A Game Theoretical Approach to Communication Security

Assane Gueye

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2011-19
http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-19.html

March 14, 2011

Acknowledgement

I would like to thank my adviser Prof. Jean Walrand and all my dissertation committee members.
Thanks to my parents for all they have done for me.

A Game Theoretical Approach to Communication Security

by

Assane Gueye

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering–Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jean C. Walrand, Chair
Professor Venkat Anantharam
Professor Shachar Kariv
Professor Vern Paxson

Spring 2011

A Game Theoretical Approach to Communication Security

Abstract

A Game Theoretical Approach to Communication Security

by

Assane Gueye

Doctor of Philosophy in Engineering–Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Jean C. Walrand, Chair

The increased reliance on the Internet has made information and communication systems more vulnerable to security attacks. Many recent incidents demonstrate this vulnerability, such as the rapid propagation of sophisticated malwares, the fast growth of botnets, denial-of-service (DoS) attacks against business and government websites, and attacks against the power grid system.

Experts must design and implement security solutions to defend against well organized and very sophisticated adversaries such as malicious insiders, cybercriminals, cyberterrorists, industrial spies, and, in some cases, nation-state intelligence agents.

Instead of designing a defense against a specific attack, Game Theory attempts to design a defense against a sophisticated attacker who plans in anticipation of a complex defense. By including this 'second-guessing' element into the design process, Game Theory has the potential of crafting improved security mechanisms. In addition, Game Theory can model issues of trust, incentives, and externalities that arise in security systems.

This thesis illustrates the potential usefulness of Game Theory in security. By modeling the interactions between defenders and attackers as games in three types of common communication scenarios, we predict the adversaries' attacks, determine the set of assets that are most likely to be attacked, and suggest defense strategies for the defenders.

The first example is a communication scenario where some components might be compromised. Specifically, we consider Bob who is receiving information that might be corrupted by an attacker, Trudy. We model the interaction between Trudy and Bob as a zero-sum game where Trudy chooses whether and how to corrupt the data and Bob decides how much he should trust the received information. By analyzing the Nash equilibrium of the game, we have determined when Bob should trust the received information, and how Trudy should corrupt the data. We have also shown how a *challenge-response* option for Bob can deter Trudy from corrupting the information.

The second example is a scenario where an intelligent virus is attempting to infect a network protected by an Intrusion Detection System (IDS). The IDS detects intrusions by analyzing the volume of traffic going inside the network. We model the interaction of the intelligent virus and the IDS as a zero-sum game where the IDS chooses the detection threshold, while the virus is trying to choose its infection rate to maximize its ultimate spreading. Using a Markov chain model, we compute the Nash equilibria of the game and analyze them. In general, a more aggressive virus is more damaging but is also faster to detect. Hence, in its best attack, the intelligent virus chooses an infection rate that balances between an aggressive attack that can be easily detected and a slow attack that causes less damage. The best defense strategy against such a sophisticated virus is to quarantine the traffic and analyze it prior to letting it go inside the network (in addition to setting the optimal threshold).

The third example is a blocking security game. For this game, given a finite set of resources $\mathcal{E} = \{e_1, \ldots, e_m\}$, a defender needs to choose a feasible subset $T \subseteq \mathcal{E}$ of resources to perform a mission critical task. The attacker, at the same time, tries to disrupt the task by choosing one resource $e \in \mathcal{E}$ to attack. Each resource $e \in \mathcal{E}$ has a cost of attack $\boldsymbol{\mu}_e$. The defender loses some value $\boldsymbol{\lambda}_{T,e}$ whenever the task is disrupted (i.e. the attacked resource e belongs to his subset T). This loss goes to the attacker. We analyze the game by using the combinatorial tools of blocking pairs of matrices (hence the name blocking security game). We introduce the notion of critical subset of resources and use this notion to define a vulnerability metric for the task. We find that, in Nash equilibrium, the attacker always targets a *critical set* of resources and the defender chooses a feasible subset that minimally intersects that critical subset. We illustrate the model with two examples of communication scenarios that consider design of network topology in the presence of a strategic adversary. The first example studies a scenario where a network manager is choosing a spanning tree of a graph while an attacker is trying to cut the tree by attacking one link of the graph. One of our findings in this scenario is that, the usual *edge-connectivity* metric for a graph is not the appropriate vulnerability measure in a network where strategic adversaries are present. The second example deals with a *supply-demand* network where a network manager is choosing a feasible flow to transport the maximum amount of goods from a set of sources to a set of destinations, and an attacker is trying to minimize this by attacking an arc of the network. In this case, we find that critical subsets of links are cutsets that maximize the minimum fraction of goods carried per link of the cutset. In most cases, these correspond to minimum cutsets of the graph.

Although computing Nash equilibria of a two-player game is generally complex, we have shown how, for a class of blocking games, one can compute a critical set of resources (hence a Nash equilibrium) in polynomial time.

To my parents, my family, my wife, my son, my friends...

To my late mother...who did not live long enough to witness this moment...

Black woman, African woman,
O mother, I think of you...

O Dâman, O mother,
Who carried me on your back, who nursed me,
Who governed my first steps,
Who was the first to open my eyes to the beauties of the world,
I think of you...

Woman of the fields, woman of the rivers, woman of the great river,
O mother, I think of you...
O Dâman, O mother, who wiped my tears,
Who cheered up my heart, who patiently dealt with my caprices,
How I would love to still be near you, be a child next to you.

Simple woman, woman of resignation,
O mother, my thoughts are always of you.
O Dâman, Dâman of the great family of true believers,
My thoughts are always of you,
They accompany me with every step, O Dâman, my mother,
How I would love to still feel your warmth,
Be a child next to you.

Black woman, African woman,
O mother, thank you; thank you for all that you have done for me, your son,
So far away yet so close to you!

Translated from the original French text of Camara LAYE
by Deborah Weagel, University of New Mexico

Contents

List of Figures

List of Tables

Acknowledgments

First, I thank The Lord who has given me life, health, and countless blessings. He has provided me with the strength to go through with this experience for the past 6 years and I would like to praise Him and thank Him.

I am heartily thankful to my supervisor Prof. Jean Walrand for his encouragement, guidance, support, and patience. A student's life does not end on campus and Jean's support has reached me far beyond the Berkeley halls and conference rooms. Thank you for all you have done for me Jean, and thanks to your family who has joined you in that.

Prof. Venkat Anantharam has generously given me his time and expertise at a crucial time of this dissertation. Without Venkat's help, finishing this thesis would have been a lot more difficult. I would like to show him my endless gratitude.

I am grateful to Prof. Vern Paxson for his valuable feedback, his input, his guidance, and his kindness. Prof. Paxson's sound understanding of the security problem has made me love the subject after I took his class. Thank you for all the efforts you have spent in teaching us this very important subject.

I am very thankful to Prof. Shachar Kariv who has guided my steps in Game Theory and has, from the initial to the final level, enabled me to develop a good understanding of the subject. Thank you for your patience, your kindness, and your feedback.

I would like to show my gratitude to all the professors who have taught me. Also, many thanks to the entire Berkeley NetEcon group. John, Shyam, Abhay, Nikhil, Libin, Jiwoong, Galina, David, Barlas, Vijay, your suggestions and feedback were very valuable to me.

My endless thanks go to Prof. Eric Brewer and the TIER group who have supported my research on Information and Communication Technologies for Development (ICT4D). Your devotion to bringing technologies to people in the less developed nations is noble and has inspired me a lot. I would specially like to thank George Sharffenberger and Scott McNeil for their guidance and many suggestions.

I need to express my deep gratitude to Daniel Mouen-Makoua and the Gateway Foundation for their financial support. Daniel did not only provide finance, he has been a friend and a mentor to me. Thank you for that. Also, I must acknowledge Ambassador Martin Brennan, Josiane Siegfried and the entire International House board for having given me the opportunity to live in such a wonderful environment. Thanks to Martin for his kindness and thanks to all the friends I have met at I-House.

My thanks must go to all my friends who have helped me maintain a balanced on-campus/off-campus life. Especially, I need to express my deepest gratitude to the members of the Dahiratoul Tayssiroul Assir, Touba-Oakland. They have helped me develop an enriching spiritual life and have supported me every time I needed them. Many thanks to Khadim, Lahad, and Issa who have welcomed me to the Bay Area and have introduced me to this great community. I am always

grateful to my spiritual guides Serigne Amsatou Mbacke and Serigne Khalifa Diop for their guidance and prayers.

I am forever grateful to my family for their unconditional love and support. My parents have given me the best of everything in this world, they have made endless sacrifices for my education and my success, and they have taught me and raised me to always be a good person. I certainly cannot thank them enough.

Last but not least, endless thanks to my wife for her love, care, and patience for all those long years that I have been away from her. Thank you for taking care of our beloved son while I was finalizing this dissertation.

Chapter 1

Introduction

"The art of war teaches us to rely not on the likelihood of the enemy's not coming, but on our own readiness to receive him; not on the chance of his not attacking, but rather on the fact that we have made our position unassailable."

-The Art of War, Sun Tzu[1]

Over the last three decades, Information and Communication Technologies (ICT) have revolutionized our access to information and how we communicate (see [2],[3], [4]). The Internet, the prime symbol of this digital revolution, is being used across all industries to conduct all kind of transactions. Unfortunately, these progresses in technology also created an increasing number of security concerns that are well illustrated by recent incidents.

1.1 Some security incidents

Very sophisticated Internet viruses* such as Nimda [5], Slammer [6] and Conficker [7] propagated rapidly and infected thousands of machines around the globe and cost millions of dollars in defense, clean up, and rebuilding good reputation [8]. According to the Conficker Working Group [9], between 6 and 13 millions hosts have been infected by different variants of the virus. In a recent blog post, the Cyber Secure Institute [10] claims that the economic loss due Conficker could be as high as $9 billion. More recently, the Stuxnet virus has gained a lot of attention from researchers and media. A Symantec white paper [11] has presented Stuxnet as "*a threat targeting a specific industrial control system such as a gas pipeline or power plant, likely in Iran. The ultimate goal of Stuxnet is to sabotage that facility...*" Stuxnet is one of the first malwares known to specifically target industrial control systems.

*Throughout this thesis, we *casually* use the term "virus" to designate a generic malware. For instance, we do not make the *usual* distinction that is made between a worm and a virus. This distinction is not particularly relevant for our analysis.

Botnets have recently been identified as being among the most important threats to the security of the Internet. A botnet is a network of compromised machines (bots) under the control of an attacker (the botnet herder). In a ten days infiltration into the Torpig botnet network, the study in [12] has gathered information from more than 180000 infected machines around the globe. Botnet herders typically get the bots to act on their behalf to send spams, spread viruses, or carry out phishing attacks. Botnets have also served to launch distributed denial of service attacks (DDoS).

