Rev.R.Acad.Cienc.Exact.Fís.Nat. (Esp)
Vol. 108, Nº. 1-2, pp 127-138, 2015
XVII Programa de Promoción de la Cultura Científica y Tecnológica

Matemáticas para un mundo más seguro: Del análisis de riesgos al análisis de riesgos adversarios

David Ríos Insua[1] y José A. Rubio[2]

[1] Real Academia de Ciencias Exactas, Físicas y Naturales e ICMAT-CSIC
[2] Universidad Rey Juan Carlos

Abstract

Muchos de los principales problemas a los que debe enfrentarse la Humanidad en este siglo se refieren a cuestiones de seguridad, p.ej. frente al cambio climático, el terrorismo, los ciberataques o los accidentes aéreos. Revisamos aquí algunos modelos y metodologías matemáticas que permiten tratar este tipo problemas, haciendo un recorrido que nos lleva del análisis de riesgos (AR) al análisis de riesgos adversarios (ARA). El AR es un proceso analítico sistemático para evaluar, gestionar y comunicar los riesgos que se realiza para reducir o eliminar las consecuencias no deseables de ciertas amenazas. El ARA expande al AR teniendo en cuenta que puede haber adversarios inteligentes dispuestos a incrementar nuestros riesgos. Las ideas se ilustran con ejemplos de seguridad aérea y ciberseguridad.

Palabras clave: Análisis de riesgos, Gestión de riesgos, Análisis de riesgos adversarios, Seguridad, Diagramas de influencia, Ciberseguridad.

1. INTRODUCCIÓN

Desde sus principios, una de las preocupaciones fundamentales de la Humanidad ha sido su seguridad. Inicialmente, la especie humana debía protegerse de sus depredadores y de fenómenos meteorológicos y geológicos extremos. Con nuestro progreso, hemos pasado a esperar seguridad, además, frente a amenazas asociadas al terrorismo, la energía nuclear, los accidentes aéreos, o Internet, entre muchas otras.

Aunque no exenta de polémicas, esta relevancia de la seguridad queda reflejada en la pirámide de motivaciones de Maslow (1943), ocupando el segundo nivel de la misma. Análogamente, el Mapa de Riesgos Globales elaborado anualmente por el World Economic Forum, véase p.ej. WEF (2016), suele identificar entre las principales amenazas algunas referidas a la seguridad (y la ciberseguridad). También se refleja en los principales programas de investigación: por ejemplo, uno de los siete pilares del programa H2020 se refiere a Sociedades Seguras; análogamente, la Estrategia Nacional de Investigación incluye entre sus retos uno referido a Seguridad y Defensa.

Las Matemáticas, como ocurre con otros muchos aspectos de nuestras vidas, pueden contribuir a desarrollar sociedades más seguras, prósperas y justas. Algunos aspectos matemáticos relevantes en seguridad se refieren a la implantación y desarrollo de los seguros, la criptografía, el procesamiento de señales adversarias, la biometría de seguridad o la detección de intrusos en redes. Aquí nos centraremos en cómo las Matemáticas pueden ayudar a tomar mejores decisiones frente a amenazas de seguridad. En concreto, haremos una introducción al análisis de riesgos (AR), que facilita la toma de decisiones para reducir la verosimilitud y/o mitigar el impacto de amenazas no adversarias, y al análisis de risegos adversarios (ARA), que tiene un rol similar frente a amenazas adversarias.

El AR se describe como un proceso analítico para evaluar, gestionar y comunicar los riesgos que pueden afectar a la vida humana, la salud, nuestras propiedades o el medio ambiente, véase Bedford y Cooke (2001) para una revisión. Para ello, adoptaremos la caracteri-

zación clásica de riesgo en Kaplan y Garrick (1981) en términos de posibles escenarios de amenazas, sus consecuencias y sus probabilidades de ocurrencia. Conlleva un proceso para identificar y evaluar las amenazas a las que se expone un sistema que, como consecuencia, permitirá minimizar o evitar la ocurrencia y el impacto de algunas pérdidas. Así, los impactos negativos de las amenazas pueden gestionarse o reducirse al menor nivel posible.

El ARA, véase Banks et at (2015), ha sido propuesto recientemente para tratar amenazas que se originan a partir de acciones intencionadas de adversarios. Motivado por aplicaciones en la lucha contra el terrorismo, la ciberseguridad y la toma de decisiones competitivas, se ha renovado el interés por el desarrollo de herramientas prácticas y teoría para analizar los cálculos estratégicos de oponentes inteligentes que toman decisiones en escenarios con resultados aleatorios. El ARA construye un modelo de análisis de decisiones para uno de los participantes, que denominamos Defensor, que, además, debe construir un modelo de predicción de las acciones de los adversarios, que denominamos Atacantes.

2. CONCEPTOS BÁSICOS DE TOMA DE DECISIONES

Puesto que el énfasis se pondrá en apoyar la toma de decisiones para la gestión de riesgos, comenzamos revisando algunos conceptos básicos de toma de decisiones en condiciones de incertidumbre. Para más información puede verse French y Ríos Insua (2000), entre muchos otros.

La estructura básica de estos problemas incluye un conjunto A de alternativas a entre las que debe elegir el decisor. Las consecuencias de sus decisiones dependen también de factores que éste no controla, que denominamos estados θ, con valores en un conjunto Θ. Así, supuesto que tomamos una decisión a y el estado es θ, recibiremos una consecuencia $c(a, \theta)$, muchas veces de carácter multivariante.

