Mechanism Design for Cooperation in Cybersecurity *

Alistair Simmons[1]

Duke Kunshan University, Kunshan, Jiangsu 215316, China
your email
LinkedIn

Abstract. This paper proposes a novel mechanism design aimed at increasing cooperation among companies in joint cybersecurity efforts, introducing an unprecedented Automated Indicator Sharing (AIS) system rooted in economic and game-theoretic principles. By incorporating structured incentives, blockchain technology for security and transparency, and the strategic utilization of trusted third-party mediators, this mechanism not only addresses traditional barriers such as the fear of exposing sensitive information and trust deficits but also pioneers a model for seamless and secure data exchange. The AIS system designed here stands out in the cybersecurity literature for its automated, transparent, and incentive-compatible approach to threat intelligence sharing across industries, promoting a higher level of collaboration and mutual benefit in cybersecurity defenses.

Notes: In submission to Problem Set 2 for COMPSCI/ECON 206 Computational Microeconomics, 2024 Spring Term (Seven Week - Second) instructed by Prof. Luyao Zhang at Duke Kunshan University.

Keywords: computational economics · game theory · innovative education · cybersecurity

1 Introduction

Cyber threats continue to evolve in complexity and scale, posing significant risks across various industries. Despite the clear benefits of collaborative cybersecurity, numerous barriers—including fear of exposing sensitive information and lack of trust—hamper effective intercompany cooperation. This paper introduces a mechanism design that addresses these challenges through structured incentives, enhanced by the immutable and transparent nature of blockchain technology, and the strategic use of trusted mediators.

* **Acknowledgments**: I would like to thank Professor Luyao Zhang for her incredible insights and guidance in learning about game theory and structuring this research.

2 Background

The Trust Game, introduced by Joyce Berg, John Dickhaut, and Kevin McCabe in 1995, serves as a foundational concept in behavioral economics and game theory, analyzing trust and reciprocity through interactions where a sender decides how much of an initial stake to trust a receiver with, which is then potentially multiplied and partially returned. Another crucial model, the Stag Hunt, exemplifies situations requiring mutual trust and cooperation to achieve optimal collective outcomes, such as when two hunters must decide independently whether to collaborate on hunting a stag for a higher reward or to pursue individual hares for a guaranteed but lower payoff. This concept is particularly relevant to cybersecurity, where similar cooperation is facilitated by initiatives like Information Sharing and Analysis Centers (ISACs). ISACs help companies within an industry share crucial data on threats and vulnerabilities, enhancing collective defenses much like hunters in a Stag Hunt deciding to pursue the stag together, thereby applying these theoretical models to real-world challenges in cybersecurity to foster a cooperative environment where information sharing is both encouraged and strategically beneficial.

“‘

– [1]
– Berg et al. [1]
– Berg et al.
– 1995

– [2]
– Skyrms [2]
– Skyrms
– 2004

– [3]
– ISA [3]
– ISA
– 2004

A Scholarly Intervention

Traditional game theory applications in cybersecurity have primarily concentrated on the adversarial aspect of cyber threats—predicting and responding to cyberattacks. This body of work often conceptualizes cybersecurity as a series of strategic games between attackers and defenders, each with their own set of strategies and payoffs. his adversarial perspective involves analyzing various strategies and their associated payoffs to determine the most effective defensive tactics against

potential threats (Pawlick et al., 2015). The literature has extensively utilized the concept of Nash Equilibrium to predict the actions of rational adversaries, aiming to optimize individual defensive strategies rather than encourage collective security measures (Roy et al., 2010). The literature has focused on understanding the cost-benefit analysis of different defensive moves, the probability of attacks, and the strategic behavior of malicious actors. It employs game-theoretic concepts such as Nash Equilibrium to ascertain the likely actions of rational adversaries, with the overarching goal of identifying optimal defensive postures and resource allocation to thwart potential attacks. However, this adversarial focus has historically meant less emphasis on mechanisms to promote collaboration among different entities in the cybersecurity landscape. While the traditional approach is adept at modeling conflict and defense, it does not inherently provide tools for fostering cooperation or trust, such as sharing threat intelligence or pooling resources for mutual benefit. Such cooperative actions are essential for enhancing overall cybersecurity resilience but require a shift in focus from individualistic strategy optimization to collective security benefits. Consequently, there's a growing recognition in the field of the need to create and analyze game-theoretic mechanisms that not only defend against attackers but also build and sustain alliances and shared defense infrastructures.