In recent years, DDoS attacks have been launched against business and government websites in incidents that are attributed to organized and terrorist groups, and sometimes to nation-state intelligence agents. The Russia-Estonia conflict [13], the Google-China saga [14], the recent Wikileak "Operation Payback" incident [15], and the many attacks on the White House and other US government agencies' websites [16] are just a few examples. In the Stuxnet case, many bloggers and security specialists have speculated that the virus was designed in Israel to target nuclear power plants in Iran.

This list of security incidents is certainly not exhaustive; [17] gives a broad overview of cyber security incidents in the last three decades.

1.2 Information and communication systems' security challenges

Being able to defend against and survive cyber attacks is *vital*. However, despite the important efforts spent by the many security companies, researchers, and government institutes, information systems' security is still a great concern. Many security challenges still need to be addressed. Some of the most important ones are discussed next.

1.2.1 Strategic and sophisticated adversaries

As information and communication technologies have now become an integral and indispensable part of our daily life, there has been a significant shift in the global threat landscape. Today, the greatest security threats are no longer coming from those *script-kiddies* and *fame-driven* hackers [18] who attack communication systems just to impress their peers. In this information age, security systems have to be designed and implemented to defend against very sophisticated adversaries. These sophisticated adversaries include malicious insiders, cybercriminals, hacktivists, cyberterrorists, industrial spies, and in some cases, nation-state intelligence agents. Such adversaries are very knowledgeable about the communication technologies and protocols. They are highly organized, well-resourced, and capable of operating across the full spectrum of online attacks. Their goals, most of the time, go beyond *minor* security failures like crashing a computer or defacing a webpage. Their motivations are for example to make big monetary gain, cause mass destruction to countries' infrastructure and economy, steal highly classified document and information, and establish long-term intelligence-gathering presence.

In addition of being very knowledgeable about the communication technologies and protocols,

these adversaries *strategically* and *dynamically* choose their targets and attack methods in order to achieve their goals. They skillfully plan their attack after deep investigations of the available technologies and defense techniques. According to this Symantec report [11], Stuxnet was highly targeted, took more than six months of implementation, and was looking for systems using a specific type of network adapter card. Today's hackers also tune their strategies by balancing the effort spent and the chances of success[†] [20]. An insider will try to be off the radar as long as possible while a terrorist is looking for making the maximum damage in one attack. A botnet herder would like to maintain a zombie machine under his control as long as possible. Similarly, an industrial spy will try to dissimulate his presence while stealing valuable information from the victim's information databases.

Defending against such sophisticated attackers is a challenging task. An effective defense solution requires not only high technical skills, but also a good understanding of these attackers motivations, strategies, and capabilities. The war against today's hackers is an *strategic* war and viable cyber security solutions should be implemented with built-in intelligence.

1.2.2 Complexity and interconnectedness

Another fundamental source of difficulty in developing secure information systems is their complexity and interconnected nature. Because of this complexity and interconnectedness, networked information systems are difficult to fully observe and control [21]. On every machine, many automated processes are running without the knowledge of the users. Furthermore, many of these processes involve communications with other computers across networks. As a consequence, full control and observability is impossible, leading to systems that are vulnerable to local as well as remote attacks.

Information and communication systems are interconnected and distributed, and they are wanted to be. Metcalfe's law [22] states that the value of a communication network is proportional to the square of the number of connected users or systems. Interconnected however means interdependence; and interdependence implies that the security of any given network system depends on that of the others. Hence, once a computer is infected by a virus or is under the control of an attacker, additional devices in the same network can become more susceptible.

Related to interconnectedness and distributiveness, is the inherent anonymity offered by networked systems, especially the Internet. The easiness to spoof an email or an IP address and to disguise once identity and location has made it difficult to go after attackers and impossible to prevent attacks; this, despite the advances in forensics [23] and other tracing technologies [24].

Defending today's information systems is to be done in this complex, interconnected and distributed environment. Any successful security solution needs to be designed and implemented by considering those factors.

[†]E.g. according to reports about the recent Wikileak "Operation Payback" incident, hackers abandon attacking Amazon's website because *they did not have the "forces" to bring it down.*[19]

1.2.3 Presence of compromised and/or malicious agents and components

Trustworthiness and full cooperation have been traditionally *assumed* among the different elements responsible of storing, sending, and processing information of communication systems. This was certainly realistic in the days of the ARPANET [25] when all parties involved were *friendly* and *trustful*, and when information was stored in and communicated through devices that were under full control of a few administrators. Many reasons explain why such assumption can no longer be made with today's information systems.

As we mentioned earlier, inside attackers are one of the most important threats today. Insiders have privileged access rights and detailed knowledge about the information system, services, and processes. They participate in the communication protocols as trusted parties and maliciously take advantage of this trust to send bogus messages and corrupt information. They dissimulate their presence while slowly damaging the system and stealing critical information.

Also, with the loss of controllability and observability due to the complexity and interconnectedness of networked information systems, the trustworthiness assumption between the network devices becomes very questionable. In many occasions we have witnessed the network infrastructures (routers, servers) being compromised by malicious adversaries [26].

The potential presence of compromised components explains why the "U.S. government's main code-making and code-cracking agency now works on the assumption that foes may have pierced even the most sensitive national security computer networks under its guard" [27]. Debora Plunkett, head of the NSA's Information Assurance Directorate has put it in these terms: "*We have to build our systems on the assumption that adversaries will get in*". Thus, security mechanisms should be able to ensure the *survivability*[‡] of information systems despite the presence of compromised components.

1.2.4 Human factors!

Security administrators have spent a lot of effort to guarantee and maintain security of information systems. In this, they set security policies to prevent attacks, perform tests to foresee future vulnerabilities, dictate responses to be taken in case of attacks, and lead recovery after security incidents. This has lead to some relatively "good" level of security of networked systems. However, relying on human expertise has some shortcomings. One of such is time scale. Human action is timed in minutes and hours while attacks such as virus infection are carried at a microsecond level. Furthermore, relying on human expertise is also not a scalable solution [21]. In fact, given the distributed nature of communication systems, it is simply not possible to have a "security guard" at each checkpoint. Thus, intelligent automated solutions are needed.

Security administrators also sometimes fail to apply appropriate security measures because of

[‡]Survivability is now considered as a separate field of study. However, in practice, security and survivability have to be implemented together.

incentive misalignment. They set policies and implement security measures according to the risk they have evaluated [28] both for the organization and for their own job position. In some cases the decision involves *some conflict of interest*. For example, managers often buy products and services which they know to be suboptimal or even defective, but which are from big name suppliers, to deflect blame when things go wrong [28].

In sum, attackers and defenders choose their course of action based on some cost-benefit analysis. Attackers choose their targets and tune their attack strategies by balancing the effort spent and the chances of success [20]. System administrators set security policies and implement security measures according to the risk they have evaluated [28] (both for the organization and for their own job position). Users choose their passwords by balancing between the constraints set by the security policies and the ease to remember a password. All those tradeoffs are set in accordance to the incentives of the decision makers (simple users, attackers, and defenders), and this is needed to be taken into account by security solutions.

1.2.5 Risk assessment and Management

The different challenges mentioned above have made the security risk assessment problem a difficult one.

With the presence of sophisticated attackers, one cannot *just* defend against the *most likely* event; which is a solution to the related but different problem of *reliability*. Such a solution is probably efficient against random "bad" events (faults) such as human errors and/or machine failures (reliability problem), but it is ineffective against strategic attackers. Intelligent adversaries will very likely launch attacks that are exceptions to the general rules. As attributed to John Doyle, 'reliability deals with events of small probability; security is about events of probability zero.' Evaluating the risk of such zero likelihood events is quasi-impossible with traditional probability models.

As was mentioned earlier, networked systems are interconnected and distributed, and as a consequence, the security of the different subsystems as well as the network's global security depend on the individual organizations' security. These external factors (or externalities) are difficult to measure and potentially result in under-investment as each organization relies on the investment of the others. This phenomenon is known as *free riding*. Its consequence is poor global network security.

Several other factors have to be considered when evaluating the security risk of information systems. The behavior and incentives of users, the imperfect nature of software and hardware developers, and the potential presence of insiders are determinant factors that are not easy to understand and model.

As a consequence, information systems need to be able to perform well against unanticipated attacks: the "when" and the "how" of the very next attack are always unknown. This makes it difficult to determine how much money and hours of work one should allocate to defending a

system. Yet, when an attack hits an organization, it can cost millions of dollars of defense, cleaning, and eventually rebuilding a good reputation. Viewed from this angle, under-investment, be it in labor time or in money spent, is very undesirable for any information system.

1.3 Security solutions

As can be inferred from the above, the challenges to information and communication security range from the complexity of the underlying hardware and software and the interdependence of the systems, to human, social, and economic factors [21]. Addressing those challenges is *vital* as networked information systems have become an integral part of our daily life. Very aware of that, security researchers and practitioners have developed a variety of defense mechanisms.

1.3.1 Practical security solutions

Physical security and tamperproof techniques have long been used to ensure protection of information and network components. Biometric solutions are being proposed for user (human) authentication [29], and steganography and digital watermarking [30] are used to conceal information within usual mediums.

Cryptography has been used for a long time to provide authentication for both user and data, encryption to protect the confidentiality of information, and signature schemes to guarantee the integrity of information. Cryptographic solutions usually are a combination of symmetric schemes, asymmetric schemes, and hashing methods.

Detection and prevention techniques include Antivirus software, Firewalls, and Intrusion Detection/Prevention Systems (IDS-IPS). They also include solutions such as penetration testing, sandboxing, and honeynets which, in addition to preventing attacks from reaching the system, try to identify the vulnerabilities that attackers are exploiting.

Antivirus softwares perform regular scanning of the storage devices and communication mediums to detect signs of malware presence and subsequently remove them. Firewalls protect the organization's perimeters by inspecting the incoming and/or outgoing traffic operating between an organization network and the outside world, and filtering suspicious packets. Intrusion Detection Systems (IDS) carry out live monitoring of traffic both at network and at host levels in order to detect, in realtime, attacks that are directed against a networked system.

Firewalls, Antiviruses, and IDS are constantly making security decisions. For instance, Antiviruses regularly classify programs as malicious or good, Firewalls decide whether to drop a suspicious packet or not, and IDSs have to determine if a given login attempt is an intrusion. Security administrators also have to make strategic decisions when they set security policies, dictate response actions, or establish security budget. Similarly, attackers have to choose their targets, their attack strategies, and the right moment to launch an attack. Consequently, there is a fun-

damental relationship between network security problems and the decision making of the different agents involved in the security process.

1.3.2 Analytical approaches

Security decisions have recently been investigated analytically in a methodical way. Analytical approaches present a number of advantages compared to heuristic and ad hoc approaches [21]. First, decisions made in an analytical way are well grounded and persistent. A threshold set optimally according to a mathematical model remains optimal as long as the model and parameters do not change. Second, the decision can be generalized and made at large scale, as opposed to heuristic methods that are problem specific. Third, the decision can be numerically implemented which enables them to be run at machine speed. Fourth, decisions made analytically can be checked experimentally and improved upon, enabling the possibility of a feedback loop between the high quality theoretical studies and the real-life problems experienced by security practitioners in a daily basis.

