

Los nodos que son relevantes para las decisiones de un solo agente están en colores sólidos, con cada agente representado por un color diferente. En cambio, los nodos relevantes para las decisiones de varios agentes se representan rayados. Los diagramas también incluyen dos tipos de arcos:

- *Punteados*, dirigidos a un nodo de decisión. Indican que las decisiones correspondientes se toman conociendo los valores de todos los nodos que las preceden.
- Só*lidos*, dirigidos a nodos de incertidumbre o valor. Indican que los valores del nodo correspondiente (probabilidades o utilidades) dependen de los valores de los nodos que las preceden.

A continuación presentamos algunos detalles del modelo

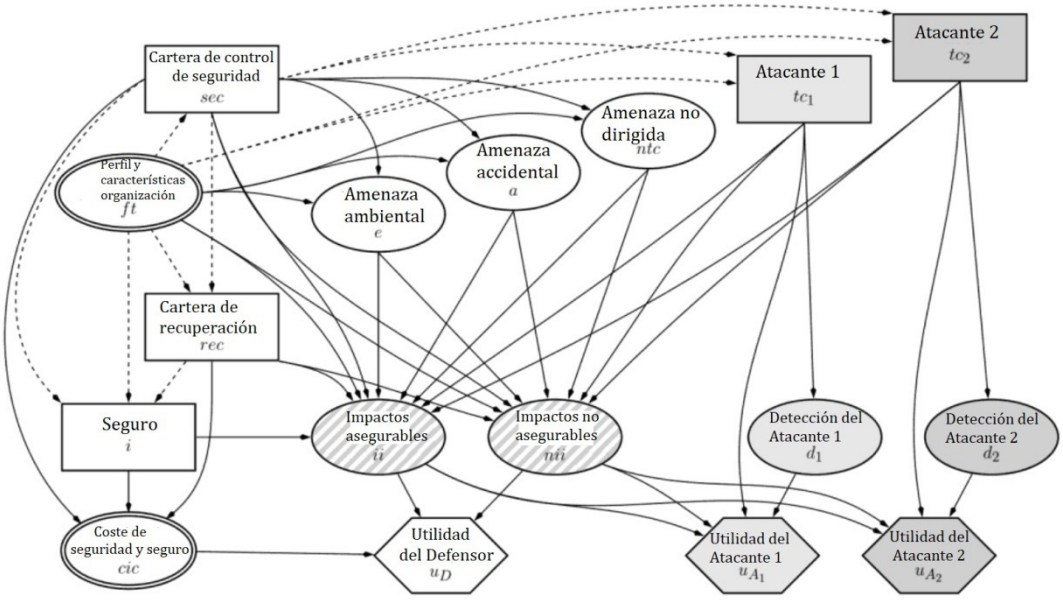

Figura 4. DITA del problema CSRM para una organización.

de CSRM, en el contexto de un responsable de la toma de decisiones de una PYME que debe decidir qué cartera de productos de ciberseguridad implantar. Para ello, el decisor considera varios factores, incluida la importancia de sus activos. Por ejemplo, si la puesta en peligro de un activo determinado de la cadena de valor de la PYME puede afectar su solvencia, el decisor puede considerar que es esencial asegurarse frente a este riesgo. También debe tener en cuenta las obligaciones reglamentarias y otras obligaciones legales de la empresa.

El diagrama de influencia tri-agente de la Figura 4 modeliza este problema. Los tres agentes en este diagrama son la PYME (designada como Defensor) que decide sobre su asignación de recursos de ciberseguridad y dos tipos de atacantes (designados Atacante 1 y Atacante 2). Estos pueden ser desde ciberdelincuentes hasta competidores o estados nacionales, mientras que sus motivos para atacar al Defensor pueden incluir desde obtener un beneficio económico (por ejemplo, el robo de datos de clientes, incluida la información de las tarjetas de crédito) hasta la obtención de una ventaja competitiva (por ejemplo, el robo de información comercial de un competidor). Para distinguir entre los agentes, los nodos que implican al Defensor son de color blanco; los que implican sólo al Atacante 1 son de color gris claro; y, finalmente, los que implican al Atacante 2 son de color gris oscuro.