- [4]
- Pawlick et al. [4]
- Pawlick et al.
- 2015

- [5]
- Roy et al. [5]
- Roy et al.
- 2010

- [6]
- Alpcan and Başar [6]
- Alpcan and Başar
- 2010

- [7]
- Loukas [7]
- Loukas
- 2013

B Constructing Mechanism Design

B.1 Theoretical Concept

The proposed mechanism design leverages a contribution-based pricing model to incentivize companies to share critical cybersecurity information, with rewards scaled according to the severity and scope

of disclosed vulnerabilities (Anderson, 2001). The compensation structure differentiates between critical industries, core businesses, and small businesses, offering higher rewards for root vulnerabilities in critical sectors and lower rewards for specific vulnerabilities in smaller businesses (Pawlick and Zhu, 2015). Implementation involves an independent panel that verifies and values each reported vulnerability to ensure fair compensation, supported by a dynamic pricing strategy that adjusts based on the cybersecurity landscape (Varian, 2004). Beyond monetary rewards, contributors gain reputation benefits and access to a collective intelligence pool (Schneier, 2012). This model not only enhances overall cybersecurity by encouraging the sharing of vital information but also targets incentives to maximize prevention of cyberattacks, fostering a proactive, collaborative security environment (Akerlof, 1970).

- [8]
- Akerlof [8]
- Akerlof
- 1970

- [9]
- Schneier [9]
- Schneier
- 2012

- [10]
- Varian [10]
- Varian
- 2004

- [11]
- Anderson [11]
- Anderson
- 2001

B.2 Market Value

The proposed mechanism design leverages a contribution-based pricing model to incentivize companies to share critical cybersecurity information, with rewards scaled according to the severity and scope of disclosed vulnerabilities (Anderson, 2001). The compensation structure differentiates between critical industries, core businesses, and small businesses, offering higher rewards for root vulnerabilities in critical sectors and lower rewards for specific vulnerabilities in smaller businesses (Pawlick and Zhu, 2015). Implementation involves an independent panel that verifies and values each reported vulnerability to ensure fair compensation, supported by a dynamic pricing strategy that adjusts based on the cybersecurity landscape (Varian, 2004). Beyond monetary rewards, contributors gain reputation benefits and access to a collective intelligence pool (Schneier, 2012). This model not only enhances

overall cybersecurity by encouraging the sharing of vital information but also targets incentives to maximize prevention of cyberattacks, fostering a proactive, collaborative security environment (Akerlof, 1970).

– [12]
– IBM [12]
– IBM
– 2020

– [13]
– Morgan [13]
– Morgan
– 2020

B.3 Application

The cyber information sharing system utilizes blockchain technology to create a decentralized, immutable ledger, enhancing trust by preventing data tampering and ensuring transparency. Smart contracts automate the verification and compensation processes, providing automatic and transparent payments to contributors when predefined criteria are met, thus improving efficiency and reducing administrative burdens. These technologies also secure data transactions against unauthorized access and enable a complete, auditable trail of all information exchanges, crucial for integrity and compliance checks. This integration of blockchain and smart contracts fundamentally transforms cyber information sharing into a more reliable, efficient, and secure practice, guaranteeing fair compensation and bolstering trust among participants.

B.4 Modeling The Benefits

Stag and Hare Model I am using the Stag Hunt game to model the economic benefits of creating incentives for exchanging cybersecurity information, transforming partnerships from noncooperative to cooperative. This approach helps demonstrate how tailored incentives can effectively alter the payoff landscape, encouraging entities to collaborate in sharing vital cybersecurity data. By adjusting the risk-reward equation to favor cooperation, my model predicts a shift in strategic behaviors, leading to enhanced mutual gains and stronger collective defenses against cyber threats. This transformation is crucial for developing resilient cybersecurity networks where shared information leads to improved situational awareness and more robust responses to emerging threats.

Stag and Hunt Without Incentives This matrix captures the decision-making scenario where two companies, particularly Company 1 with three times the cybersecurity knowledge of Company 2, consider whether to share their information. The benefit for sharing is gaining access to the other company's cybersecurity knowledge. However, sharing entails a risk of -0.5, which could result from competitors exploiting the shared information or it leading to a devaluation of the company's market standing. This model represents the basic cooperative dynamics without additional incentives, reflecting the natural trepidations companies might have about sharing sensitive information.

Table 1. Matrix 1: No Incentives

	C	D
C	$(3.5, 3.5)$	$(0.5, 4)$
D	$(4, 0.5)$	$(1, 1)$

Stag and Hunt With Incentives n this upgraded scenario, companies are enticed to share information through proportional incentives valued at 0.2 percent of their overall cybersecurity knowledge. Additionally, risks associated with sharing are mitigated to -0.3 due to the implementation of blockchain and smart contract technologies, which provide secure, transparent mechanisms for sharing and reduce the potential for misuse or devaluation. This matrix demonstrates how the introduction of incentives and secure sharing infrastructures can potentially alter the decision calculus of the companies, making the option to cooperate more appealing by offsetting inherent risks.

Table 2. Matrix 1: No Incentives

	C	D
C	$(3.9, 4.3)$	$(0.9, 3)$
D	$(4, 0.9)$	$(1, 1)$

B.5 Comparative Analysis

Using Nash.py and QuantEcon, the equilibrium for the payoff matrices were calculated to be as follows:

No Incentives Model (Matrix 1) The Nash Equilibrium calculations result in three outcomes: one where both companies choose to cooperate (C, C), one where both choose not to cooperate (D, D), and a mixed-strategy equilibrium where they randomize their strategies with certain probabilities. This indicates that without additional incentives, the company with more information does not have a clear incentive to share, as the risk of sharing may outweigh the benefits.