Many mathematical models have been used to model and analyze the decision making problems in security. Decision Theory is the classical mathematical tool to study decision problems. Machine Learning [31], Control Theory [32], and Pattern Recognition [33], are other mathematical models that have been utilized to solve the security decision problem. More recently, Game Theory has been considered to study network security problems.

Among all these methods, Game Theory seems very appealing because, in addition to providing a principled way to understand security problems, game theoretic models capture the adversarial nature of the security problem. Instead of designing a defense against a specific attack, Game Theory attempts to design a defense against a sophisticated attacker who plans in anticipation of a complex defense. As of such, both the defender and attacker's actions can be in principle computed and analyzed. Furthermore, Game Theory can model issues of trust, incentives, and externalities that arise in security systems. Section 1.6 present a review of the basics of Game Theory. In the next section, we discuss some potential usefulness of Game Theory in communication security.

1.4 Game Theory Potentials

Recently, there have been an increased interest to applying game theoretic models to address network security issues, (see the surveys [34],[35]) and some of these approaches look promising.

Game Theory shares many common concerns with the information security problem [36]. In Game Theory, one player's outcome depends not only on his decisions, but also on those of his opponents. Similarly, the success of a security scheme depends not only on the actual defense strategies that have been implemented, but also on the strategic actions taken by the attackers to launch their attacks. It also depends on the actions of the users that are sharing the systems, and on the actions of their peers situated in other networks. All these agents act rationally according

to their various incentives. Game Theory provides means to represent these complex, competitive, and multi-agent interactions into mathematical models that allow a rigorous analysis of the problem [36]. Game Theory also helps the agents predict each other's behavior and suggests a course of action to be taken in any given situation.

Humans intervene (e.g., security administrators) into in the security process, and as it has been discussed earlier this can be slow, it is not scalable, and it is mostly ad-hoc-based. The mathematical abstraction of Game Theory provides a quantitative framework that can generalize and combine heuristic solutions under a single umbrella [21]. Furthermore, Game Theory provides the capability of examining hundreds of scenarios and offers methods for suggesting several potential courses of action with accompanying predicted outcomes [34]. Computer implementations of those methods can result in intelligent and automated decision engines that are fast and scalable. Indeed, computers can efficiently analyze complex combinations and permutations of actions to derive player's strategies. In contrast with computers, humans can handle only a limited level of complexity and tend to overlook possibilities.

As was stated earlier, attackers intelligently choose their targets and alter their attack strategies based on the defensive schemes that are put in place. Traditional security approaches such as Decision Theory fail to capture this fact because they assume that the defender views the opponent's action as *exogenous* [37]. For instance, in a decision theoretic model, the strategy of the attacker (e.g., the probability of attack) is given as an input to the model. In a game theoretic model, both the defense strategies and the hacker's actions are *endogenously* determined. Accordingly, game theoretic models are well suited to model the interaction with dynamic, pro-active, and cognitive adversaries.

Uncertainty is also another challenge that security has to deal with. No one knows when the next vulnerability exploit will arise. Game theory allows to model situations where only partial knowledge about the game is accessible to some players. It helps predict opponent's behavior in such scenarios and dictates actions to be taken. Additionally, dynamic game models account for the learning ability of all players. Using such learning one can build a strong and useful feedback loop between the high quality theoretical studies and the real-life problems experienced by security practitioners in a daily basis.

Trust is yet another important aspect in the design and analysis of secure solutions. Without some level of trust, any security scheme will be very difficult to implement. Building trust relationships and deciding whether to trust received information is particularly relevant in the presence of (potentially) compromised network agents. Moreover, the trust problem can be view as a game. Indeed, an intelligent liar would not tell a lie if he risks to be never believed again. As was said by Samuel Butler "*Any fool can tell the truth, but it requires a man of some sense to know how to lie well.*"[§] Now, when such sophisticated liars are potentially part of an information exchange, how much should one trust received information? By using a game theoretical approach for the trust problem, one can define appropriate trust rules that network agents will follow when exchanging information in adversarial environments [38].

[§]Samuel Butler was an English composer, novelist, and satiric author (1835 - 1902).

1.5 Thesis summary

These many potentials of Game Theory explain the recent surge in applying game theoretical models to the security problem. Books [21]-[39], and many journal, conference (e.g., GameSec, GameNets) and workshop (e.g., the Workshop on the Economics of Information Security (WEIS)) papers are being devoted to this subject.

Going in this same direction, this thesis illustrates the important role that Game Theory can play for information security problems. By modeling the interactions between defenders and attackers as games in communication scenarios, we predict the adversaries' attacks, determine the set of assets that are most likely to be attacked, and suggest defense strategies for the defenders. We study three illustrative examples. We start by a brief discussion of the abstracted models considered in the thesis.

1.5.1 Communication security games

We considered abstracted communication scenarios where an abstract defender (which can be thought of as a network agent e.g. human, process, or a group of agents) is interacting with an abstract attacker, which also can model a human, a process, or a group of those. The interaction happens over the network where the attacker is trying to attack the network assets that the defender aims to protect. We assume that both actors are strategic and choose their actions in anticipation of the opponent's reactions. In all game models, we assume that all the game parameters are known (payoff, cost, network structure).

We put ourself in a world where some known security solutions such as cryptography are inapplicable. For instance in the intruder game model studied in chapter 2, the intruder has full access of the communication channel and can observe and modify all messages. There is no encryption or message authentication process involved. This is not totally unrealistic because an intruder could have access to all encryption and authentication schemes. Today's Wifi is a good example where an intruder can completely take over the security protocol.

In the intelligent virus game model considered in Chapter 3, we study a simplified scenario where the intrusion detection system's only capability is to analyze network traffic and set a detection threshold. The attacker's unique strategy is to choose an infection rate.

In the blocking game model (Chapter 4) we discuss an abstract attack where the adversary "disturbs" the communication by attacking one network asset. This can be seen as jamming a communication link, compromising a server or router, or inserting bogus messages into a protocol.

We do not aim to analyze fully realistic and detailed scenarios. Rather, our goal in this thesis is to illustrate the potential usefulness of Game Theory for the security problem. Section 5.1.1 of the conclusion discusses the issues of applying Game Theory to security. Later in the chapters we argue that, far from being trivial, the simplified models considered here permit us to draw conclusions that would be difficult to draw in a non-game theoretic setting.

1.5.2 Chapter 2: The intruder game

The first game considers a communication scenario where not all network components are trustworthy –because some of them might be compromised. More concretely, we consider the case where a given network agent (node) is receiving some information via another agent (or relay) which might be compromised by an attacker. If the relay is compromised, the goal of the attacker is to corrupt the information in order to deceive the receiver. The receiver (defender) cannot distinguish between a compromised relay and a normal one. Upon receiving the information, the receiver then has to decide how much to trust it.

We model the interaction between the attacker and the defender as a zero-sum game where the attacker chooses whether and how to corrupt the data and the receiver decides how he should trust the received information. By analyzing the Nash equilibrium of the game, we have found that when the chances that the relay is compromised are not small, the receiver should never trust the received message. If those chances are relatively small, the receiver should always trust the received message, despite the fact that the best attacker always corrupts the message. The equilibrium strategies also suggest that, in general, the most aggressive attackers are not always the most dangerous ones.

By introducing a *challenge-response* mechanism, we have shown how the receiver can completely deter the attacker from attacking. The challenge-response is a process under which the receiver has the option to pay a certain cost and use a secure channel to verify the message (and detect the presence of the attacker if there is any). When detected, the attacker incurs some penalty.

1.5.3 Chapter 3: The intelligent virus game

The second game models a scenario where an intelligent virus is attempting to infect a network protected by a simple Intrusion Detection System (IDS). The IDS detects intrusions by analyzing the volume of traffic going inside the network. We model the interaction between the intelligent virus and the IDS as a zero-sum game where the IDS chooses the detection threshold, while the virus selects its infection rate.

Using a Markov chain model, we compute the Nash equilibrium of the game and analyze it. We have found that, not surprisingly, a more aggressive virus is generally potentially more damaging but is also faster to detect. Hence, in its best attack, the intelligent virus always chooses the infection rate by balancing between an aggressive attack that can be easily detected and a slow attack that causes less damage. The best defense strategy against such a sophisticated virus is to quarantine the traffic for analysis purposes prior to letting it go into the network (in addition to setting the optimal threshold). If the defender does not quarantine but let the traffic in while making decision, the best attack is for the virus to be as aggressive as possible. In fact in this case, while the IDS is analyzing, the virus can cause as much damage as possible. The damage can be enormous if the IDS needs human intervention, which operates at a very slow rate compared intelligent automated intervention.

1.5.4 Chapter 4: Blocking games

In our blocking security game models, we solve a class of security problems where a defender selects a set of resources and the attacker attacks one resource. Specifically, we consider that there is a finite set S and two collections $\mathcal{T}$ and $\mathcal{E}$ of subsets of S. The defender selects a subset $T \in \mathcal{T}$ to perform a *mission critical* task. Each subset $T \in \mathcal{T}$ needs some set of resources $e_{T_1}, e_{T_2}, \ldots, e_{T_p} \in \mathcal{E}$ in order to fulfill the task. The attacker, at the same time, tries to disrupt the task by choosing one resource $e \in \mathcal{E}$ to attack. The attacker spends a cost $\boldsymbol{\mu}_e$ to attack $e \in \mathcal{E}$. The net reward of the attacker is $\boldsymbol{\lambda}_{T,e} - \mu_e$ and the loss of the defender is $\boldsymbol{\lambda}_{T,e}$, where $\boldsymbol{\lambda}_{T,e} \geq 0$ is given and, typically, $\boldsymbol{\lambda}_{T,e} = 0$ if $e \notin \mathcal{T}$. This framework applies directly to some communication security problems as discussed below.

We consider mixed strategy Nash equilibria where the defender chooses T with a distribution α on $\mathcal{T}$ and the attacker chooses e with a distribution β on $\mathcal{E}$. The defender wants to minimize the expected loss $E(\boldsymbol{\lambda}_{T,e})$. The attacker, on the other hand, is trying to maximize the expected net attack reward $E(\boldsymbol{\lambda}_{T,e} - \mu_e)$.

We analyze the game by using the combinatorial tools of *blocking pair of matrices* (hence the name blocking security game). We introduce the notion of critical subset of resources and use this notion to define a vulnerability metric for the defender's task. For the non-zero sum game where the attack costs ($\boldsymbol{\mu}_e$) are positive, we show the existence of a set of Nash equilibria under which the attacker will always target critical subsets of resources. When the attack costs are all equal to zero, we characterize all Nash equilibria of the game. In each NE, the attacker will target a particular critical subset of resources. The defender chooses feasible subsets that *minimally* intersect the critical subsets.