Como ejemplo, en planificación de la seguridad aérea en el ámbito de una agencia nacional, una alternativa sería un plan de seguridad o, equivalentemente, una asignación de recursos de seguridad; un estado se referiría al número de incidentes de distintos tipos y severidades en el periodo de planificación; las consecuencias podrían ser, por ejemplo, el número de víctimas, el de heridos, los retrasos inducidos, el número de aeronaves destruídas de distintos tipos y, finalmente, el impacto de imagen para el país, supuesto que éstos fuesen los criterios seleccionados por el gestor de seguridad aérea.

En orden de esfuerzo y sofisticación crecientes, frente a un problema de decisión podemos optar por seleccionar las alternativas requeridas mediante la intuición, el uso de reglas o el análisis. La intuición se suele apoyar en una serie de heurísticas que demandan escaso esfuerzo computacional y cognitivo, pero conducen a sesgos y paradojas bien conocidas a partir de los trabajos pioneros de Kahnemann y Tversky (1974). Con un nivel algo más sofisticado, se pueden emplear reglas que resumen experiencia, decisiones y consecuencias previas. Sin embargo, estas reglas suelen adolecer de falta de adaptabilidad en situaciones cambiantes y altamente inciertas. Finalmente, en el nivel mayor de sofisticación tenemos los métodos del análisis de decisiones que requieren mucho mayor esfuerzo cognitivo y computacional. Esta clase de métodos es la más relevante en AR, pues debemos enfrentarnos a situaciones cambiantes con alta incertidumbre y consecuencias potencialmente muy negativas.

Tal incertidumbre típicamente proviene de que carecemos de conocimientos suficientes para predecir el estado que se va a producir. Un primer elemento esencial para resolver el problema sería construir un modelo de predicción sobre los futuros estados, que denominamos $p(\theta)$. En el ejemplo de seguridad aérea antes mencionado, necesitaríamos predecir los números de sucesos de distintos tipos para las distintas clases de aeronaves y severidades y, para cada uno de ellos, las correspondientes víctimas, heridos, retrasos, si la nave se destruyó o no, y, finalmente, el impacto de imagen para el país. Después, por agregación tendríamos una predicción de las consecuencias globales asociadas al plan pertinente de seguridad. El modelo se expresará, típicamente, mediante una distribución de probabilidad y en su construcción se utilizarán juicios de expertos y datos, siendo la metodología principal la de predicción bayesiana, véase French y Ríos Insua (2000).

El otro elemento relevante en el análisis de decisiones es la modelización de preferencias del decisor

sobre las consecuencias $c(a,\theta)$. Distintos ejemplos como la Paradoja de San Petersburgo hacen ver cómo el valor monetario esperado no es criterio adecuado para la toma de decisiones en condiciones de riesgo. Esto motiva la introducción del concepto de utilidad, que modeliza no sólo las preferencias del decisor sobre las consecuencias, sino también sus actitudes frente al riesgo. Así, se puede tener aversión al riesgo, cuando se prefiere el valor esperado de una lotería a la propia lotería, de forma consistente en el rango de consecuencias; afición al riesgo, cuando se prefiere la lotería al valor esperado; y, finalmente, ser neutros al riesgo, cuando nos resultan indiferentes la lotería y su valor esperado. Finalmente, integrando los elementos anteriores de probabilidad y utilidad se puede calcular la utilidad esperada asociada a cada alternativa, escogiéndose aquélla de máxima utilidad esperada. Este principio de toma de decisiones será fundamental en el AR.

En la discusión posterior será especialmente relevante el uso de diagramas de influencia, véase Shachter (1986). Esta herramienta permite estructurar problemas de toma de decisiones mediante grafos acíclicos dirigidos con tres tipos de nodos y dos tipos de arcos. Tenemos así nodos de decisión, representados mediante un cuadrado, que modelizan decisiones a tomar; nodos de azar, representados mediante un círculo, que modelizan incertidumbres presentes en el problema de toma de decisiones; y, finalmente, nodos de valor, representados mediante exágonos, que representan la evaluación de consecuencias asociadas al problema de toma de decisiones. Además, tenemos arcos que inciden en nodos de decisión, que indican que la correspondiente decisión se toma conociendo los valores de los nodos antecesores, y arcos que inciden en nodos de azar o de valor, que indican, respectivamente, si las probabilidades y las utilidades dependen de los valores de los nodos antecesores.

Ejemplo: El siguiente diagrama de influencia, Figura 1, ilustra un problema típico de AR. Consideramos una empresa que desea proteger sus instalaciones. Contempla dos amenazas potenciales: que se produzca un incendio o que tenga lugar un robo. Asociamos un nodo de azar a cada una de ellas. Podemos tomar medidas para reducir la probabilidad de que se produzca alguna de las amenazas (por ejemplo, un sistema de alarma hace mucho menos probable un robo) y/o las consecuencias de las mismas, en caso de producirse (por ejemplo, un sistema de detección de incendios, permite apercibirse de la presencia de uno de ellos antes y poner en marcha los procedimientos de extinción más rápidamente, reduciendo en consecuencia su impacto); asociamos un nodo de decisión al mismo. Además de las medidas anteriores, dirigidas a reducir el riesgo, podemos tomar medidas de transferencia de riesgo, referidas a adoptar un seguro, como se refleja en el diagrama de influencia con otro nodo de decisión.

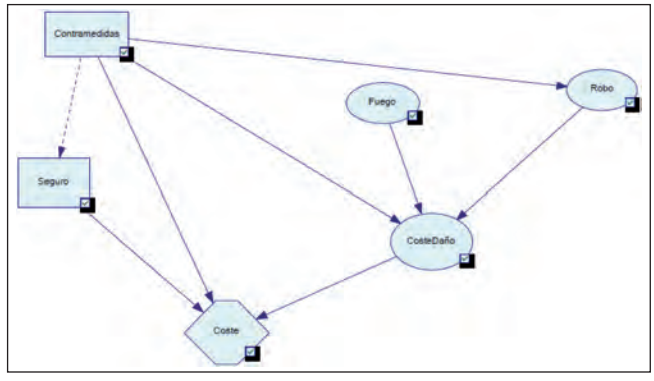

Figura 1. Diagrama de infuencia para un problema de gestión de riesgos de seguridad en una empresa.