Incentives Model (Matrix 2) The Nash Equilibrium is consistently at (C, C), where both companies choose to cooperate. This suggests that with the incentives in place, it becomes in the interest of both companies to share information. The provision of incentives aligns their strategies towards mutual cooperation, illustrating the effectiveness of the incentives in promoting information sharing.

B.6 Usefulness and Application

The Nash Equilibria obtained from the payoff matrices reveal a significant shift in strategy due to the mechanism design, highlighting its influence on cybersecurity information sharing and economic utility. In the no incentives model, the existence of multiple equilibria, including a mixed-strategy one, indicates a state of indecision and risk aversion among companies, particularly for the one with more information. This aligns with Akerlof's concept of information asymmetry, where the risk of sharing valuable data can lead to a non-cooperative equilibrium (Akerlof, 1970).
However, when incentives are introduced, the equilibrium shifts to consistent cooperation, as the proportional rewards and reduced risks enhance the benefits of sharing, outweighing the potential costs. This change in strategy can be understood through Nash's cooperative game theory, which suggests that individuals are capable of adjusting their strategies for mutual benefit when properly incentivized (Nash, 1951). The integration of blockchain and smart contracts in this model further mitigates risks, building trust and cementing cooperation as the rational, utility-maximizing choice. Therefore, the incentive-based mechanism design not only promotes increased information sharing but also provides substantial economic utility by fostering a more secure and collaborative cyber environment.

- [8]
- Akerlof [8]
- Akerlof
- 1970

- [14]

– Nash [14]
– Nash
– 1951

C Game Theory Glossary Tables

Table 3. Game Theory Glossary

Term	Definition	Source
Stag and Hare Game	A game that illustrates the conflict between safety and social cooperation. Players must choose between a risky joint venture (stag) for a great reward or a safe individual action (hare) with a lesser payoff.	
Utility	Utility in game theory is a representation of preferences over some set of goods and services.	Von Neumann & Morgenstern (1944)
Strategy	A strategy is a complete plan of action a player will follow throughout the game, irrespective of the game's history.	Von Neumann & Morgenstern (1944)
Nash Equilibrium	A situation where no player can benefit by changing their strategy while the other players keep theirs unchanged.	Nash (1950)
Dominant Strategy	A strategy that is optimal for a player, no matter what the opponents play.	Nash (1950)
Zero-Sum Game	A situation in which one participant's gains or losses are exactly balanced by the losses or gains of the other participants.	Von Neumann & Morgenstern (1944)
Cooperative Game	A game where players can form coalitions and make binding agreements to achieve collective action that influences individual payoffs.	Von Neumann & Morgenstern (1944)

Bibliography

[1] J. Berg, J. Dickhaut, and K. McCabe, "Trust, reciprocity, and social history," *Games and Economic Behavior*, vol. 10, no. 1, pp. 122–142, 1995.

[2] B. Skyrms, *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, 2004.

[3] "Information sharing and analysis centers (isacs)," https://www.nationalisacs.org/, 2004, accessed: date-of-access.

[4] J. Pawlick, S. Farhang, and Q. Zhu, "A game-theoretic flipit model with multiple defenders," *International Conference on Decision and Game Theory for Security*, pp. 135–154, 2015.

[5] S. Roy, C. Ellis, S. G. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A survey of game theory as applied to network security," in *43rd Hawaii International Conference on System Sciences*. IEEE, 2010, pp. 1–10.

[6] T. Alpcan and T. Başar, "Network security: A decision and game-theoretic approach," *Cambridge University Press*, vol. 1, 2010.

[7] G. Loukas, *Cyber-Physical Attacks: A Growing Invisible Threat*. Butterworth-Heinemann, 2013.

[8] G. A. Akerlof, *The Market for Lemons: Quality Uncertainty and the Market Mechanism*. Oxford University Press, 1970, vol. 84, no. 3.

[9] B. Schneier, *Liars and Outliers: Enabling the Trust that Society Needs to Thrive*. John Wiley & Sons, 2012.

[10] H. R. Varian, "System reliability and free riding," *Economics of Information Security*, vol. 12, pp. 1–15, 2004.

[11] R. Anderson, *Why Information Security is Hard-An Economic Perspective*. ACM, 2001.

[12] IBM, "Cost of a data breach report 2020," https://www.ibm.com/security/digital-assets/cost-data-breach-report/, 2020.

[13] S. Morgan, "Cybercrime to cost the world $10.5 trillion annually by 2025," *Cybercrime Magazine*, 2020.

[14] J. Nash, "Non-cooperative games," *Annals of Mathematics*, pp. 286–295, 1951.