We provide two illustrative examples for this model. The first example studies a scenario where a network manager is choosing a spanning tree of a graph while an attacker is trying to cut the tree by attacking one link of the graph. One of our findings in this scenario is that, the usual *edge-connectivity* metric for a graph is not the appropriate vulnerability measure in a network where strategic adversaries are present. The second example deals with a *supply-demand* network where a network manager is choosing a feasible flow to transport the maximum amount of goods from a set of sources to a set of destinations, and an attacker is trying to minimize this by attacking an arc of the network. In this case, we find that critical subsets of links are cutsets that maximize the minimum fraction of goods carried per link of the cutset. In most cases, these correspond to minimum cutsets of the graph.

Computing Nash equilibria of a two-player game is known to be in general complex. In this thesis, we have shown how, for a class of security games, one can compute a critical set of resources (hence a Nash equilibrium) using algorithms that run in polynomial time.

1.5.5 Challenges to applying Game Theory to security

Although Game Theory presents a lot of potentials for security, there are a certain challenges that need to be addressed to make a viable approach to security. The challenges include the complexity of computing an equilibrium strategy and the difficulty to quantify security parameters such as risk, privacy, and trust. Choosing the appropriate game model for a given security problem is another challenge to Game Theory for security. So far, choosing a game is solely based on intuition and there is a lack of data to substantiate those choices. Also, many game models assume full *unbounded rationality* for the players. This assumption contrasts with experimental studies, which have shown that players do not always act rationality. Another challenge that game theorists for security need to address is the interpretation of notions like mixed strategy Nash equilibrium. In fact, even within the Game Theory community, these is no consensus on how to interpret a mixed strategy. These challenges, discussed in more details in the conclusion section 5.1.1, need to be addressed in order to convert game theoretic results into practical security solutions.

1.6 Game Theory basics

This section gives a brief review of the basics of Game Theory. For a more complete presentation, the book by Osborne and Rubinstein [40] and that of Fudenberg and Tirole [41] are classical references. A tutorial on Game Theory is presented in [39, Appx. B] in the context of wireless networks. Section 2 of [34] also gives a brief and very concise overview of Game Theory.

Game Theory models the interaction(s) of decision makers who have to choose actions that might be (but not necessarily) conflicting. The first game theoretic studies appeared in the economics literature and are attributed to Cournot (1838) and Bertrand (1883). John Von Neumann and Oskar Morgenstern ([42], 1928-1944) introduced the general theory of games in their book: *"The Theory of Games and Economic Behavior"*. In particular, they showed the existence of mixed strategy equilibrium for a two-player zero-sum game. John Nash (1950) extended their analysis and proposed the concept of "Nash Equilibrium". The following decade has seen the game theoretic research broadened by the works of Selten (1965) who introduced the notion of "subgame perfect equilibria" and Harsanyi (1967-68) who developed the concepts of "incomplete information" and "Bayesian games". Many studies have followed since then, and most of them focus on the subject of Economics (see [41, Chap. Introduction]). In the late 80's (at the end of the cold war), Game Theory was used for the analysis of nuclear disarmament negotiations [43]. The first application of game theoretic approaches to security was by Burke [36] in the domain of information warfare. Recently, applications of Game Theory to networked information security have driven a lot of interest (see section 1.7).

The basic assumption of a game theoretic model is that decisions makers are *rational* and take into account the *rationality* of other decision makers. In a game, there are typically two or more decision makers called *players*. Players constitute the basic entities of a game. A player can represent a person, a machine or a group of persons. Within a game, players perform *actions* that

they draw from their respective *action sets*. The plan of actions that a given player takes during a game play is called the *strategy* of that player. When all players play their strategies, it leads to an *outcome*, which is the vector of the actions taken by the players in that game play. An outcome gives a *payoff* (i.e. positive or negative reward) to each player. Being rational, each player is trying to choose the strategy that maximizes the received payoff. The payoff of a given player is derived from the *preference* that the player has of some outcome compared to others. A player's preference, in the Von Neumann-Morgenstern sense, is given by a utility function that assigns to each outcome a real number. The more the player prefers an outcome, the higher the assigned number is.
In summary, a game is described by:

- Players $(P_1, \ldots, P_N)$: finite number $(N \geq 2)$ of decision makers.
- Action sets $(A_1, \ldots, A_N)$: player P_i has a nonempty set A_i of actions.
- Payoff functions $u_i : (A_1, \ldots, A_N) \to \mathbb{R},\ i = 1, \ldots, N$: materialize each player's preference, take a possible action profile and assign to it a real number.

If the action sets A_i are finite for all players $i = 1, \ldots, N$, then the game is *finite*. The game is said to be *zero-sum* if the payoff functions are such that $u_1 + u_2 + \cdots + u_N = 0$. It is a *constant-sum* game if the sum is equal to a fixed constant. If the sum can take arbitrary values, the game is *variable-sum*.

Different models of games have been considered, depending on what kind of interaction the players have, how much they know about each other, and how long a game play is. If, within a game, each player is maximizing his own payoff regardless of the results of the others, the game is said to be *non-cooperative*. A *cooperative* game is one in which players can form *coalitions* in order to coordinate their strategies. Each coalition has a given value for the players who form it. The game is then between coalitions rather than between individual players.

When a game is played once and for all, and that all actions are taken simultaneously, then the game is called *static* or *strategic*. If, on the other hand, the game has more than one stage where players can take action, it is an *extensive* game. Extensive games can be *finite* or *infinite*.

The games considered in this thesis are all non-cooperative and static. Also, they are such that each player knows his action set and payoff as well as other players' action sets and payoffs. Such games are defined as *complete information* games. *Incomplete information* games are those in which at least one player is unaware of the possible actions and/or payoffs for at least one other player.

Knowing the action sets does not mean knowing the actual actions taken by the players. When, in a game, each player is aware of the "previous" actions taken by all other players, then the game is of *perfect information*. Otherwise, it is an *imperfect information* game. In this sense, a static game is of imperfect information. Indeed, any static game can be written as an extensive form game where the the players do not observe the actions taken by other players.

A static game can also be of incomplete information when some players are not certain about the action sets and the payoffs of others. Harsanyi [44] has proposed a way to model and understand such games by introducing an *initial* move by "nature" that determines the"types" (action set and payoff) for the players. This initial move is done according some prior probability. Such games are called *Bayesian* because of their inherent probabilistic nature.

Probabilistic moves also characterize models of *stochastic* games. Such games progress as a sequence of states. In each state, players take actions and receive payoffs that depend on the current state of the game, and then the game transitions into a new state with a probability based upon the players' actions and the current state.

Each game model considers a solution concept that determines the outcome of the game and its corresponding strategies. *Nash equilibrium* (NE) is the most famous concept of game solution. In this thesis, we are only concerned with NE. For details about other solution concepts, we refer the interested reader to [40] and [41].

A NE is a set of strategies of the game under which no single player is willing to *unilaterally* change his strategy if the strategies of the other players are kept fix. Formally, the action profile $a^* = (a_1^*, a_2^*, \ldots, a_N^*)$ is a Nash equilibrium if

$$u_i(a^*) \geq u_i(a_i, a_{-i}^*), \quad \text{for all} \quad i = 1, \ldots, N, \quad \text{and all} \quad a_i \in A_i; \tag{1.1}$$

where a_{-i}^* denotes the profile of actions of all players except P_i.
Whenever, for a fixed profile a_{-i}, an action $\tilde{a} \in A_i$ for player P_i verifies $u_i(\tilde{a}, a_{-i}) \geq u_i(a, a_{-i}^*)$ for all other $a \in A_i$, then $\tilde{a}$ is said to be a *best response* to profile a_{-i} for player P_i[¶]. From this, we derive the following equivalent definition for a Nash equilibrium: *an action profile $a^* = (a_1^*, a_2^*, \ldots, a_N^*)$ is a Nash equilibrium if and only if for each player P_i, the action a_i^* is a best response to the action profile a_{-i}^*.*

If the equilibrium strategy designates a particular action for each player, then we refer to it as *pure strategy equilibrium*. Pure strategy equilibria do not always exist. However, for a finite, 2-player game an equilibrium can always be achieved by randomizing the choice of actions on the action sets (John Nash [45]-1951). Such equilibrium is called a *mixed strategy* equilibrium. Formally, a pure strategy is defined as a deterministic choice of action, while a mixed strategy chooses a probability distribution on the set of actions.

In this thesis, because of the competitive aspects of the games we consider, pure strategy equilibrium will, most of the time, not exist. As a consequence, we are particularly interested in mixed strategy Nash equilibria. We typically have two players: a *defender* and an *attacker*. The defender is trying to choose an action to minimize the potential damage that an attacker could cause by taking some attack action. The attack's goal is to maximize the damage. In this sense, we will most of the time be using *zero-sum* games. For zero-sum games, the Nash equilibrium strategies are known to correspond to the *max-min* equilibria (Von Neumann-Morgenstern 1928, John Nash

[¶]Notice that in this definition, we do not not make the difference between action and strategy. However, the *correct* definition considers strategies (action plans) and not actions.

1951). A max-min strategy for player P is a strategy that maximizes the worst outcome that other players could impose to P by their strategy choices. We use this result to compute NE, mostly in the first two game models considered in this thesis.

Another way to find NE is by analyzing best responses. We use this tool to analyze NE in the third game model studied in this thesis. The main result about best response is that in equilibrium, *every action that is assigned a strictly positive probability by a mixed strategy for some player is a best response to the other player's strategy.*

Computing Nash equilibria is in general a difficult problem. For 2-player, finite, zero-sum games, we know that the NE strategies correspond to the max-min equilibrium strategies. Max-min strategies can be solved by casting the problem as a *linear program* [46, Chap. 11]. Linear program can be solved by algorithms that run in polynomial time [46, Chap. 4] in the size of the problem. As a consequence, finite, 2-player, zero-sum games can be solved in polynomial time. For general finite, 2-player games, Papadimitriou [47] has shown that the problem of computing a Nash equilibrium is PPDA-complete ("Polynomial Parity Arguments on Directed graphs"). In this thesis, we will show how to find polynomial-time algorithms for a class of *quasi-zero-sum* games.

1.7 Related work

As was stated earlier, there are currently many activities into applying game theoretic tools to the security problem. The approaches used to study the communication security problem through Game Theory can be classified in two groups (although a lot of papers combine the two). *Socio-Economic* approaches consider the social and economic aspects of security and have studied (among others), the incentives ([48],[20]) of agents involved in the security process (attackers, network administrators, and simple users), their behaviors ([49],[50]), and the economics of information security ([37], [51],[52]). The second approach uses Game Theory to understand the theoretical foundation of the security problem and to derive technical solutions ([53],[54],[55],[56],[57]).
This thesis focuses on this second line of thoughts.

The book by Buttyán and Hubaux [39] considers the application of Game Theory to Wireless Network security problems. That of Alpcan and Basar [21] covers a wide range of security aspects through the lenses of Game Theory. [58] is a compilation of the GameSec2010 conference proceedings. The surveys [34]-[35] give a broad overview of the current status of this "new" field of research. [34] classifies the research activities according to the game models used, while in [35], current research efforts are grouped based on the network security problems that are considered. Some of the research efforts mentioned in those surveys are discussed next.