3. ANÁLISIS DE RIESGOS

El AR puede describirse como un proceso analítico sistemático para evaluar, gestionar y comunicar el riesgo. Se realiza para entender la naturaleza de las consecuencias negativas, no deseables para la vida humana, la salud, nuestros activos y/o el medio ambiente, y así reducirlas o eliminarlas. Suele describirse como un proceso con tres fases:

- *Evaluación de riesgos,* en la que se obtiene información sobre las características de los riesgos atribuidos a una amenaza, esencialmente su probabilidad de ocurrencia y la distribución de sus consecuencias.

- *Gestión de riesgos,* que comprende las actividades dirigidas a controlar las amenazas, sugiriendo qué decisiones tomar, y, finalmente,

- *Comunicación de riesgos,* referida al intercambio de información y opiniones en relación con el

riesgo y sus factores entre evaluadores de riesgo, gestores y otros participantes.

Los usos habituales del AR incluyen la gestión de riesgos para una instalación existente o propuesta; el desarrollo de regulaciones; la demostración de que se cumplen regulaciones o de la necesidad de mejorar en el cumplimiento de las mismas; su empleo en litigios; o, finalmente, la investigación científica.

Esta disciplina matemática relativamente reciente fue inicialmente predada por la necesidad en el sector de los seguros de hacer predicciones sobre los sucesos asegurados y sus consecuencias. Después, tras los éxitos en la II Guerra Mundial de la Investigación Operativa, se vió fuertemente influenciada por las Ciencias de la Decisión. En los años 70 se produjo un interés creciente por el estudio de la seguridad en sistemas, fundamentalmente en los campos militar, de la ingeniería aero-espacial y de la industria nuclear. Una contribución conceptual principal del inicio de los años 80 se refiere a la gestión de riesgos: una vez identificados y evaluados los riesgos a que estamos expuestos (Kaplan y Garrick, 1981), podemos evitar la ocurrencia de ciertas pérdidas y minimizar el impacto de otras. Así el coste del riesgo puede gestionarse y reducirse a su nivel mínimo. Finalmente, desde principios de este siglo, motivado por el incremento de acciones terroristas a gran escala, se ha comenzado a poner el énfasis en la presencia de adversarios inteligentes, combinando el análisis de riesgos con diversas aportaciones de disciplinas afines a la teoría de juegos, véase Banks et al (2015).

En este punto debemos destacar la metodología de matrices de riesgos (Cox, 2008), muy empleada en numerosos campos, desde la seguridad aérea, pasando por la prevención de riesgos laborales, la auditoría o la ciberseguridad. Consiste en identificar una serie de niveles cualitativos ordinales de verosimilitud (p.ej., *muy infrecuente, infrecuente, normal, frecuente, muy frecuente*) e impacto (p.ej., *muy bajo, bajo, intermedio, alto, muy alto*), dispuestos en una matriz, típicamente de orden bajo (p.ej., 5x5). Después, se asocia a cada amenaza su verosimilitud y su impacto y se despliega en la matriz. Finalmente, se introduce un sistema de colores (habitualmente, verde, amarillo, rojo) para mostrar alarmas en relación con las amenazas en la zona roja y avisos en la zona amarilla. Basados en esta representación gráfica, los decisores pueden discriminar las principales amenazas e identificar recursos a asignar para reducir la situación de riesgo. Sin embargo, aunque se usen con gran profusión, como hemos dicho, han recibido fuertes críticas, véase, por ejemplo, Cox (2008). La Figura 2 ilustra una matriz de riesgos.

Risk		Probability				
		VL	L	M	H	VH
Impact	VH	H	VH	VH	VH	VH
	H	M	H	H	VH	VH
	M	L	M	M	H	H
	L	VL	L	L	M	M
	VL	VL	VL	VL	L	L

Figura 2. Matriz de riesgos.

Describimos ahora un esquema alternativo que formaliza el AR, cuya versión más sencilla se despliega en la Figura 3 a través de diagramas de influencia. En el primero de ellos, el nodo c representa los costes asociados al funcionamiento del sistema bajo circunstancias normales, que se modelizan mediante una densidad $\pi(c)$. El hexágono representa las consecuencias netas en función de la utilidad u del decisor. Evaluamos globalmente los resultados del sistema a través de la utilidad esperada $\psi_n = \int u(c)\pi(c)\mathrm{d}c$.

En la práctica, el dueño del sistema debería hacer una evaluación de riesgos para: (1) identificar los posibles sucesos disruptivos $E_1, E_2, \ldots, E_m$, que supondremos mutuamente excluyentes; (2) evaluar sus probabilidades de ocurrencia, q_j, $j = 1, \ldots, m$; y, finalmente, (3) evaluar los costes (aleatorios) asociados a la ocurrencia del suceso E_j, mediante la densidad $\pi_j(c)$, $j = 1, \ldots, m$. Además, incluímos un suceso E_0 bajo el que no hay disrupciones, con una probabilidad asociada q_0. Entonces el dueño del sistema estimaría la utilidad esperada $\psi_r = \sum_{j=0}^{m} q_j \int u(c)\pi_j(c)\,\mathrm{d}c$, según se ilustra en el segundo diagrama de la Figura 3. La diferencia $\psi_n - \psi_r$, típicamente, será no negativa, puesto que ψ_n describe un problema sin incluir los costes asociados a los sucesos disruptivos, mientras que ψ_r se basa en una evaluación de riesgos.

