Teoría de juegos y la dinámica de atacantes-defensores en ciberseguridad: Una revisión de literatura

Julian De Leon
Universidad del CEMA
Ciudad Autónoma de Buenos Aires, Argentina

Abstract— A medida que los sistemas informáticos ganan complejidad y los ciberataques se vuelven más sofisticados, las aproximaciones tradicionales para la ciberseguridad resultan cada vez más insuficientes. Si bien su objetivo sigue siendo prevenir intrusiones no autorizadas en redes y sistemas conectados, así como proteger datos sensibles, se requieren nuevos enfoques capaces de modelar las interacciones estratégicas entre atacantes y defensores. La teoría de juegos emerge como un marco analítico prometedor para comprender y abordar de manera más efectiva las preocupaciones de seguridad cibernética. El análisis realizado en este trabajo espera revelar las ventajas que ofrece esta disciplina para garantizar la seguridad de entornos informáticos complejos frente a amenazas en constante evolución.

Keywords: ciberseguridad, ciberataques, teoría de juegos.

I.INTRODUCCIÓN

En las últimas dos décadas, internet ha traído consigo nuevas oportunidades, pero también ha abierto la puerta a vulnerabilidades nunca vistas, convirtiendo el ciberespacio en un campo de batalla propicio donde individuos, empresas e incluso naciones-estados pueden operar de forma anónima, aprovechando la falta de leyes y regulaciones internacionales, así como el anonimato inherente a este ámbito. [1]

Defender sistemas contra intrusiones no autorizadas e inteligentes en el mundo cibernético supone un desafío aún mayor, dada la hiperdimensionalidad de este entorno, los tipos de armas digitales utilizadas, la velocidad de las operaciones y la gran cantidad de nodos que proteger frente a un número relativamente alto de potenciales atacantes. [2]

Ante este escenario, la teoría de juegos ofrece un marco natural para capturar las interacciones tanto defensivas como adversariales entre defensores y atacantes. Al abordar problemas donde varios jugadores con objetivos opuestos compiten entre sí, esta disciplina brinda un fundamento matemático para modelar y analizar cuestiones de seguridad cibernética. [3]

Este enfoque resulta idóneo para representar distintos escenarios y ha demostrado su utilidad en el ámbito de la ciberseguridad. De hecho, las aplicaciones más recientes de la teoría

de juegos abarcan temas tan diversos como la seguridad de infraestructuras críticas, la gestión de riesgos cibernéticos, la defensa de objetivos móviles, las amenazas internas, el aprendizaje automático adversario y las técnicas de engaño cibernético, entre otros, todos ellos en constante evolución. [1]

El resto del trabajo se organiza en varias secciones. En la Sección 2 se discuten brevemente los aspectos fundamentales de la ciberseguridad y sus objetivos. La relación entre la teoría de juegos y la ciberseguridad se presenta en la Sección 3. En la Sección 4 se realiza un análisis exhaustivo de la literatura, y finalmente, en la Sección 5, se presentan las conclusiones derivadas de los hallazgos.

II.CIBERSEGURIDAD

Antes de comenzar nuestra revisión de literatura, es conveniente definir de antemano los términos que utilizaremos en las secciones subsecuentes. Uno de los términos que se repetirá frecuentemente es el de ciberseguridad, por lo que entender su significado nos proporcionará un modelo mental adecuado para discutir qué tipo de activos intentarán proteger los defensores y cuáles intentarán corromper los atacantes. Para ello, utilizaremos la definición proporcionada por la Unión Internacional de Telecomunicaciones (ITU), que define la ciberseguridad como:

> La colección de herramientas, políticas, conceptos de seguridad, salvaguardas, directrices, enfoques de gestión de riesgos, acciones, capacitación, mejores prácticas, garantías y tecnologías que se pueden utilizar para proteger el entorno cibernético y los activos de los usuarios y las organizaciones. Estos activos incluyen dispositivos informáticos conectados, personal, infraestructura, aplicaciones, servicios, sistemas de telecomunicaciones y la totalidad de la información transmitida y/o almacenada en el entorno cibernético. La ciberseguridad busca asegurar la consecución y el mantenimiento de las propiedades de seguridad de los activos de la organización y del usuario contra los riesgos relevantes en el entorno cibernético, con objetivos generales de seguridad que comprenden disponibilidad, integridad (que puede incluir autenticidad y no repudio) y confidencialidad. [4]

De esta definición se desprende la multidimensionalidad del concepto de ciberseguridad y las múltiples aristas que un defensor debe considerar para proteger los activos. En la siguiente sección, exploraremos la relación entre Teoría de Juegos y Ciberseguridad.