Intrusion Detection Systems are among the subjects of network security to which Game Theory has been applied the most. This logically follows from the fact that traditional IDSs are based on Decision Theory and, as we have discussed earlier, for the security problem, Game Theory seems to be more appropriate than Decision Theory. In [21], game theoretic approaches to IDS are discussed for various game models and two chapters (9 and 10) are solely devoted to this subject. [35, Sec.

5] devoted an entire section to game theoretic approaches to IDS. Cavusoglu and Raghunathan [57] compare the use of Game Theory and Decision Theory to the detection software configuration problem when the attackers are strategic. Their study has confirmed the conjecture that in general, the game theoretic approach is more appropriate than the decision theoretic approach.

The many game theoretic applications to IDS mentioned in [34] and [35] are mostly different by the type of games considered in the models. In [59], Alpan and Basar study the interaction between a distributed IDS and an attacker and model it as a two-player, non-zero sum, one-shot game. The model is refined in [60] in an extensive form game where a network sensor playing the role of "nature" moves first by choosing a probability of attack (and no-attack), then the IDS and the attacker move simultaneously. In [61], a multi-stage dynamic game model is used to study the intrusion detection problem in a mobile ad hoc network. Zhu and Baser [62] model the configuration problem of policy-based IDS as a stochastic dynamic game. A stochastic game model is also considered by Liu *et al.* [56] for the insider attack problem. Zonouz *et al.* [63] model the interaction between a response engine and an attacker as a Stackelberg game where the engine is the *leader* and the attacker the *follower*. A Bayesian game approach is proposed in [64] to study the intrusion detection problem in wireless ad hoc networks.

As proofs of concept (application of Game Theory to security), we believe that all these works are complementary to our intelligent virus game model. The intruder game treated in chapter 2 of this thesis departs a bit from those models. In our model, we are not interested in detecting the presence of an intruder; rather, we focus our attention on how much trust a receiver should have in a received message that has transited via a potentially compromised relay. In that sense, our intruder game is closer to [65], where the author treats the problem of false dissemination of information in a vehicular network. Our work is also closely related to the attacks on ad hoc network routing protocols discussed in [39, Chap. 7].

Trust and Deception is an important subject that has received a lot of attention in the years. Kevin Mitnick[‖], the infamous hacker and "social engineer" explains in his book, the *Art of Deception*, how to bypass the most efficient firewalls and encryption protocols by fooling users in a network to gather confidential information. The book by Castelfranchi [66] treats the subject of trust and deception in virtual societies.

Availability is one of the security requirements that requires that network resources and assets are available to legitimate users when needed. In an availability (or denial-of-service) attack, the hacker targets a resource that the defender (might) need to perform a certain task. In that sense, our blocking security game model can be interpreted as an availability game. The jamming games presented in ([67]-[68]) are other examples of availability games. In [67], Zander considers the problem of multihop packet radio networks (PRNs) in the presence of an active jammer. The situation is modeled as a two-player constant-sum game where both the jammer and the attacker are subject to an average power constraint. The optimum jamming strategies as well as the best medium

[‖]David Kevin Mitnick was one of the most notorious cyber thieves. Over a 13 year period, he broke into the computer systems of more that 35 international organizations. He was on the FBI's list of most wanted criminals and was arrested in 1995 and sentenced to 5 years prison. Mitnick now runs "Mitnick Security Consulting".

access policies are derived. Kashyap *et al.* [69] consider the jamming problem in a MIMO Gaussian Rayleigh fading channel scenario. The interaction between the jammer and the transmitter-receiver pair is modeled as a zero-sum game where the attacker is trying to minimize the mutual information between the transmitted and received signal, while the defenders are trying to maximize it. Altman *et al.* consider a similar problem where the value of the game is the signal to interference-plus-noise ratio (SINR).

The work in [55] is another denial-of-service attack game where an attacker on the Internet is trying to deface the homepage on a given server. A stochastic game approach is proposed between the network administrator and the attacker where at each time step, both players choose their actions and the game moves to a new state according to some probability that depends on the chosen actions. Through simulations, the authors have shown that the game admits multiple Nash equilibria. Since a NE gives to the defender an idea about the attacker's best strategies, finding more NE means having more information about the attack.

All these games (and others cited in [34] and [35]) head to the same direction as our blocking security game models in the sense that resources that might be (or are) needed by the network are the target of attacks. They consider a variety of game models and different techniques to compute Nash equilibria. In that, we consider them complementary to our work. Our model is however more general, permitting us to cover a broad range of problem. Furthermore, we have analytically characterized the set of Nash equilibria in our models.

Although not dealing with an availability issue, the work presented in [70] is particularly close to the examples of network design we have presented in this thesis. In that paper, the authors consider a non-cooperative, multi-player game on a graph, with two kinds of players, a set of *attackers* and a *protector* player, representing the viruses and the system security software, respectively. Each attacker chooses a node of the graph to infect and the protector chooses a simple path (or edge) to protect. As in our case, the authors relate the mixed Nash equilibria of the game to graph-theoretic notions (edge cover and vertex cover) and suggest polynomial time algorithm to compute Nash equilibria. Our work however departs from their paper in two ways. First, the graph examples presented in this thesis are just applications of a general model to particular cases. Our model covers a broad range of security problems which include the one considered in their paper. Second, in our example the network is interested in choosing links, while in their paper, the network protects nodes.

The Nash equilibria characterization provided in this thesis can be considered as an application of the result in [71] to the particular case of *quasi zero-sum game*. Although Avis *et al.* were not interested in characterizing Nash equilibria (which would be very laborious for an arbitrary two-player matrix game) and did not explicitly consider the notion of blockers, all the ingredients we have used in our NE characterization can be derived from their results. Our use of the combinatorial notion of blocker was the key to our success in characterizing the mixed strategy Nash equilibria of the game. To our knowledge, such a notion was not used before in the context of computing Nash equilibria.

We derive polynomial time algorithms by relating the Nash equilibria computation problem to

the minimization of submodular functions ([72, Chap. 9],[73]). Polynomial time algorithms are known to exist for zero-sum bimatrix games via Convex Optimization methods [74, Chap. 11]. However, computing Nash equilibria for a general two-player matrix game is known to be PPAD-complete* [47]. In that sense, our work extends (although very modestly) the class of games that are known to be solvable in polynomial time.

*Polynomial Parity Arguments on Directed graphs

Chapter 2

Intruder Game

This chapter presents a communication scenario where not all network components are trustworthy– because some of them might be compromised. More concretely, we consider the situation where a given network agent (node) is receiving some information via another agent (or relay) which might be compromised by an attacker. If the relay is compromised, the goal of the attacker is to corrupt the information in order to deceive the receiver. The receiver (defender) cannot distinguish between a compromised relay and a normal one. Upon receiving the information, the receiver then has to decide how much to trust it.

In section 2.1, we model the interaction between the attacker and the defender as a zero-sum game. We study a simple model where the message expected by the receiver is a binary random variable drawn from $\{0, 1\}$ according to a known probability distribution. In this case, the attacker is just trying to figure out whether to flip a received bit of not. When the receiver gets some data, he will decide whether to trust it or not. Indeed, the model holds for non-binary messages. However, characterizing the NE in those cases can easily become overwhelming. Software programs like GAMBIT [75] can help compute Nash equilibria in certain cases.

By analyzing the Nash equilibrium of the game, we have found that, unless the chances that the relay gets compromised are small, the receiver should never trust the received message. If those chances are relatively small, the receiver should always trust the received message, despite the fact that the best attacker will always corrupt the message. The equilibrium strategies also suggest that, in general, the most aggressive attackers are not always the most dangerous ones.

In section 2.3, we introduce a *challenge-response* mechanism. The challenge-response is a process under which the receiver has the option to pay a certain cost and use a secure channel to verify the message (and detect the presence of the attacker if there is any). When detected, the attacker incurs some penalty. The analysis of the NE in this case shows that, by using a challenge, the receiver can completely deter the attacker from attacking.

2.1 Simple Intruder Game

We consider the communication model depicted in figure 2.1.
We use names in the security jargon (Alice, Trudy, Bob), to designate the agents involved in our security games. These agents can, however, be any network component capable of storing, transferring, or processing information.

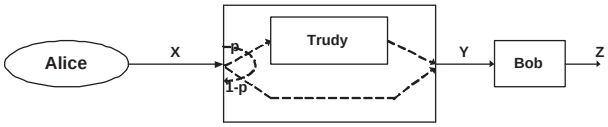


Figure 2.1: Intruder game model. Alice sends a message to Bob. The message might transit via an attacker: Trudy. The game is between Bob who wants to correctly detect the message and Trudy who wants to corrupt it.

Alice sends a message X to Bob through a channel. The channel is *insecure* in the sense that an intruder, Trudy, might be present and is able to corrupt the message. This captures the cases where a malicious user is present in a communication system or that communication components can be compromised by malicious users.

The message X is considered to be a binary random variable ($X \in \{0,1\}$) drawn from the set $\{0,1\}$ according to a probability distribution $Pr[X = x] = \pi(x), x \in \{0,1\}$ that is assumed to be known to Bob and Trudy*. We let Y designate the message that Bob receives.

Trudy is present with a probability p known to both Trudy and Bob. This can be interpreted as if Trudy has only p chance of getting the message or that a fraction equal to p of the network components is corrupted (either by a malicious user, or by being themselves malicious).

When Trudy is not present, Bob correctly receives the message sent by Alice $Y = X$. When Trudy is present, we model her possibly randomized strategy that modifies X into Y by the probabilities $P(x,y) = Pr[Y = y \mid X = x]$ for $x, y \in \{0,1\}$. P is hence a 2×2 stochastic matrix.

Upon receiving a message Y from the channel, Bob guesses that Alice sent the message Z. We model Bob's decision by the probabilities $Q(y,z) = Pr[Z = z \mid Y = y]$ for $y, z \in \{0,1\}$. Let $C(x,z)$ be the cost that Bob faces when he decides $Z = z$ when $X = x$, for $x, z \in \{0,1\}$. For given choices of the probabilities P and Q, the expected cost $J(P,Q) = E(C(X,Z))$ can be calculated as follows:

$$J(P,Q) = \sum_{x,y,z} \pi(x) R(x,y,z) C(x,z) \tag{2.1}$$

where R is the matrix given by

$$R(x,y,z) = [(1-p)I + pP(x,y)]\, Q(y,z), \tag{2.2}$$

*Alice is not part of the game

where I is the 2×2 identity matrix. The term $(1-p)I + pP(x,y)$ in this expression means that with probability $1-p$ the message is faithfully received, and with probability p, it gets modified by the intruder (according to the stochastic matrix $P(x,y)$).

Trudy's goal is to choose P to maximize $J(P,Q)$ whereas Bob's goal is to choose Q to minimize $J(P,Q)$. This scenario is modeled by a Bayesian zero-sum game ([40],[41]).