Si esa diferencia es suficientemente grande, podemos reducirla realizando la gestión de riesgos, introdu-

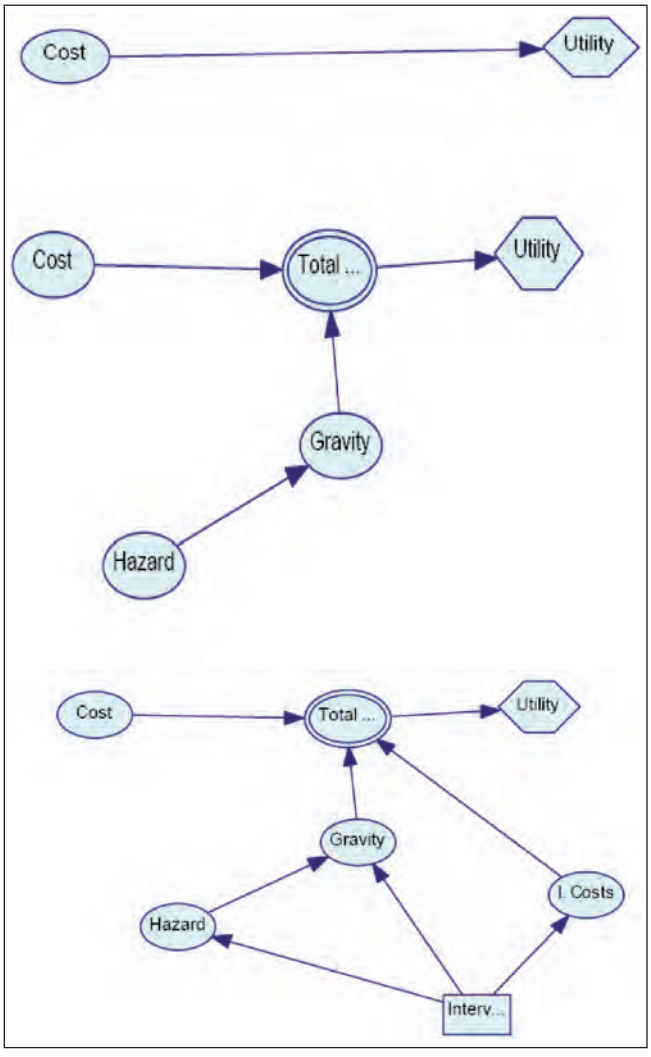

Figura 3. Diagramas de influencia para el análisis de riesgos.

ciendo un conjunto $\mathcal{M}$ de opciones que puede incluir planes de contingencia o seguros, entre muchos otros. Estas pueden reducir los costes asociados con algunos sucesos y/o reducir las probabilidades de disrupción, como se muestra en el tercer bloque de la Figura 3. La solución de gestión de riesgos es la cartera de contramedidas que maximiza la utilidad esperada, esto es, $\psi_m = \max_{m\in\mathcal{M}} \psi_r(m)$, donde

$$\psi_r(m) = \sum_{j=0}^{n} q_j(m) \int u(c)\pi_j(c|m)\,\mathrm{d}c.$$

Puesto que la gestión de riesgos aumenta el conjunto de opciones, será $\psi_m \geq \psi_r$.

4. UN ESQUEMA DE ANÁLISIS DE RIESGOS PARA CIBERSEGURIDAD

En la actualidad, la gran mayoría de organizaciones conectadas se ven impactadas de forma crítica por ciberamenazas, véase Andress y Winterfeld (2013). De igual modo, en el ámbito militar se habla del ciberespacio, un quinto espacio de operaciones en el que son frecuentes movimientos operativos por parte de gran número de países, Leak Source (2014). Para paliar este problema, una herramienta fundamental es el AR, con el que las organizaciones pueden analizar los riesgos que les afectan, priorizar sus activos, ver qué tipo de amenazas y vulnerabilidades pueden estar presentes en los mismos, y qué salvaguardas deben implantar para reducir la materialización de incidencias. Existen ya marcos de análisis de riesgos en ciberseguridad, como, por ejemplo, MAGERIT, y marcos para evaluación, control y cumplimiento como el de la ISO, 27001. Sin embargo, tienden a basarse en matrices de riesgos para la gestión, con los defectos que hemos mencionado en la Sección 3.

Como alternativa describimos un esquema basado en el de la Figura 3, que puede sofisticarse en varias direcciones Por ejemplo, puede haber más de una función de evaluación. Un caso típico en ciberseguridad es considerar como atributos de evaluación la disponibilidad (a), la integridad (i) y la confidencialidad (s). En este caso, $p(a, i, s)$ será la distribución de probabilidad que modeliza la incertidumbre sobre los costes relacionados con los tres criterios antes mencionados. Si $u(a, i, s)$ representase la utilidad multiatributo, la utilidad esperada sería

$$\psi_n = \int\int\int p(a, i, s)\, u(a, i, s)\, da\, di\, ds.$$

Obsérvese que empleamos el modelo $p(a, i, s)$ si se esperan interrelaciones entre tales atributos.

Para considerar el problema de evaluación de riesgos en ciberseguridad, nos centraremos para simplificar la exposición, en el esquema de la Figura 4. Para simplificar se supondrán dos amenazas, una física (p.ej., fuego) y otra representativa de las amenazas tecnológicas (p.ej., sufrir un ataque DDoS). Las denominamos t_1 y t_2 respectivamente. El esquema se extiende de manera inmediata a más de dos amenazas.