En la Figura 4, identificamos los elementos relevantes del problema. Designamos mediante *ft* las características de la empresa. También se identifican distintas amenazas que pueden afectar a la empresa, incluyendo ciberamenazas dirigidas, ataques cibernéticos dirigidos intencionadamente a la organización por el Atacante 1 (tc_1) o el Atacante 2 (tc_2). Se modelizan como nodos de decisión, ya que implican decisiones tomadas por los atacantes. Por otro lado, el diagrama también incluye las siguientes amenazas no intencionadas:

- *Ciberamenazas no dirigidas* (*ntc*), ataques cibernéticos que no están directamente dirigidos a la empresa, como una campaña de malware.
- *Amenazas accidentales* (*a*), incidentes cibernéticos causados sin intención maliciosa como, por ejemplo, un empleado que borra accidentalmente archivos.
- *Amenazas medioambientales* (*e*), que implican inundaciones, terremotos y otros desastres que pueden afectar a los activos informáticos de la organización.

Las amenazas no intencionadas se modelizan mediante nodos de incertidumbre. También podemos observar en el diagrama los impactos relevantes para la empresa, que incluyen los impactos asegurables (*ii*), parcialmente cubiertos por el seguro, y los no asegurables (*nii*), no cubiertos por el seguro. Se modelizan mediante nodos de incertidumbre rayados porque son relevantes para las decisiones de varios agentes.

Además, la figura también incluye una serie de herramientas que la empresa puede elegir para mitigar los riesgos. Estos incluyen dos tipos diferentes de productos de seguridad: la *cartera de control de seguridad* (*sec*), y la *cartera de recuperación* (*rec*). También incluyen un *seguro* (*i*), empleado para la transferencia de riesgos. Todos estos instrumentos se modelizan mediante nodos de decisión (rectángulos), ya que implican decisiones tomadas por la organización.

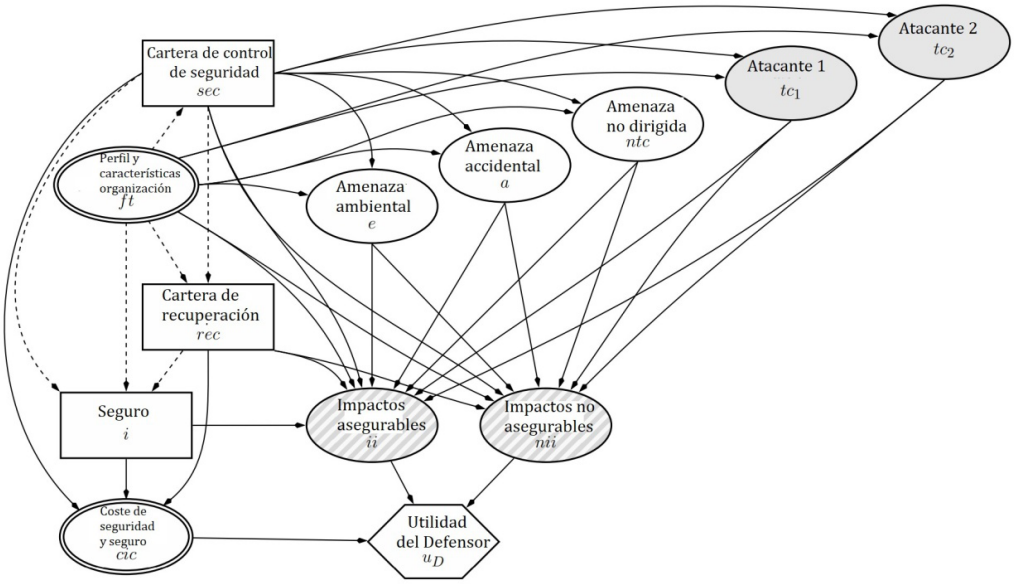

Figura 5. Identificación del problema CSRM para el Defensor.