2.2 Nash Equilibrium

Recall that in a two-player zero-sum game, the Nash equilibrium corresponds to the max-min equilibrium ([40, Chap.2]) and is a pair (P^{NE}, Q^{NE}) such that

$$J(P^{NE}, Q^{NE}) = \max_P \min_Q J(P,Q) = \min_Q \max_P J(P,Q). \tag{2.3}$$

That is, neither Alice nor Bob can gain by deviating from those choices.

To analyze the Nash equilibria of the game, we assume that $C(0,0) = C(1,1) = 0$, which means that Bob does not incur any cost when he correctly guesses Alice's message. Also, we assume, without loss of generality, that $\pi(0)C(0,1) \leq \pi(1)C(1,0)$. (We could interchange the roles of 0 and 1 if the inequality does not hold.)

Theorem 1 *The Nash equilibria, as a function of the parameter p, are shown in figure 2.2, where* $\mathcal{P}_1$ *is a set of stochastic matrices* P *such that*

$$p\left(\pi(0)P(0,1)C(0,1) + \pi(1)P(1,0)C(1,0)\right) \geq \pi(0)C(0,1). \tag{2.4}$$

$\mathcal{P}_2$ *is a set of stochastic matrices* P *such that the previous inequality holds and, moreover,*

$$p\left(\pi(0)P(0,1)C(0,1) + \pi(1)P(1,0)C(1,0)\right) \leq \pi(1)C(1,0). \tag{2.5}$$

The values r_1 *and* r_2 *are defined by*

$$r_1 = 1 - r_2 = \frac{\pi(0)C(0,1)}{\pi(0)C(0,1) + \pi(1)C(1,0)}. \tag{2.6}$$

2.2.1 Understanding the NE

The meaning of this result is as follows.
When the probability p that Trudy is present is smaller than r_1, Bob should always trust the received message (and choose $Z = Y$). The corresponding strategy for Trudy is to always corrupt the message as $Y = 1 - X$ whenever she is present. This might seem counter-intuitive (for Bob) but it is not. In fact, Bob does not know when Trudy is present or not, but he knows that Trudy

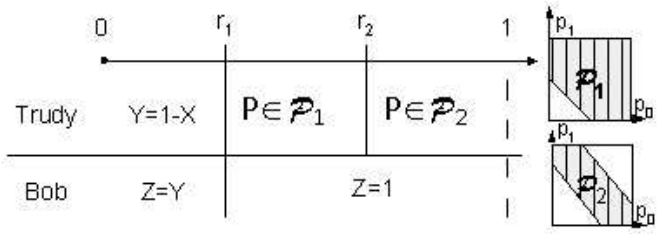


Figure 2.2: The Nash Equilibria of the Intruder Game. The figures in the right show the regions $\mathcal{P}_1$ and $\mathcal{P}_2$.

is absent most of the time (p small). Thus, although Trudy is always corrupting the message, the risk of trusting the channel is small for Bob (think of $p \approx 0$).

When p is larger than r_1, Bob should always ignore the message Y because it is "too" untrustworthy. For this strategy to hold as NE strategy for Bob, Trudy should corrupt the message with *appropriate probabilities*, which corresponds to the conditions (2.4)-(2.5) on P. These conditions are represented by the regions $\mathcal{P}_1$ and $\mathcal{P}_2$ of figure 2.2, respectively.

To understand (2.4), note that the left-hand side is Bob's expected cost if he trusts the received message and chooses $Z = Y$. The right-hand side is his cost if he (ignores the message and) chooses $Z = 1$. Also, we will show later that for $r_1 \leq p \leq r_2$, choosing $Z = 0$ will always be suboptimal for Bob, independently of Y. Thus, the condition means that Bob should chooses $Z = 1$.

Similarly, the right-hand side of (2.5) is Bob's expected cost if he chooses $Z = 0$. That condition implies that Bob should not choose $Z = 0$ and the first condition implies that he should not choose $Z = Y$. Hence, Bob's best strategy is $Z = 1$ independently of the received message Y. The upper bound $\pi(1)C(1,0)$ imposes some maximum values for $P(0,1)$ and $P(1,0)$. This implies that Trudy cannot be too aggressive. To see this, assume that Trudy is so aggressive that condition (2.5) is violated. Then, consider a strategy where Bob always chooses $Z = 1 - Y$. The corresponding cost is equal to $Pr[X = 0, Y = 0]C(0,1) + Pr[X = 1, Y = 1]C(1,0)$, which, by extending the terms, gives

$$\pi(0)C(0,1) + \pi(1)C(1,0) - p\left(\pi(0)P(0,1)C(0,1) + \pi(1)P(1,0)C(1,0)\right). \tag{2.7}$$

This is less than $\pi(0)C(0,1)$ if condition (2.5) is violated. Thus, Bob can obtain a better payoff by changing his strategy from $Z = 1$ to $Z = 1 - Y$. As a consequence, for Trudy to achieve the highest possible payoff, she must limit her aggressiveness to satisfy (2.5).

2.2.2 Proof of theorem 1

To simplify notation, we let $C(0,1) = A$, $C(1,0) = B$, $P(i,j) = 1 - P(i,i) = p_i$, and $Q(i,j) = 1 - Q(i,i) = q_i$ for $i \neq j \in \{0,1\}$ (which are the probabilities of flipping).

With these new notations, the expected cost can be written as:

$$\begin{aligned} J(P,Q) &= \pi(0)\left((1-pp_0)q_0 + (1-q_1)pp_0\right)A + \pi(1)\left((1-pp_1)q_1 + (1-q_0)pp_1\right)B \quad (2.8)\\ &= q_0\left(\pi(0)A - T\right) + q_1\left(\pi(1)B - T\right) + T \quad (2.9)\end{aligned}$$

where identity (2.9) is obtained by rearranging the terms in the RHS of (2.8) and T is defined as $T = p\left[\pi(0)p_0A + \pi(1)p_1B\right]$.

First, notice that Bob can always achieve a cost of $\pi(0)A \leq \pi(1)B$ by completely ignoring the message and choosing $Z = 1$ (this is equivalent to $q_0 = 1$ and $q_1 = 0$–always flip a 0 and always accept a 1).

Now, minimizing the expression (2.9) over Q, we see that Bob's best response to any strategy P is to choose (q_0, q_1) such that $q_0 = q_1 = 0$ if $T < \pi(0)A$; $q_0 = 1 - q_1 = 1$ if $\pi(0)A < T < \pi(1)B$; and $q_0 = q_1 = 1$ if $T > \pi(1)B$. The corresponding costs are T, $\pi(0)A$, and $\pi(0)A + \pi(1)B - T$ respectively for $T < \pi(0)A$, $\pi(0)A < T < \pi(1)B$, and $T > \pi(1)B$. If Q is such that $T = \pi(0)A$ (resp. $T = \pi(1)B$), Bob is indifferent and can choose any $0 \leq q_0 \leq 1$ (resp. $0 \leq q_1 \leq 1$). Thus, we find that

$$\min_Q J(P,Q) = \begin{cases} T, & \text{if } T < \pi(0)A \\ \pi(0)A, & \text{if } \pi(0)A \leq T \leq \pi(1)B \\ \pi(0)A + \pi(1)B - T, & \text{if } T > \pi(1)B. \end{cases} \quad (2.10)$$

Trudy's best strategy is obtained by maximizing $\min_Q J(P,Q)$ with respect to P. For that, we let p vary between 0 and 1, and examine how the best responses are changing.

- If $0 \leq p < r_1$, then for all values of (p_0, p_1), $T < \pi(0)A$. We have already shown that Bob's best response in this case is $Z = Y$ (i.e. $q_0 = q_1 = 0$). Trudy's best strategy is obtained by maximizing the corresponding cost (T). The optimal value is achieved at $p_0 = p_1 = 1$, so that $Y = 1 - X$. Thus, for $0 \leq p < r_1$, the strategies given in the theorem are best responses to each other. As a consequence, they form a NE.

- If $r_1 \leq p \leq r_2$, then for some values of (p_0, p_1), T is less then $\pi(0)A$ and for others, $T \geq \pi(0)A$. Notice that the maximum value of T is $\pi(1)B$. If (p_0, p_1) is such that $T \leq \pi(0)A$, then the corresponding cost is $T \leq \pi(0)A$ (see (2.10) second row). If, on the other hand $\pi(0)A < T \leq \pi(1)B$, then the cost is $\pi(0)A$. This second strategy dominates the first, thus, the best strategy for Trudy is to choose (p_0, p_1) such that $\pi(0)A < T \leq \pi(1)B$. Bob's best bet in this case is to choose $Z = 1$. Consequently, any $P \in \mathcal{P}_1$ and $Z = 1$ is a Nash equilibrium.

- If $r_2 < p \leq 1$, then, depending on the choice of (p_0, p_1), one can have $T \leq \pi(0)A$, $\pi(0)A < T \leq \pi(1)B$, or $\pi(1)B < T$. A similar reasoning as in the previous case shows that if (p_0, p_1) are chosen such that $\pi(0)A \leq T \leq \pi(1)B$, then Bob's best response is $q_0 = 1 - q_1 = 1$. This leads to a payoff of $\pi(0)A$. If Trudy chooses (p_0, p_1) such that $\pi(1)B < T$, we have seen in our discussion above (2.7), that Bob can choose a strategy ($Z = 1 - Y$) to achieve a payoff that is less than $\pi(0)A$. Thus, the best bet for Trudy is to make her choice satisfy $\pi(0)A \leq T \leq \pi(1)B$. This implies that any $P \in \mathcal{P}_2$ and $Z = 1$ is a Nash equilibrium.

Notice that at the boundary $T = \pi(0)A$, Bob can choose any value for q_0, but $q_1 = 0$. Similarly, when $T = \pi(0)A$, Bob can choose any value for q_1, but $q_0 = 1$. All these choices yield to the same cost of $\pi(0)A$ for Bob.

2.3 Challenging the Message

In this section we introduce a challenge-response mechanism for the receiver (Bob). The challenge-response is a process under which Bob has the additional choice of challenging the received message by paying a fixed cost of V and getting an acknowledgement from Alice (see figure 2.3). We assume that Trudy cannot corrupt the challenge communication. For example, in the case of a phishing

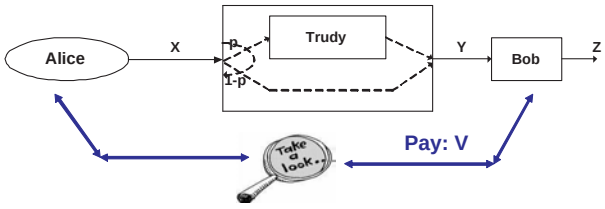


Figure 2.3: The Intruder Game with Challenge.

email that pretends to come from a bank, this challenge can be interpreted as the recipient of the email making the additional effort of taking the phone and calling the bank to verify the authenticity of the email. In general, the challenge can be thought of as Bob being able to use a *secure* side channel to verify the message with the source (Alice) and detect the presence of the attacker if there is any. When detected, the attacker incurs some penalty.