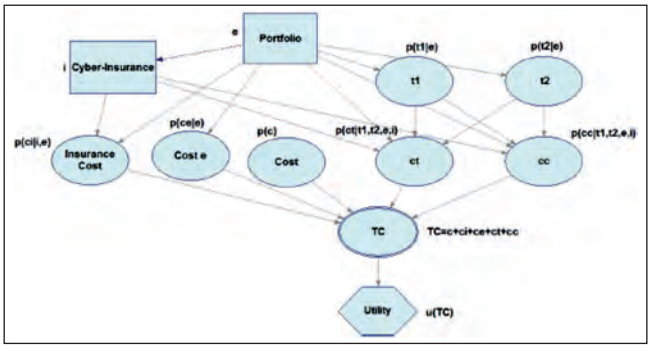

Figura 4. Esquema en análisis de riesgos para ciberseguridad.

Además, consideramos dos tipos de activos, siendo, de igual modo, uno tradicional (por ejemplo, instalaciones) y otro ciber (por ejemplo, equipos informáticos). Los impactos sobre los activos son c_t y c_c y serán, típicamente, inciertos. De nuevo, las ideas se extienden de forma sencilla a un caso con más de dos activos. En caso de que haya alguna relación, bien entre los costes dadas las amenazas, bien entre las amenazas, el correspondiente modelo probabilístico sería de la forma

$$p(c_t,\ c_c\ |t_1,\ t_2)\ p(t_1,\ t_2).$$

Los costes se agregan en el nodo coste total $T\,C$, que incluye los costes bajo condiciones normales, junto con aquéllos asociados a ambos tipos de incidentes. Entonces, la utilidad esperada si se tienen en cuenta las amenazas sería

$$\psi_r = \int \ldots \int u(c + c_t + c_c)\ p(c)\ p(c_t,\ c_c|t_1,\ t_2\ )\ p(t_1,\ t_2)\ dt_1\ dt_2\ dc_t\ dc_c\ dc,$$

supuesto que las consecuencias c, c_c y c_t son aditivas. Más generalmente, tendríamos una función de utilidad $u(c,\ c_c,\ c_t)$. Como antes, cuando $\psi_n - \psi_r$ sea grande, la pérdida en utilidad esperada por tener en cuenta las amenazas es considerable. Así, los incidentes potencialmente perjudican nuestros resultados de manera notable, por lo que debemos intentar gestionar tales riesgos. Introducimos entonces una cartera de contramedidas para reducir la probabilidad de las amenazas y/o reducir su impacto, como se describe en la Figura 4. Algunos ejemplos de contramedidas serían implantar firewalls, dar formación a los empleados, o realizar copias de seguridad.

En el diagrama de influencia de la Figura 4, suponemos, para simplificar, que todas las medidas actúan sobre todos los sucesos e impactos. El nodo e describe la cartera de contramedidas, que tendrán un coste que modelizamos mediante la distribución $p(c_e\ |e)$. Las contramedidas impactan sobre las amenazas $p(t_1|e)$ y $p(t_2\ |e)$, así como sobre los impactos sobre los activos $p(c_t\ |t_1\ ,\ t_2\ ,\ e)$ y $p(c_c|t_1\ ,\ t_2\ ,\ e)$. Todos los costes se agregan en el nodo coste total CT. En este caso, la utilidad esperada cuando se implanta la cartera e es

$$\psi(e) = \int \ldots \int u(c_e+c_t+c+c_c)\ p(c)\ p(c_e|e)p(t_1\ |e)\ p(t_2\ |e)\ p(c_t\ |t_1,\ t_2,\ e)\ p(c_c\ |t_1,\ t_2,\ e)\ dt_1\ dt_2\ dc_t\ dc_c\ dc_e\ dc,$$

donde, de nuevo, suponemos aditividad en los costes. Después buscaríamos la cartera de máxima utilidad esperada resolviendo el problema

$$\psi_e^* = \max_{e\in E} \psi(e),$$

donde E representa el conjunto de carteras de contramedidas factibles: partiendo del conjunto de contramedidas individuales, definiremos carteras que satisfagan distintas restricciones, que pueden ser de tipo económico (por ejemplo, no superar cierto presupuesto), legal (como, por ejemplo, en el cumplimiento de la LOPD, cuando una organización no puede permitirse un daño reputacional por incumplimiento), logístico, político o físico.

Como contramedida de creciente interés, podemos introducir la adopción de un ciberseguro, cuyo coste dependerá, típicamente, de las otras contramedidas implantadas, como se refleja en la Figura 4: cuanto mayores o mejores sean las contramedidas, menor será la prima que, igualmente, dependerá de los activos y arquitectura a proteger. Obsérvese que podríamos incluir el ciberseguro dentro de las contramedidas. Sin embargo, preferimos separarlo, puesto que las primas dependerán de las contramedidas incluidas. El nodo s de decisión describe, pues, la contramedida ciberseguro, derivándose un coste $c_s\ |s,\ e$, habitualmente determinístico. Asimismo, el seguro y las contramedidas inciden en los impactos sobre los activos, que se modelizan mediante $p(c_t\ |t_1\ ,\ t_2\ ,\ e,\ s)$ y $p(c_c\ |t_1,\ t_2,\ e,\ s)$. Todos los costes descritos se agregan en el nodo coste total CT. La utilidad esperada, si se implanta la cartera e y el seguro s, es:

$$\psi(e,\ s) = \int \ldots \int u(c_s + c_e + c + c_t + c_c)\ p(c)\ p(c_e|e)\ p(t_1|e)\ p(t_2|e)\ p(c_t|t_1,\ t_2,\ e,\ s)\ p(c_c|t_1,\ t_2,\ e,\ s)\ p(c_s|s,\ e)\ dt_1\ dt_2\ dc_t\ dc_c\ dc_e\ dc\ d_s$$

Buscamos, entonces, la cartera-seguro de máxima utilidad esperada bajo las restricciones correspondientes, esto es,

$$\max_{e,s}\ \psi(e,\ s)$$

Obsérvese que el problema podría resolverse en dos etapas, aplicando programación dinámica.