III.TEORÍA DE JUEGOS Y CIBERSEGURIDAD

En el campo de la ciberseguridad, la teoría de juegos es una herramienta útil para entender cómo reaccionan los defensores ante los atacantes y viceversa. Mediante un simple juego de dos jugadores, se puede modelar la interacción estratégica entre el atacante y el defensor, en la cual ambos buscan maximizar sus propios beneficios.

La siguiente tabla, ilustra un juego genérico no cooperativo de información perfecta con dos jugadores, pudiendo cada uno de ellos realizar dos acciones posibles:

TABLA 1. REPRESENTACIÓN NORMAL DE UN JUEGO NO COOPERATIVO [5]

	ATACANTE		
DEFENSOR		H	S
	S	(3, -3)	(-2,2)
	N	(-10,10)	(0,0)

El defensor, puede elegir entre dos estrategias: implementar medidas de seguridad (S) o no implementar medidas de seguridad (N), mientras que el atacante puede atacar (H) o no atacar (N). Para mantener el análisis sencillo, no involucraremos cierta probabilidad p ($0 < p < 1$) de que el hackeo tenga éxito a pesar de la vigilancia.

Si el defensor implementa medidas de seguridad (S) y el atacante decide atacar (H), el pago es 3 (1). Si el defensor implementa medidas de seguridad (S) y el atacante decide no atacar (N), el pago es -2 (2). Siendo (2) el costo de implementar medidas de seguridad sin enfrentar un ataque. Si el defensor no implementa medidas de seguridad (N) y el atacante decide atacar (H), el pago es -10 (3). Siendo (3) la pérdida del defensor debido a un ataque exitoso. Por último, si el defensor no implementa medidas de seguridad (N) y el atacante decide no atacar (N), el pago es 0 (2). No hay ganancia ni pérdida ya que no hay medidas de seguridad implementadas ni ataque realizado.

En este contexto, el atacante y el defensor interactúan de manera no cooperativa: los jugadores tienen intereses contrapuestos (las ganancias de uno corresponden a las pérdidas del otro) caracterizados por una propiedad de suma cero. En otras palabras, no hay valor en cooperar en tales interacciones porque ningún jugador puede ganar sin que el otro pierda. Además, el juego es completamente estratégico, ya que el mejor movimiento de cualquier jugador depende

estrictamente del movimiento del otro jugador. [5]

Aunque el juego presentado no refleja completamente una situación real, es crucial destacar que las acciones del defensor afectan las estrategias del atacante y, de igual forma, las acciones del atacante influyen en las estrategias del defensor. Por lo tanto, la eficacia de un mecanismo de defensa puede evaluarse en función de estos comportamientos estratégicos. En consecuencia, la teoría de juegos es esencial para estudiar las decisiones estratégicas de los defensores y para analizar los incentivos de los atacantes. [6]

IV. REVISIÓN DE LITERATURA

Como se ha expresado en secciones anteriores, el enfoque de la teoría de juegos puede capturar la interacción entre defensores y atacantes. A continuación, mostraremos las principales aplicaciones de la teoría de juegos en el ámbito de la ciberseguridad.

A. Teoría de Juegos & Ataques DoS (Denial-of-Service)

La teoría de juegos ofrece un enfoque estructurado y cuantitativo para modelar y analizar las interacciones estratégicas entre atacantes y defensores en el contexto de los ataques de denegación de servicio (DoS) y denegación de servicio distribuida (DDoS) [7]. En [8] los autores exploran la aplicación de esta disciplina para desarrollar un robusto mecanismo de defensa contra estos ataques que intentan agotar el ancho de banda de la red mediante un intenso tráfico malicioso. Su propuesta representa la interacción estratégica entre atacante y defensor como un juego de suma cero de dos jugadores, donde el primero busca maximizar la interrupción de los servicios de red mediante inundaciones mientras que el segundo apunta a minimizarla optimizando las estrategias de gestión del tráfico.

El modelo se construye en torno al concepto de equilibrio de Nash, donde la estrategia de cada jugador es óptima dada la del contrario. El atacante debe decidir la tasa óptima de ataque y el tamaño de la botnet para maximizar el impacto minimizando costos y detección. El defensor, por su parte, debe establecer un umbral óptimo de tráfico para diferenciar flujos legítimos de maliciosos, ajustándolo dinámicamente según las condiciones de red en tiempo real. Así, se formulan las estrategias y pagos de un juego cuya resolución conduce a estrategias de equilibrio de Nash. Las simulaciones demuestran que este enfoque mitiga eficazmente los ataques

DoS/DDoS al convergir a dichas estrategias óptimas, mejorando las capacidades de toma de decisiones defensivas e introduciendo un marco cuantitativo para evaluar su efectividad amte este tipo de amenazas en red.