As in the previous sections, we model the problem as a Bayesian game and study the set of Nash equilibria. We consider the same setting as in the previous section with the following additional features:

1. Whenever Bob receives a message, he decides whether to challenge it or not. Challenging requires the payment of a fixed cost of V. We let $\alpha(y)$ be probability that Bob challenges a message $Y = y$, for $y = 0, 1$.

2. If he decides not to challenge the message, then Bob is back to the decision problem considered in the previous section: should he trust the message or not? In that case, we model his strategy with the matrix Q.

3. If the challenged message turns out to be corrupted, then the intruder (Trudy) is punished by having to pay a fine V [†].

[†]Typically the punishment is larger than the challenge cost. We have considered these costs to be equal for simplicity.

4. If Bob challenges a received message and finds out that it was correct (i.e. was not changed), then he looses the challenge cost of V.

5. As before the strategies for Trudy is to decide whether to corrupt the message before relaying it. We model her strategy with the matrix P, as before.

For this game, Bob pays $C(X,Z)\mathbf{1}\{\text{no challenge}\}+V\mathbf{1}\{X=Y \text{ and challenge}\}$. Accordingly, the expected value $J(P,Q,\alpha)$ of Bob's cost is given by the following expression:

$$\begin{aligned} J(P,Q,\alpha) &= \sum_{x,y,z}\pi(x)C(x,z)R(x,y,z)(1-\alpha(y)) \\ &\quad + V\sum_{x,y}\pi(x)\left((1-p)I(x,y)+pP(x,y)\right)\alpha(y)Pr\{X=y|Y=y\}. \quad (2.11) \end{aligned}$$

Bob is trying to minimize this value by his choice of α_i and q_i for $i=0,1$. Trudy's objective is to choose $p_i,\ i=0,1$ to maximize the expected value of her reward $C(X,Z)\mathbf{1}\{\text{no challenge}\}-V\mathbf{1}\{X\neq Y \text{ and challenge}\}$. This expected reward $K(P,Q,\alpha)$ is as follows:

$$\begin{aligned} K(P,Q,\alpha) &= \sum_{x,y,z}\pi(x)C(x,z)R(x,y,z)(1-\alpha(y)) \\ &\quad - V\sum_{x,y}\pi(x)\left((1-p)I(x,y)+pP(x,y)\right)\alpha(y)Pr\{X\neq y|Y=y\}. \quad (2.12) \end{aligned}$$

To analyze the Nash equilibria, we assume, as before, that $\pi(0)C(0,1)\leq\pi(1)C(1,0)$. Furthermore, in this section we assume that $A=C(0,1)<B=C(1,0)$. We also assume that $\frac{\pi(1)B}{\pi(0)}\geq\frac{\pi(0)A}{\pi(1)}$. These assumptions have a *slight* effect in the result presented below. More precisely, changing these assumptions will slightly change the decision regions. However, the same approach can be carried to compute the Nash equilibrium for the other cases.

Theorem 2 *Figure 2.4 shows the different decision regions for the Nash equilibrium, and Table 2 shows the corresponding strategies for Bob and Trudy.*

In the table , $\beta(0)=\frac{V(\pi(1)B-\pi(0)V)}{p\pi(0)(AB-V^2)}$, $\beta(1)=\frac{V(\pi(0)A-\pi(1)V)}{p\pi(1)(AB-V^2)}$, *and* $\gamma=\frac{(1-p)\pi(0)V}{p\pi(1)B}$.

A proof of the theorem is provided in section 2.3.2

2.3.1 Understanding the NE

The meaning of this result is as follows.
The region $R0$ corresponds to the case when $V>\max(A,B)=B$. In this case the challenge cost is too high, and the receiver is better off to never use it. Instead, Bob will optimally decode as if there were no challenge possibility. This gives the same result as in section 2.2.

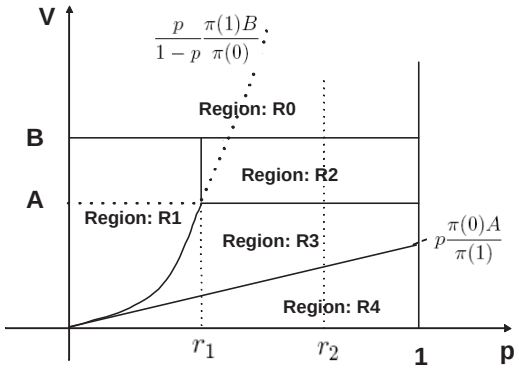


Figure 2.4: The Nash equilibria decision regions of the Intruder Game with challenge.

Region	*Bob*	*Trudy*
R0	As in previous section	
R1	$\alpha(0) = \alpha(1) = 0$ $Z = Y$	P = I
R2	$\alpha(0) = \alpha(1) = 0$ $Z = 1$	$P \in \mathcal{P}_3$
R3	$\alpha(0) = B/(V + B)$ $\alpha(1) = 0$ $Z = Y$	$P(0,1) = 1,\ P(1,0) = \gamma$
R4	$\alpha(0) = B/(V + B)$ $\alpha(1) = A/(V + A)$ $Z = Y$	$P(0,1) = \beta(0), P(1,0) = \beta(1)$

Table 2.1: Nash equilibria strategies for the Intruder Game with challenge.

In $R2$ the challenge cost is high but not high enough to be totally ignored as in region $R0$. Here, the challenge plays the role of a *threat* and obliges the attacker to reduce her aggressiveness compared to $R0$. Hence, the intruder has less degrees of freedom compared to the previous section: the optimal strategy set is now $\mathcal{P}_3$ which is $\mathcal{P}_1$ (resp. $\mathcal{P}_2$ depending on the value of p) of the previous section with the additional requirement that $\pi(0)V \geq p(P(0,1)\pi(0)V + P(1,0)\pi(1)B)$. The optimal strategy for Bob is then to not use the challenge and ignore the message.

Region $R1$ is the set of V that satisfies $V > \frac{p}{1-p}\frac{\pi(1)B}{\pi(0)}$ (or $(1-p)\pi(0)V > p\pi(1)B$). Notice that such V also satisfies $V > \frac{p}{1-p}\frac{\pi(0)A}{\pi(1)}$ (or $(1-p)\pi(1)V > p\pi(0)A$) because of the assumption $\pi(0)A \leq \pi(1)B$. The term $(1-p)\pi(0)V$ (resp. $(1-p)\pi(1)V$) is the cost of challenging when a 0 (resp. 1) is sent and there is no intruder, while the term $p\pi(1)B$ (resp. $p\pi(0)A$) represents the average cost of accepting a 1 while a 0 (resp. 1) was sent and the intruder is present. The inequalities above tell that, for Bob, in either case, the expected cost of challenging a message is higher than the that of accepting it. Thus, Bob is better off to not use the challenge ($\alpha(0) = \alpha(1) = 0$) and

to optimally decode as if there were no challenge. Since in this region $p \leq r_1$, the best decoding strategy is $Z = Y$ (always accepting the message as in the previous section) and Trudy will always change the message. As was discussed in the previous case, Bob can trust the message because most of the time Trudy is absent and the challenge is too high to be used. The average cost for Bob is equal to $p(\pi(0)A + \pi(1)B) \leq \pi(0)A$, which is equal to the expected reward that Trudy receives.

$R3$ can be described as $(1-p)\pi(0)V < p\pi(1)B$ and $\pi(1)V > p\pi(0)A$. In this region, the risk of challenging a 0 ($(1-p)\pi(0)V$) if Trudy is not present, is less than the cost of wrongly accepting a 1 ($p\pi(1)B$) if she is present and happens to flip the bit. Thus, Bob is better off challenging a received $Y = 0$. However, since the intruder is present only a fraction of time, challenge should also occur a fraction of time ($\alpha(0) = \frac{B}{B+V}$). When Bob does not challenge a received $Y = 0$, he will trust it ($Z = Y$). With the assumption $\pi(0)A \leq \pi(1)B$, the expected cost of accepting a 1 is always less than the cost of not accepting it. And, since the challenge cost is relatively high compared to $p\pi(0)A$, Bob is better off to never challenge 1 and to always accept it ($\alpha(1) = 0$ and $Z = Y$ when $Y = 1$). The corresponding best strategy for the intruder is to always flip a 0, and to flip a 1 only a fraction of time γ. The average cost for Bob is $\pi(0)(pA + (1-p)V) < \pi(0)A$, while the expected net attack reward received by Trudy is $p\pi(0)A$.

In region $R4$ the challenge risk is small enough for Bob to challenge both messages. As a consequence, Trudy will flip only a fraction of time ($\beta(0), \beta(1)$). Interestingly, in this region, the average cost for Bob is $\frac{AB-(\pi(0)A+\pi(1)B)}{AB-V^2}V < V$[‡] while the average reward for Truder is equal to zero; i.e the intruder has no incentive to attack. Thus, by using the challenge (as a *credible threat*), Bob can deter Trudy from attacking the communication.

Compared to the previous case, we have seen that the challenge gives to the receiver the possibility to always trust the channel ($Y = Z$) when $V \leq \min\{C(0,1), C(1,0)\}$ without having to pay the worst case cost of $\pi(0)A$. Furthermore, with relatively cheap challenge, the receiver can deter the intruder from attacking. This tells that with a simple challenge-response scheme, one can implement a *perfect*[§] communication over an insecure channel.

2.3.2 Proof of theorem 2

As in the proof of the theorem in the previous case, we simplify the notation by letting $C(0,1) = A,\ C(1,0) = B,\ P(i,j) = 1 - P(i,i) = p_i$, and $Q(i,j) = 1 - Q(i,i) = q_i$ for $i \neq j \in \{0,1\}$.

By extending and arranging the terms in equation (2.11) (resp. (2.12)), we can rewrite the expected cost (resp. reward) of the receiver (resp. attacker) as

$$\begin{aligned} J(P,Q,\alpha) &= T + \alpha(0)\left(\pi(0)\left(1 - pp_0\right)V - \pi(1)pp_1B\right) + (1-\alpha(0))q_0\left(\pi(0)A - T\right) \\ &\quad + \alpha(1)\left(\pi(1)\left(1 - pp_1\right)V - \pi(0)pp_0A\right) + (1-\alpha(1))q_1\left(\pi(1)B - T\right), \quad (2.13) \end{aligned}$$

[‡]Note that as p goes to zero, V also goes to zero and so does the attack cost for Bob.

[§]This requires the channel used for the challenge to be secure itself. Also, the meaning given to *perfect* is that the attacker does not have an incentive to attack.

and

$$\begin{aligned} K(P,Q,\alpha) &= p_0\left(((1-\alpha(1))(1-q_1)-(1-\alpha(0))q_0\right)A-\alpha(1)V)\,p\pi(0) \\ &\quad +p_1\left(((1-\alpha(0))(1-q_0)-(1-\alpha(1))q_1\right)B-\alpha(0)V)\,p\pi(1) \\ &\quad +(1-\alpha(0))q_0\pi(0)A+(1-\alpha(1))q_1\pi(1)B. \end{aligned} \tag{2.14}$$

In the equations above, T is defined as $T = p(\pi(0)p_0A + \pi(1)p_1B)$.