5. ANÁLISIS DE RIESGOS ADVERSARIOS

En muchos problemas, algunas de las amenazas se asocian a adversarios inteligentes y proactivos. El procedimiento general descrito en las Secciones 3 y 4 seguirá siendo válido. Sin embargo, predecir lo que los adversarios van a hacer resulta complicado, por el elemento estratégico involucrado. El ARA facilita esta asignación, considerando el problema que resolvería el adversario.

Para explicarlo, emplearemos el modelo de la Sección 4. Suponemos ahora que el parámetro t_2 se refiere a amenazas adversarias, véase Rios Insua et al (2009), Banks et al (2015): se supone que hay un atacante que, de forma intencionada, puede desarrollar tales amenazas, una vez observadas las contramedidas del defensor. En concreto, se corresponde al denominado modelo secuencial de defensa-ataque. Dicho atacante tiene su propio nodo de utilidad $u(a)$, buscando maximizar la eficacia de su ataque, como se describe en la Figura 5 que incluye un diagrama de influencia bi-agente. La amenaza t_1 sigue siendo no intencionada. El ejemplo original, en el que t_1 se refería a fuego y t_2 a un ataque DDoS, seguiría siendo representativo.

El problema al que se enfrenta el defensor está descrito en la Figura 4 y la forma completa de resolución se describió en la Sección 4. Sin embargo, en ella, modelizar $p(t_2|e)$, que describe las probabilidades que la organización otorga a que el atacante implemente el ataque t_2 si se ha introducido la defensa e es difícil por sus características estratégicas. Para facilitar su asignación, pensamos en el problema del atacante, descrito en la Figura 6 mediante un diagrama de influencia.

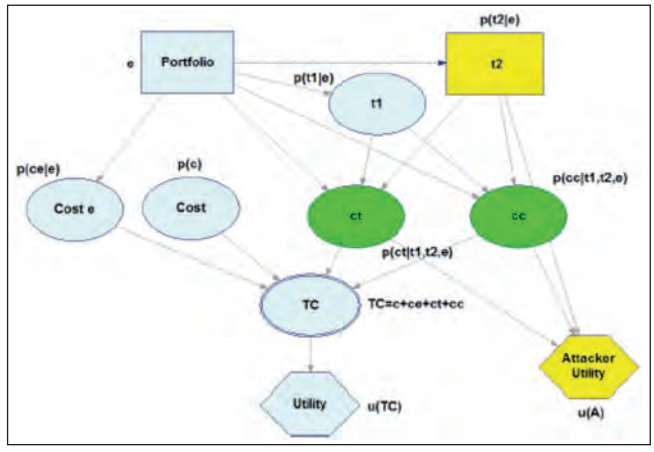

Figura 5. Análisis de riesgos adversarios en ciberseguridad.

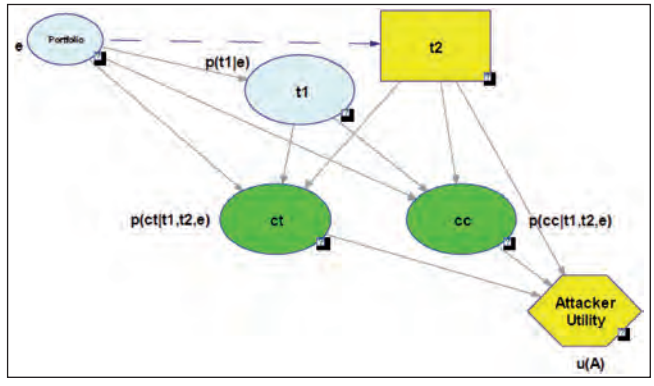

Figura 6. Problema del atacante.

Para cada cartera e, el atacante puede calcular para cada ataque t_2 la correspondiente utilidad esperada

$$\psi_A(t_2|e) = \int\int\int u_A(t_2,\ c_t,\ c_c)\ p_A(t_1|e)\ p_A(c_t|t_1,\ t_2,\ e)\ p_A(c_c|t_1,\ t_2,\ e)\ dt_1 dc_t\ dc_c,$$

suponiendo que intenta maximizar ésta. Después debe encontrar el ataque t_2^* de máxima utilidad esperada, definido mediante

$$\max_{t_2}\ \psi_A(t_2|e),$$

que le proporcionaría su mejor ataque t_2, dada la defensa e.

Sin embargo, puesto que no conocemos u_A y p_A, utilidades y probabilidades del atacante, empleamos utilidades U_A y probabilidades P_A aleatorias que describan nuestra incertidumbre sobre u_A y p_A. Definimos,

después, el ataque óptimo aleatorio, dada la defensa e, mediante

$$T_2^*(e) = \arg\max_{t_2} \int\int\int U_A(t_2, c_t, c_c)\, P_A(t_1|e)\, P_A(c_t|t_1, t_2, e)\, P_A(c_c|t_1, t_2, e)\, dt_1\, dc_t\, dc_c.$$

Se tendría entonces la distribución requerida, que satisface

$$p(t_2|e) = P(T_2^*(e) = t_2),$$

supuesto que T_2 es discreta y, análogamente, si T_2 fuese continua (por ejemplo, cuando se refiere a un esfuerzo de ataque). Tal distribución puede estimarse mediante simulación.

6. UN EJEMPLO DE CIBERSEGURIDAD

Ilustramos ahora el marco para el análisis y gestión de riesgos para ciberseguridad. Mostraremos los principales elementos que intervienen en un caso real simplificado. Nuestro objetivo es dar una idea de la estructura típica de un problema real y cómo se especifica a partir de las plantillas presentadas en las secciones anteriorre. Empezamos estructurando el problema planteado a través de un diagrama de influencia bi-agente que puede observarse en la Figura 7, y describimos a continuación.