B. Dilema del Prisionero & Ciberataques

En [9] se aplica el concepto clásico de la teoría de juegos del dilema del prisionero para analizar las complejidades de las relaciones entre naciones-estados en el ámbito de los ciberataques. Al analizar las interacciones entre naciones poderosas, naciones menos poderosas y naciones "trampolín", la investigación resalta las complejidades y consideraciones estratégicas en las relaciones cibernéticas internacionales.

El marco del dilema del prisionero revela por qué las naciones-estado podrían optar por no cooperar, incluso cuando la cooperación mutua arrojaría mejores resultados. En los conflictos cibernéticos, esta falta de confianza y temor a la explotación a menudo conduce a decisiones subóptimas. Los autores analizan los ciberataques de 2007 contra Estonia, que ilustraron un escenario clásico del dilema del prisionero. A pesar de las sospechas de Estonia de que los ataques se originaron en Rusia, no hubo cooperación entre las dos naciones para abordar el problema debido a la desconfianza política y estratégica. La nación menos poderosa se enfrenta al dilema de acusar a la nación poderosa arriesgándose a una escalada, o guardar silencio. Mientras tanto, la nación poderosa tiene pocos incentivos para cooperar y ayudar al estado más débil. La respuesta de Estonia fue buscar asistencia de la OTAN en lugar de intentar cooperar directamente con Rusia.

TABLA 2. DILEMA DEL PRISIONERO PARA RUSIA Y ESTONIA [9]

		Estonia	
		Cooperar	No cooperar
Rusia	Cooperar	**Escenario improbable**: 1) Los hackers son castigados; 2) Se disuaden futuros ataques	**Escenario Altamente Improbable:** 1) Rusia niega responsabilidad; 2) Las relaciones entre Rusia y Estonia empeoran
	No cooperar	**Escenario probable:** 1) Estonia busca ayuda de Rusia; 2) El Tratado de Asistencia Mutua no tiene valor	**Escenario Altamente Probable:** 1) Los ataques se intensifican; 2) Los países son incapaces de vigilar su ciberespacio 3) Estonia busca ayuda de la OTAN y la UE.

De manera similar, cuando se involucran dos naciones poderosas como Estados Unidos y China, las apuestas y las complejidades aumentan. Ambas naciones participan en el ciberespionaje, motivadas por la falta de confianza y el deseo de salvaguardar sus intereses. Aquí, el dilema del prisionero se manifiesta en el ciclo continuo de ataques y negaciones, con ambas partes temiendo las repercusiones de la cooperación.

A través de estos modelos de teoría de juegos, se demuestra cómo se manifiesta el dilema del prisionero en diferentes escenarios de ciberataques, con los estados-nación tendiendo a un comportamiento no cooperativo debido a la falta de mecanismos de confianza y transparencia en el intercambio de información. El análisis argumenta que sólo mediante el fomento de la cooperación, las naciones pueden superar este dilema cibernético, aunque la cooperación es un desafío debido a la desconfianza inherente y los incentivos estratégicos para no hacerlo.

C. Teoría de Juegos & Modelización de estrategias

En [3] se propone un marco teórico de juegos dinámicos llamado "hiper defensa" para analizar las interacciones entre atacantes y defensores como un juego de seguridad no cooperativo. El modelo captura la naturaleza continua y evolutiva de la ciberguerra, donde tanto los atacantes como los defensores son jugadores racionales que ajustan dinámicamente sus estrategias para maximizar sus beneficios.

El marco modela a los atacantes y defensores con múltiples niveles de estrategias, que difieren en efectividad, costo y las posibles ganancias o daños. Por ejemplo, los atacantes pueden optar por no atacar, un ataque de baja intensidad o un ataque de alta intensidad, cada uno requiriendo diferentes niveles de recursos y generando distintos grados de impacto. De manera similar, los defensores pueden elegir no defender, una defensa de bajo nivel o una defensa de alto nivel, con sus correspondientes costos y efectividad. El modelo de teoría de juegos está diseñado para encontrar el equilibrio de Nash, donde tanto atacantes como defensores adoptan estrategias que equilibran sus gastos de recursos contra sus ganancias o pérdidas esperadas.

A través de estudios de caso de diferentes tipos de ataques en redes, como ataques de inundación "hello", ataques de malware y ataques de adivinanza de contraseñas, el documento demuestra la aplicabilidad práctica del modelo propuesto. El sistema de hiper defensa ajusta dinámicamente las estrategias defensivas en función de las intensidades de ataque observadas y las restricciones de recursos, lo que conduce a una protección más efectiva y eficiente en

comparación con los mecanismos de defensa estáticos.