Además, pueden introducirse restricciones sobre estas carteras de seguridad de varios tipos. También conllevan *costes* (*cic*), representados mediante un nodo determinista.

Una vez identificados todos los elementos relevantes para el Defensor, podemos construir el modelo de preferencias del defensor, que se determina mediante su función de utilidad u_D con un nodo de valor. Esta función cubre también los aspectos de aversión al riesgo del decisor.

El diagrama incluye además el resto de elementos que son relevantes para los atacantes. Tomando como ejemplo al Atacante 1, un factor relevante es la *detección del Atacante 1* (d_1), que simboliza si la empresa es capaz de identificar al mismo como responsable del ataque. Esto se modeliza mediante un nodo de incertidumbre. También cabe destacar la posible presencia de atacantes sofisticados, que emplean la información obtenida invirtiendo tiempo y esfuerzo durante una fase de reconocimiento. Esta información puede incluir el perfil de la empresa y sus características, así como la cartera de controles de seguridad. Estos atacantes emplean esta información para perpetuar ataques dirigidos. Cuando también se han identificado todos los elementos relevantes para los atacantes, podemos construir análogamente el modelo de preferencia de un atacante, modelizando la utilidad del Atacante 1 (u_{A1}) como un nodo de valor. El proceso es análogo para el Atacante 2.

Para determinar la asignación óptima de recursos de ciberseguridad de la organización, seguimos estos pasos:

1. Se examina el problema CSRM desde la perspectiva del defensor, incluyéndo la creación del diagrama de influencia del defensor. Se construye el modelo de preferencias del defensor respecto a su asignación óptima de recursos, teniendo en cuenta las alternativas de decisión del defensor entre diferentes tipos de productos de seguridad y opciones de seguro, así como las probabilidades condicionadas de que se produzcan diversas amenazas o impactos.

2. El problema CSRM se observa desde la perspectiva de los atacantes, teniendo en cuenta el pensamiento estratégico de los mismos. Además, se crean los diagramas de influencia de los atacantes así como sus modelos de preferencias respecto a la realización de determinados ataques dirigidos. Además, se tiene en consideración sus incertidumbres respecto a factores como si serán detectados y los impactos del ataque, así como sus alternativas. De esta forma, determinamos las probabilidades condicionales de las amenazas de los atacantes.

3. Por último, se utiliza optimización para maximizar la utilidad esperada del defensor teniendo en cuenta el pensamiento estratégico de los atacantes. Tenemos en cuenta la probabilidad del defensor de ser atacado por uno o ambos atacantes, así como cada cartera de ciberseguridad factible. Así, se obtiene la cartera de ciberseguridad que proporciona al defensor la máxima utilidad esperada, revelando su elección óptima de una cartera de ciberseguridad para defenderse de los atacantes.

No daremos los detalles matemáticos de los problemas anteriores, que pueden verse en (Ríos Insua et al., 2020),