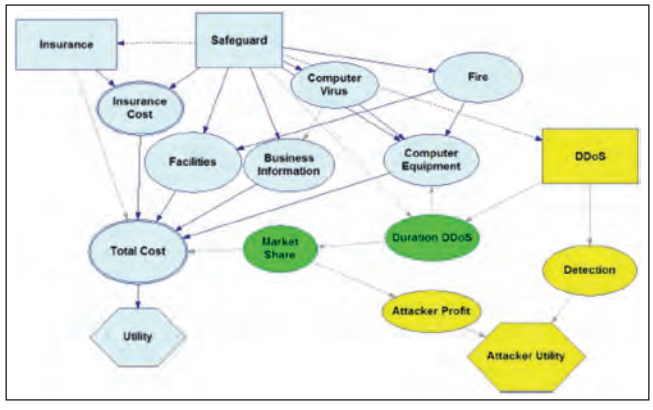

Figura 7. Estructura del problema aplicado.

6.1. Activos relevantes

Para construirlo, en primer lugar, identificamos los activos de la organización. Típicamente, podríamos partir de las listas de alguna de las metodologías estándar, como, por ejemplo, MAGERIT. En este ejemplo consideramos:

- *Instalaciones:* Oficinas de la organización bajo estudio que pueden verse afectadas por las amenazas. Constituyen un ejemplo de activo no informático sin el que la organización no podría operar.
- *Equipos informáticos:* El CPD y los puestos de trabajo de las oficinas resultan un activo esencial para la organización. Si se viese afectado por una amenaza, los costes podrían ser muy cuantiosos, dando lugar a contratiempos como tener que sustituir los equipos afectados o la paralización de actividades.
- *Información de negocio:* Activo esencial para la organización. Se incluyen en el mismo las bases de datos de clientes, la información sobre operativa de proyectos, la propiedad intelectual, los datos de recursos humanos, etc.

Otros posibles activos serían el software de desarrollo propio, los dispositivos móviles o el personal de la organización, pero no los consideraremos aquí, para simplificar la exposición. Asociamos a cada activo un nodo de azar que representa el coste asociado al impacto de las amenazas sobre los mismos.

6.2. Amenazas relevantes

Consideramos ahora las amenazas que pueden producirse sobre los activos identificados. Para ello, empleamos una simplificación de la lista recogida en MAGERIT, aunque podría optarse por listas similares de otras metodologías:

- *Fuego:* Inicio de un fuego en las instalaciones de la organización. Pondrá en peligro dichas instalaciones, junto con los sistemas informáticos. Estos activos se verían degradados esencialmente dependiendo del tiempo de exposición. No se contemplan impactos sobre el activo información de negocio, pues la organización cuenta con un

sistema de respaldo. Consideramos que un fuego puede producirse únicamente de forma accidental: no contemplamos la posibilidad de sabotaje por incendio, debido al sistema de control de accesos en la organización.

- *Virus Informático:* Software malintencionado dirigido a alterar el funcionamiento normal de sistemas informáticos. Puede afectar de diversas maneras a los mismos. Podría, por ejemplo, degradar los datos almacenados, destruyéndolos, o bien reducir el rendimiento de los servicios ofrecidos. Para su propagación hay dos vías principales. En la primera, el usuario acepta de forma inadvertida la instalación del virus; en la segunda, el programa malicioso actúa replicándose a través de redes. Consideraremos la aparición de esta amenaza con carácter accidental, dada su gran capacidad de difusión. Puede afectar a los activos información de negocio y equipos informáticos.

- *Denegación de servicio:* Ataque informático lanzado externamente para socavar la disponibilidad de los sistemas de la organización. Se trata de una amenaza informática intencionada. La parada de un sistema por este tipo de ataque puede hacer que otras partes de la infraestructura TIC se vean comprometidas. Sólo afecta a los activos equipos informáticos y a la información de negocio. Se identifica un único competidor como posible causante del ataque DDoS.

Algunas otras amenazas podrían ser daños por agua, cortes del suministro eléctrico, abuso de privilegios de acceso o una amenaza persistente avanzada, aunque no las consideraremos aquí. Asociamos un nodo de azar a las amenazas fuego y virus, y un nodo de decisión de distinto color, correspondiente a un atacante, a la amenaza denegación de servicio. Como hemos dicho, consideramos sólo un competidor que puede estar interesado en realizar tal ataque.

6.3. Decisiones del defensor

Finalmente, se identifican las salvaguardas relevantes para las amenazas consideradas. De nuevo, puede recurrirse a listados de metodologías como MAGERIT. En nuestro caso planteamos:

- *Salvaguardas:* Consideramos aquí las carteras de contramedidas que puede implantar la organización. Como contramedidas básicas incluimos un sistema anti-incendios; un firewall para protegernos de ataques informáticos externos, como los ataques DDoS; y, finalmente, procedimientos para la adquisición, desarrollo y mantenimiento de sistemas, como estipula la ISO 27001. Así, el nodo inicial de salvaguardas incluiría las siguientes alternativas:

  — Ninguna salvaguarda.

  — Instalar sólo un anti-incendios.

  — Instalar sólo un firewall.

  — Implantar sólo un procedimiento ADM.

  — Restantes carteras que surgen de combinar los elementos anteriores.

Además, consideramos la posibilidad de contratar un seguro, tal vez con cobertura en ciberseguridad.

- *Seguro:* Tendrá un coste dependiente de las restantes contramedidas implantadas en la organización. La prima asociada dependerá de los activos a proteger y otros factores contextuales. Las posibles alternativas son:

  — Ninguno.