Este enfoque permite una representación más realista de la naturaleza dinámica de los ataques cibernéticos y las defensas, donde ambas partes adaptan continuamente sus estrategias en respuesta a las acciones del oponente.

CONCLUSION

En este trabajo, se ha explorado la aplicabilidad de los métodos de teoría de juegos para resolver problemas de ciberseguridad. Muchos de estos problemas son juegos no cooperativos, por lo que la teoría de juegos es una herramienta eficaz para modelar las interacciones, así como las actividades de ciberguerra entre atacantes y defensores.

Como se ha descrito, investigaciones recientes han demostrado la efectividad de la teoría de juegos en diversas áreas de la ciberseguridad: desde modelar y mitigar ataques de denegación de servicio (DoS y DDoS) al optimizar las interacciones entre atacantes y defensores, hasta analizar las complejas relaciones estratégicas entre estados-nación en el contexto de ciberataques, y desarrollar marcos dinámicos que ajustan las estrategias de defensa en tiempo real. Estos enfoques no solo mejoran la comprensión del comportamiento adversario, sino que también proporcionan herramientas para diseñar estrategias defensivas más adaptativas y resilientes frente a amenazas en constante evolución.

Aunque la teoría de juegos es relevante para la seguridad, enfrenta varios desafíos para desarrollar enfoques viables, desde la complejidad de calcular una estrategia de equilibrio teórico, hasta cuantificar parámetros de seguridad como el riesgo, la privacidad y la confianza, que son esenciales para definir las funciones de utilidad para los participantes en un juego [10]

Por lo tanto, no deberíamos considerar la teoría de juegos como nuestra única herramienta para abordar los problemas presentados, sino como un complemento para entender las acciones estratégicas entre atacantes y defensores, lo que facilitará el desarrollo de estrategias de defensa más eficaces y adaptativas, mejorando así la resiliencia de los sistemas frente a amenazas en evolución.

Referencias

[1] F. Anwar, B. U. I. Khan, R. F. Olanrewaju, B. R. Pampori, and R. N. Mir, "A Comprehensive Insight into Game Theory in relevance to Cyber Security," Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 8, no. 1, Mar. 2020, doi: https://doi.org/10.52549/ijeei.v8i1.1810.

[2] S. Shiva, D. Dasgupta and Q. Wu, "Game Theoretic Approaches to Protect Cyberspace", Final Technical Report, Department of Computer Science University of Memphis, Memphis, TN, USA, p. 86, 2010.

[3] Afraa Attiah, M. Chatterjee, and C. C. Zou, "A Game Theoretic Approach to Model Cyber Attack and Defense Strategies," May 2018, doi: https://doi.org/10.1109/icc.2018.8422719.

[4] "Definition of cybersecurity," www.itu.int. https://www.itu.int/en/ITU-T/studygroups/2013-2016/17/Pages/cybersecurity.aspx

[5] F. Moisan and C. Gonzalez, "Security under Uncertainty: Adaptive Attackers Are More Challenging to Human Defenders than Random Attackers," Frontiers in Psychology, vol. 8, Jun. 2017, doi: https://doi.org/10.3389/fpsyg.2017.00982.

[6] C. T. Do et al., "Game Theory for Cyber Security and Privacy," ACM Computing Surveys, vol. 50, no. 2, pp. 1–37, Jun. 2017, doi: https://doi.org/10.1145/3057268.

[7] O. Thakoor, P. Vayanos, M. Tambe, and M. Yu, "Game Theory for Strategic DDoS Mitigation." Accessed: May 26, 2024. [Online]. Available: https://minlanyu.seas.harvard.edu/writeup/optmas20.pdf

[8] B. Kumar and B. Bhuyan, "Using game theory to model DoS attack and defence," Sādhanā, vol. 44, no. 12, Nov. 2019, doi: https://doi.org/10.1007/s12046-019-1228-4.

[9] N. Kostyuk, "The digital prisoner's dilemma: Challenges and opportunities for cooperation," 2013 World Cyberspace Cooperation Summit IV (WCC4), Silicon Valley, CA, USA, 2013, pp. 1-6, doi: https://doi.org/10.1109/WCS.2013.7050508

[10] A. Iqbal, L. J. Gunn, M. Guo, M. Ali Babar, and D. Abbott, "Game Theoretical Modelling of Network/Cybersecurity," IEEE Access, vol. 7, pp. 154167–154179, 2019, doi: https://doi.org/10.1109/access.2019.2948356