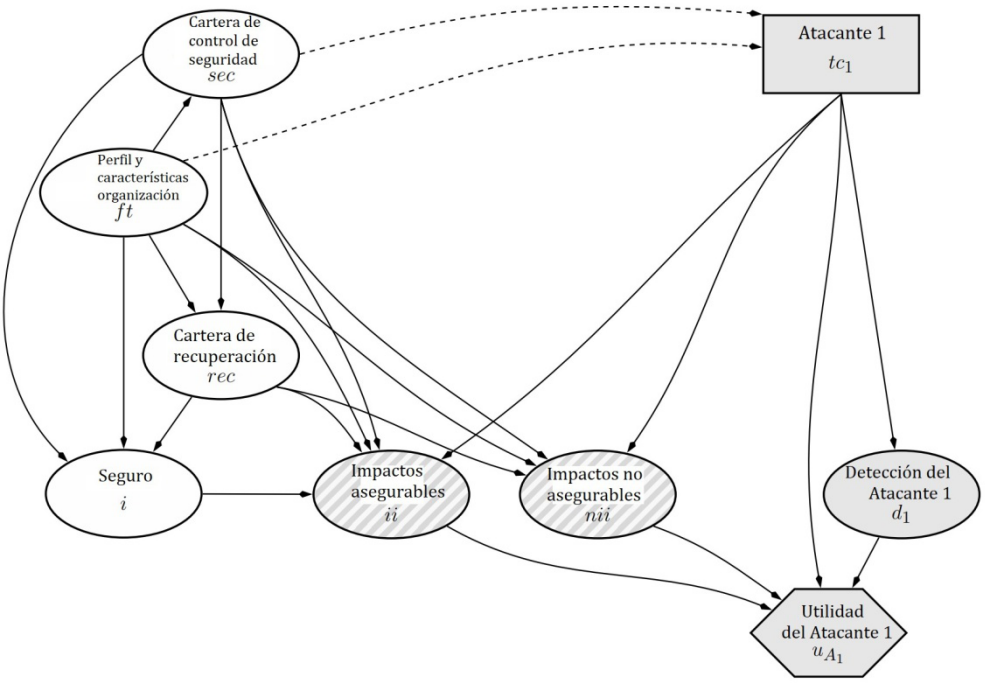

Figura 6. DI del problema CSRM para el Atacante 1.

pero sí representaremos el problema de decisión del Defensor y de uno de los atacantes según se deduce de la Figura 4. La perspectiva del Defensor sobre este problema se representa mediante el diagrama de influencia de la Figura 5.

Para evaluar la probabilidad del Defensor de ser atacado por el Atacante 1, el Atacante 2, o ambos, se deben considerar los componentes del atacante del problema CSRM presentado en la Figura 5. Utilizando al Atacante 1 como ejemplo, presentamos un diagrama de influencia para el Atacante 1 en la Figura 6. El diagrama se basa en la creencia de que el Atacante tiene medios para detectar la cartera de control de seguridad y el perfil y las características de la organización, lo que para los atacantes sofisticados que realizan ataques dirigidos es una hipótesis razonable. El diagrama de influencia del Atacante 2 es análogo.

4.5 La nueva frontera: ciberriesgos e IA

Las máquinas son cada vez más inteligentes. Como ejemplo, los coches pueden ayudarnos ya a planificar nuestro itinerario, aparcar, activar los frenos cuando sea necesario o detectar árboles, aceras y vehículos circundantes. En un futuro no muy lejano, nos transportarán de forma rutinaria de casa al trabajo sin conductor. Para estas actividades, recogen datos, los transmiten a otros vehículos o una infraestructura interconectada y aprenden online, apoyados en métodos de la IA. Entonces, ¿las máquinas inteligentes conectadas harán nuestro mundo más seguro?

Las empresas de ciberseguridad y los gobiernos emplean ya la IA para detectar vulnerabilidades desconocidas en sus sistemas de información y tratarlas antes de que los

atacantes las exploten. Por ejemplo, en la Sección 4.2 hemos descrito esquemas automáticos para comprobar el estado de cientos de miles de dispositivos conectados y enviar señales de alerta a los ingenieros cuando un dispositivo se comporta de forma anormal. Además, se expuso cómo los modelos predictivos pueden prever fallos inminentes, obteniendo un tiempo precioso para reaccionar con antelación gracias a la IA.

Igualmente, algunos sistemas de inteligencia de amenazas realizan análisis en profundidad del entorno y la postura de seguridad dentro de una organización, como esbozamos en la Sección 4.3. Conllevan un verdadero diluvio de datos que debe agregarse de forma coherente para proporcionar indicadores de riesgo significativos y útiles que, típicamente, combinan modelos de aprendizaje automático y económicos para realizar dicha agregación. Algunos de esos sistemas también pueden capturar y analizar contenidos de la web y de las redes sociales sobre una empresa.[2] Para ello, nos apoyamos en herramientas de IA, como el análisis de sentimientos, para determinar la eventual negatividad de los comentarios.