  — Seguro tradicional, con cobertura contra incendios.

  — Ciberseguro, con cobertura contra violaciones de datos, pérdidas por amenazas y extorsión, así como limpiezas de virus informáticos, procedimientos de Protección de Datos y fraude informático.

  — Seguro integral, con todas las coberturas anteriores.

Otras posibles contramedidas serían introducir un control de acceso lógico, la protección criptográfica de los datos o la protección del cableado, que no consideraremos aquí. Asociamos un nodo de decisión a las salvaguardas y un segundo nodo de decisión relativo a los seguros a contratar.

6.4. Impactos

Una vez identificados los elementos anteriores, se analizan los impactos derivados de ellos, algunos de los cuales se han mencionado previamente. Sólo consideramos costes. También podrían incluirse impactos menos tangibles, como pudieran ser los relativos a la imagen corporativa. Específicamente, tendremos en cuenta:

1. *Coste de salvaguardas:* Restringidos por el presupuesto con que se cuente.

2. *Costes de impactos sobre los activos:* Los impactos derivados sobre los activos pueden suponer su degradación total o parcial, que implicaría costes muy diversos asociados a factores como incumplimiento de contratos con las consecuentes penalizaciones, sanciones administrativas, etc.

3. *Coste total:* Engloba todos los elementos de coste anteriores.

Se asocian nodos determinísticos a los costes de tipos 1 y 3, y nodos de azar a los costes de tipo 2 relacionados con impactos.

6.5. Modelos de preferencias

Se incluyen también los modelos de preferencias de las partes implicadas.

1. *Utilidad del defensor:* Modeliza las preferencias y actitudes frente al riesgo del defensor, la organización a la que apoyamos en su gestión de riesgos.

2. *Utilidad del atacante:* Análogamente, modelizaría las preferencias y actitudes frente al riesgo del atacante. En este caso, el posible ejecutor del ataque DDoS.

Se incluye un nodo de valor para el defensor. Asimismo, se incluye un nodo de valor por atacante, que, en este caso, se limita a uno.

6.6. Arcos

El único arco no estándar sería el que une el nodo salvaguardas con el nodo DDoS, dado que el atacante realizará su acción una vez identifique y conozca las salvaguardas implantadas en la organización. El resto de arcos modelizan las relaciones de dependencia habitualmente consideradas entre los nodos de decisión, azar y utilidad, véase Shachter (1986).

6.7. Solución

Para la solución procederíamos construyendo las probabilidades y utilidades del Defensor. Algunas de ellas requieren considerar el problema del atacante, por lo que modelizaríamos las creencias del Defensor sobre sus preferencias y creencias mediante utilidades y probabilidades aleatorias. Simularíamos entonces del problema del atacante para obtener la distribución de sus ataques, que incorporaríamos al problema del Defensor. Resolveríamos finalmente éste para obtener sus decisiones óptimas. En este caso concreto, la decisión fue instalar el firewall y el sistema anti-incendios y no adquirir seguros.

7. DISCUSIÓN

Hemos hecho una breve introducción a las metodologías de análisis de riesgos y de riesgos adversarios, con énfasis en problemas de ciberseguridad. En particular, el panorama actual de las metodologías de análisis y gestión de riesgos en ciberseguridad tiene características positivas, como una adecuada catalogación de activos, amenazas y salvaguardas a considerar para plantear el análisis inicial del aseguramiento de cualquier infraestructura empresarial. No obstante, este campo no está suficientemente formalizado, no contemplando, por ejemplo, el carácter dinámico de los riesgos. Por ello, se ha mostrado la necesidad de establecer un marco más adecuado.

Las ideas son, en cualquier caso, generales, y se han aplicado en muy diversas áreas, desde la seguridad nuclear a la seguridad aérea, pasando por la lucha contra el terrorismo y frente al cambio climático, mostrando así, de hecho, la capacidad de las Matemáticas para ayudar a construir un mundo más seguro.

AGRADECIMIENTOS

El trabajo de DRI viene apoyado por el proyecto de MINECO MTM2014-56949-C3-1-R, la Acción COST-ESF Action IS1304 sobre Juicios de Expertos y la Cátedra AXA-ICMAT de Análisis de Riesgos Adversarios.

REFERENCES

Andress, J., Winterfeld, S. (2013) *Cyber Warfare: Techniques, Tactics and Tools for Security Practitioners*. Syngress.

Banks, D., Ríos, J., y Ríos Insua, D. (2015) *Adversarial Risk Analysis,* CRC Press.

Bedford, T. y Cooke, R. (2001) *Probabilistic Risk Analysis*. Cambridge University Press.

Cox, L. A. (2008) What's Wrong with Risk Matrices?. *Risk Analysis,* Vol. 28, 497-512.

French, S. y Ríos Insua, D. (2000) *Statistical Decision Theory*. Arnold.

Kahnemann, D. y Tversky, A. (1974) Judgement under uncertainty: Heuristics and biases, *Science,* 185, 1124-1131.

Kaplan, S., y Garrick, B. (1981) On the quantitative definition of risk, *Risk Analysis,* 1, 11-27.

Leak Source (2014) CSEC *Document Reveals Suspected France Intelligence Spyware "Babar".* Leak Source.

Maslow, A. (1943) A theory of human motivation, *Psych. review,* 50, 370- 396.

Ríos Insua, D., Ríos, J. y Banks, D. (2009) Adversarial risk analysis. *Journal of the American Statistical Association,* 104(486):841–854.

Shachter, R. (1986) Evaluating influence diagrams. *Operations Research,* 34 (6):871–882.

World Economic Forum (2016) *Global Risks Report,* WEF.