En todos estos casos, las predicciones de los modelos

2 Por ejemplo, los tuits negativos pueden aumentar el riesgo de ataque: una mención cuidadosamente diseñada puede irritar a los hacktivistas para que lancen dichos ataques.

de IA apoyan la toma de decisiones de ciberseguridad en presencia de adversarios. Los nuevos enfoques, como los basados en el ARA, facilitan las decisiones en línea y mejoran la precisión y la rapidez en la gestión de ciberriesgos, como describimos en la Sección 4.4. Sin embargo, la IA no es inmune a los ciberataques. Para garantizar que las aplicaciones basadas en IA sean seguras, y la lista es interminable (conducción automática, filtros de contenido, sistemas de vigilancia, ...), los algoritmos de aprendizaje automático deben ser robustos y fiables. En realidad, tales algoritmos funcionan extraordinariamente bien con datos estándar; sin embargo, son vulnerables frente a los denominados *ejemplos adversarios* (Naveiro et al., 2019a), que son datos hábilmente modificados por atacantes para engañar a los algoritmos y obtener algún beneficio. Por ejemplo, un estafador puede modificar una reclamación de seguro ilegítima para que sea aceptada por el correspondiente algoritmo y conseguir una indemnización. Además, los atacantes se adaptan rápidamente a las defensas de los sistemas de aprendizaje automático, como se ha señalado en ámbitos como la detección de spam, la detección de fraude y la visión por ordenador. Sus implicaciones en ámbitos como los sistemas de conducción automatizada, los sistemas de defensa, la aplicación de la ley y la sanidad, por nombrar sólo algunos, podrían ser dramáticas.

Tales problemas de seguridad cuestionan el empleo de algoritmos estándar de aprendizaje automático, dada la presencia de adversarios adaptativos dispuestos a intervenir en el problema para modificar los datos con el fin de obtener un beneficio. Para evitar los ataques de adversarios, está surgiendo un nuevo campo denominado *aprendizaje automático adversario* (AAA). Se ha centrado en el uso de la teoría de juegos para modelizar el enfrentamiento entre los sistemas de aprendizaje y sus adversarios. Sin embargo, al hacerlo se adoptan hipótesis de conocimiento común entre defensores y atacantes, lo cual es cuestionable en el ámbito de la seguridad, ya que los adversarios intentan esconder y ocultar información. Como alternativa, es posible adoptar una aproximación al AAA basada en el ARA que facilita la predicción del comportamiento adversario, sin adoptar hipótesis fuertes de conocimiento común.

La ciberseguridad y la IA van de la mano. Sin embargo, como ocurre con muchas herramientas y metodologías, la IA es un arma de doble filo: utilizamos las modernas herramientas del aprendizaje automático para diseñar sistemas más ciberseguros, pero necesitamos diseñarlos de forma que no se vean afectados por los ataques. Necesitamos pues una ciberseguridad más inteligente.

5. CONCLUSIONES

Una de las mayores amenazas globales se refiere a los ciberataques, que no hacen sino crecer en número, sofisticación e impacto. Requieren, por tanto, especial atención por parte de la sociedad en su conjunto para que, entre todos, mejoremos el tratamiento que se hace frente a las amenazas de ciberseguridad.

De forma interesante, existen ya marcos generales para la gestión de la ciberseguridad y la implantación de buenas prácticas. Sin embargo, aunque hacen un excelente trabajo de estructuración del problema de gestión de la ciberseguridad desde el punto de vista de la ingeniería, tales marcos adolecen de algunas carencias desde el punto de vista del análisis de riesgos.

Por ello, hemos presentado algunas contribuciones que, desde la modelización matemática, pueden ayudar a mejorar el panorama descrito. La presentación ha sido de tipo divulgativo y pueden verse aspectos técnicos adicionales en las referencias mencionadas. En cualquier caso, resultan meras propuestas iniciales que esfuerzos adicionales de investigación y desarrollo pueden contribuir a facilitar el alcanzar un mundo más ciberseguro.

AGRADECIMIENTOS

Trabajo apoyado por un proyecto de la Fundación BBVA, el proyecto MTM2017-86875-C3-1-R AEI/ FEDER EU y la Cátedra AXA-ICMAT en Análisis de Riesgos Adversarios. El trabajo de J.M. Camacho viene apoyado por una beca INPhINIT de la Fundación "La Caixa" (ID 100010434), cuyo código es LCF/BQ/DI21/11860063.

REFERENCIAS

Agence Nationale de la Sécurité des Systèmes d'Information (France) (2010). *Expression des Besoins et Identification des Objectifs de Sécurité*.

Amutio, M., Candau, J., & Mañas, J. (2012). MAGERIT–Versión 3.0. *Metodología De Análisis y Gestión De Riesgos De Los Sistemas De Información. Libro I-Método*, 22-45.

Banks, D., Ríos, J., & Ríos Insua, D. (2015). *Adversarial Risk Analysis*. Francis and Taylor.

Beek, C., Cashman, M., Fokker, J., Gaffney, M., Grobman, S., Hux, T., Minihane, N., Munson, Munson, L., Palm, C., Polzer, T., Roccia, T. Samani, R., & Schmugar, C. (2021). McAfee Labs Threats Report. McAfee.

Berr, J. (2016). "WannaCry" ransomware attack losses could reach $4 billion. *CBS News*.

Bianco, A. M., Garcia Ben, M., Martinez, E., & Yohai, V. J. (2001). Outlier Detection in Regression Models with ARIMA Errors using Robust Estimates. *Journal of Forecasting,*

20(8):565-579. doi.org/10.1002/for.768.

Brutlag, J. D. (2000). Aberrant Behavior Detection in Time Series for Network Monitoring. In *Large Installation System Administration.*, pages 139-146.

Cardenas, A., Amin, S., Sinopoli, B., Giani, A., Perrig, A., & Sastry, S. (2009). Challenges for Securing Cyber Physical Systems. In *Workshop on Future Directions in Cyber-Physical Systems Security*.

CCTA (2003). CCTA Risk Analysis and Management Method (CRAMM). Central Computer and Telecommunications Agency (CCTA).

Cloud Security Alliance(2016). Cloud Control Matrix.

Cooke, R. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press.

Cooke, R. M. & Bedford, T. (2001). *Probabilistic Risk Analysis: Foundatións and Methods*. Cambridge University Press.

Couce-Vieira, A., Rios Insua, D., & Kosgodagan, A. (2020). Assessing and Forecasting Cybersecurity Impacts. *Decision Analysis,* 17:356-374. doi.org/10.1287/deca.2020.0418.

Cox, L. A. (2008). What's Wrong with Risk Matrices? *Risk Analysis,* 28(2):497-512. Doi.org/10.1111/j.1539-6924-2008.01030.x.

Dawson, S. (2019). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal*.

Eddy, M. & Perlroth, N. (2020). Cyber Attack Suspected in German Woman's Death. *New York Times*.

Federal Trade Commission (2020). *Equifax Data Breach Settlement*.

French, S. & Ríos Insua, D. (2000). *Statistical Decision Theory.* Wiley.

Greenberg, A. (2018). The Untold Story of NotPetya, the Most Devastating Cyberattack in History. *Wired*.

Hatmaker, T. (2018). The damage from Atlanta's huge cyberattack is even worse than the city first thought. *TechCrunch*.

HM Government (2012). *Standard Numbers 1 & 2. Information Risk Management*. Number 4.

Holst, A. (2019). IoT connected devices worldwide 2019-2030. *Statista*.

Instituto Español de Estudios Estratégicos (2012). *Ciberseguridad. Retos y amenazas a la seguridad nacional en el ciberespacio.*

International Organization for Standardization (2011). ISO/IEC 27005 - Information security risk management. International Organization for Standardization (ISO).

International Organization for Standardization (2013). *ISO/IEC 27001 - Information Security Management Systems - Requirements*.

International Organization for Standardization (2015). *Societal Security - Business Continuity Management Systems - Guidelines for Business Impact Analysis*. Technical report, International Organization for Standardization.

ISF (2017). Information Risk Assessment Methodology 2. Information Security Forum (ISF).

Keeney, R. (1992). *Value Focused Thinking*. Harvard U.P.

Krutz, R. & Vines, R. (2004). *The CISP Prep Guide*. Wiley.

Latour, A. (2001). A Fire in Albuquerque Sparks Crisis For European Cell-Phone Giants. *The Wall Street Journal*.

Lewis, J. (2018). *Economic impact of cybercrime, No slowing down*. McAfee.

Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long Short Term Memory Networks for Anomaly Detection in Time Series. In *23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.*, page 89.

Markowsky, G., & Markowsky, L. (2014). From air conditioner to data breach. In *Proceedings of the International Conference on Security and Management (SAM)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

McGuinness, D. (2017). Cómo uno de los primeros ciberataques de origen ruso de la historia transformó a un país. BBC News Mundo.

Ministerio de Defensa (2017). *Estrategia de seguridad nacional. Un proyecto compartido.*

Mowbray, T. J. (2013). *Cybersecurity: Managing Systems, Conducting Testing, and Investigating Intrusions*. Wiley.

Nakashima, E., W. J. (2012). Stuxnet was work of U.S. and Israeli experts, officials say.

Naveiro, R., Redondo, A., Rios Insua, D., & Ruggeri, F. (2019a). Adversarial classification: An adversarial risk analysis approach. *International Journal of Approximate Reasoning*, 113(2):133-148. Doi.org/10.1016/j.ijar.2019.07.003.

Naveiro, R., Rodríguez, S., & Ríos Insua, D. (2019b). Large-scale automated forecasting for network safety and security monitoring. Applied Stochastic Models in Business and *Industry,* 35(3):431-447. Doi.org/10.1002/asmb.2436.

NIST (2012). Special Publication (SP) 800-30, Revision 1. Guide for conducting risk assessments. National Institute of Standards and Technology (NIST).

NIST (2021). Consumer Cybersecurity Labeling for IoT Devices: A Q&A with NIST's Katerina Megas.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow,

T. (2006). Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons.

Perlroth, N. (2016). Hackers Used New Weapons to Disrupt Major Websites Across U.S. New York Times.

Polityuk, P., Vukmanovic, O., & Jewkes, S. (2017). Ukraine's power outage was a cyber attack: Ukrenergo. Reuters.

Redondo, A., Torres-Barrán, A., Ríos Insua, D., & Domingo, J. (2019). Assessing Supply Chain Cyber Risks. arXiv preprint arXiv:1911.11652.

Rios Insua, D., Baylon, C., & Vila, J. (2020). Security Risk Models for Cyber Insurance. Francis Taylor.

Rios Insua, D., Ruggeri, F., & Wiper, M. (2012). Bayesian Analysis of Stochastic Process Models. Wiley.

Sayfayn, N. & Madnick, S. (2017). Cybersafety Analysis of the Maroochy Shire Sewage Spill.

Sedgewick, A. (2014). Framework for Improving Critical Infrastructure Cybersecurity. Technical report, NIST.

Taylor, S. J. & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1):37-45. Doi.org/10.1080/00031305.2017.1380080.

Tidy, J. (2020). Solarwinds: Why the sunburst hack is so serious. *BBC*.

Tittel, E. (2017). Comparing the top threat intelligence services. *Tech Target instead of Search Security*.

Torres, A., Redondo, A., Rios Insua, D., Domingo, J., & Ruggeri, F. (2021) Structured Expert Judgement Issues in a Supply Chain Cyber Risk Management System.

West, M. & Harrison, J. (1999). Bayesian Forecasting & Dynamic Models. Springer.

World Economic Forum (2020). The Global Risks Report.

World Economic Forum (2021). The Global Risks Report 2021.