Now, we show that for the different regions shown in Figure 2.4, the strategies given in Table 2 are best responses to each other.

- If $V \le p\frac{\pi(0)}{\pi(1)}A$, then we want to show that $\alpha(0) = \frac{B}{B+V}$, $\alpha(1) = \frac{A}{A+V}$, and $q_0 = q_1 = 0$ are in Nash equilibrium with $p_0 = \frac{V(\pi(1)B-\pi(0)V)}{p\pi(0)(AB-V^2)}$ and $p_1 = \frac{V(\pi(0)A-\pi(1)V)}{p\pi(1)(AB-V^2)}$. To see this, we replace $\alpha(0)$, $\alpha(1)$, q_0, and q_1 by their values in (2.14). We observe that

$$\begin{aligned} \left((1-\alpha(1))(1-q_1)-(1-\alpha(0))q_0\right)A-\alpha(1)V &= \left(1-\frac{A}{A+V}\right)A-\frac{A}{A+V}V \\ &= \frac{V}{A+V}A-\frac{A}{A+V}V \\ &= 0. \end{aligned}$$

  This implies that Trudy cannot improve her payoff by unilaterally changing p_0. Similarly, we have that

$$\left((1-\alpha(0))(1-q_0)-(1-\alpha(1))q_1\right)B-\alpha(0)V = 0,$$

  and hence changing p_1 does also not improve Trudy's payoff. Thus, the pair (p_0, p_1) given above is a best response to Bob's strategy.

  On the other hand, if we replace p_0 and p_1 by their values in (2.13), we observe that

$$\begin{aligned} \pi(0)\left(1-pp_0\right)V-\pi(1)pp_1B &= \pi(0)\left(1-p\frac{V\left(\pi(1)B-\pi(0)V\right)}{p\pi(0)(AB-V^2)}\right)V \\ &\qquad -\pi(1)p\frac{V\left(\pi(0)A-\pi(1)V\right)}{p\pi(1)(AB-V^2)}B \\ &= \frac{1}{AB-V^2}\left(\pi(0)(AB-V^2)V-\left(V^2\left(\pi(1)B-\pi(0)V\right)\right)\right. \\ &\qquad \left.-\left(BV\left(\pi(0)A-\pi(1)V\right)\right)\right) \\ &= \frac{1}{AB-V^2}\left(\pi(0)ABV-\pi(0)V^3-\pi(1)BV^2\right. \\ &\qquad \left.+\pi(0)V^3-\pi(0)ABV+\pi(1)BV^2\right) \\ &= 0. \end{aligned}$$

  Similarly, one can show that

$$\pi(1)\left(1-pp_1\right)V-\pi(0)pp_0A = 0.$$

Thus, Bob does not gain more by changing $\alpha(0)$ or $\alpha(1)$. To complete the proof of this case, we need to show that the choices $q_0 = q_1 = 0$ are also optimal for the the receiver. For that, it suffices to show that the term $\pi(0)A - T$ is nonnegative. Indeed, since Bob is trying to minimize the expected cost (2.13), the best value for q_0 is 0 whenever $\pi(0)A - T \geq 0$. Since $\pi(1)B \geq \pi(0)A$, this will also mean that $\pi(1)B - T \geq 0$. By expending the term $\pi(0)A - T$, we get that

$$\begin{aligned}
\pi(0)A - T &= = \pi(0)A - p\left(\pi(0)p_0 A + \pi(1)p_1 B\right) \\
&= \pi(0)A - p\left(\pi(0)\frac{V\left(\pi(1)B - \pi(0)V\right)}{p\pi(0)(AB - V^2)}A + \pi(1)\frac{V\left(\pi(0)A - \pi(1)V\right)}{p\pi(1)(AB - V^2)}B\right) \\
&= \frac{1}{AB - V^2}\big(\pi(0)A(AB - V^2) - A\left(V\left(\pi(1)B - \pi(0)V\right)\right) \\
&\qquad\qquad -B\left(V\left(\pi(0)A - \pi(1)V\right)\right)\big) \\
&= \frac{1}{AB - V^2}\big(\pi(0)A^2B - (\pi(0) + \pi(1))ABV + \pi(1)BV^2 \\
&\qquad\qquad -\pi(0)AV^2 + \pi(0)AV^2\big) \\
&= \frac{1}{AB - V^2}\big(\pi(0)A^2B - (\pi(0) + \pi(1))ABV + \pi(1)BV^2\big) \\
&= \frac{B}{AB - V^2}\left(\pi(0)A(A - V) - \pi(1)V(A - V)\right) \\
&= \frac{B(A - V)}{AB - V^2}\left(\pi(0)A - \pi(1)V\right).
\end{aligned}$$

Since $V \leq p\frac{\pi(0)}{\pi(1)}A$ for all p, we have that $(\pi(0)A - \pi(1)V) \geq 0$ (by considering the particular case of $p = 1$). Thus, the RHS in the last equality above is greater than 0 if $A - V \geq 0$, which is what we needed.

The last step to finalize the proof in this case is to show that the value p_0 and p_1 given above are probability distributions. It is readily verifiable that they are both greater than zero. To verify that p_0 is less than one, we write

$$p\pi(0)(AB - V^2) - V\left(\pi(1)B - \pi(0)V\right) = B\left(p\pi(0)A - \pi(1)V\right) + (1 - p)\pi(0)V^2,$$

which is greater than zero if $V \leq p\frac{\pi(0)}{\pi(1)}A$. Thus, $p\pi(0)(AB - V^2) \geq V\left(\pi(1)B - \pi(0)V\right)$, and as a consequence $p_0 \leq 1$. Similarly,

$$p\pi(1)(AB - V^2) - V\left(\pi(0)A - \pi(1)V\right) = A\left(p\pi(1)B - \pi(0)V\right) + (1 - p)\pi(1)V^2,$$

which is greater than zero because of the assumption $\frac{\pi(1)B}{\pi(0)} \geq \frac{\pi(0)A}{\pi(1)}$. Thus $p_1 \leq 1$.

- If $p\frac{\pi(0)}{\pi(1)}A \leq V \leq \frac{p}{1-p}\frac{\pi(1)}{\pi(0)}B$, we need to show that $\alpha(0) = \frac{B}{B+V}$, $\alpha(1) = 0$, and $q_0 = q_1 = 0$ are in Nash equilibrium with $p_0 = 1$ and $p_1 = \frac{\pi(0)(1-p)V}{p\pi(1)B}$. As in the previous case, we verify that they are best strategies to each other by showing that none of the players will gain from a unilateral deviation. Fixing $(\alpha(0), \alpha(1))$ and (q_0, q_1), we verify that the term that multiplies

p_0 in (2.14) is equal to $p\pi(0)A$. Thus, $p_0 = 1$ is a best response of the intruder. The term multiplying p_1 is equal to zero, so that the intruder does not make any gain by changing p_1. On the other hand, by fixing p_0 and p_1, we verify that the term multiplying $\alpha(0)$ in (2.13) is equal to zero, implying that a change in $\alpha(0)$ does not improve Bob's payoff. By using the values of $p_i,\ i = 0, 1$, we can write

$$\pi(0)A - T = \pi(0)(1-p)(A-V),$$

which is nonnegative for $A \geq V$, justifying the choice $q_0 = 0$ for Bob. As a consequence of this, we also have that $\pi(1)B - T \geq 0$, implying that $q_1 = 0$ is also a best response. Finally we need to show that the term that multiply α_1 in (2.13) is nonnegative (so as $\alpha(1) = 0$ is optimal). For that, we write

$$\begin{aligned} \pi(1)\left(1-pp_1\right)V - \pi(0)pp_0A &= \frac{1}{B}\left(\pi(1)BV - \pi(0)\left((1-p)V^2 + pAB\right)\right) \\ &\geq \frac{1}{B}\left(\pi(1)BV - \pi(0)AB\right) \qquad (2.15) \\ &= \pi(1)V - \pi(0)A \\ &\geq 0, \end{aligned}$$

where (2.15) is obtained by observing that $(1-p)V^2 + pAB$ is at most AB. The final step of the proof of this case is to show that $p_1 \leq 1$. This follows from the assumption $V \leq \frac{p}{1-p}\frac{\pi(1)}{\pi(0)}B$.

- If $V \geq \frac{p}{1-p}\frac{\pi(1)}{\pi(0)}B$, we first observe that $\alpha(0) = \alpha(1) = 0$ is always optimal for Bob. Indeed, in this case, the minimum value of $\pi(0)\left(1-pp_0\right)V$ (i.e. $\pi(0)\left(1-p\right)V$) is always larger that the maximum value of $\pi(1)pp_1B$ (i.e. $\pi(1)pB$). Thus, $\alpha(0) = 0$ is optimal for Bob. Similarly, the minimum value of $\pi(1)\left(1-pp_1\right)V$ (i.e. $\pi(1)\left(1-p\right)V$) is always larger that the maximum value of $\pi(0)pp_0A$ (i.e. $\pi(1)pA$), thus, the best $\alpha(1)$, is $\alpha(1) = 0$. This means that Bob will ignore the challenge, leading to the case studied in the previous section. Now since in region $R1$, p is always less that r_1, we have that $T \leq \pi(0)A \leq \pi(1)B$. As a consequence, the best choices for q_0 and q_1 are $q_0 = q_1 = 0$. This means that Bob will never challenge a message and will always trust it. The best response to such strategy for Trudy is to choose $p_0 = p_1 = 1$.

- If $B \geq V \geq A$, all we need to show in that in region $R2$ there is a choice of p_0 and p_1 that satisfies $T \geq \pi(0)A$ and $p(\pi(0)p_0V + \pi(1)p_1B) \leq \pi(0)V$. But, this is always the case because since $p \geq r_1$, we have that $p(\pi(0)A + \pi(1)B) \geq \pi(0)A$. This implies that there exist p_0 and p_1 such that $p(\pi(0)p_0A + \pi(1)p_1B) \geq \pi(0)A$. The choices $p_0 = p_1 = 1$ is always possible. The second condition $(p(\pi(0)p_0V + \pi(1)p_1B) - \pi(0)V \leq 0)$ can also always be satisfied. In fact, letting p_0 and p_1 such that $p(\pi(0)p_0A + \pi(1)p_1B) = \pi(0)A$, we get $p\pi(1)p_1B = \pi(0)A - p\pi(0)p_0A$. Replacing this value of $p\pi(1)p_1B$ in the second condition and rearranging the terms,, we get $-\pi(0)(1-pp_0)(V-A)$ which is less than zero.

- If $V \geq B$ the challenge is not helpful anymore to Bob. If $p \leq r_1$, then V is always larger than $\frac{p}{1-p}\frac{\pi(1)}{\pi(0)}B$, and we have seen that Bob should always trust and never use the challenge.

If $p > r_1$, then any choice of p_0 and p_1 possible in the case without challenge is still possible for Trudy (from the analysis of the case $B \geq V \geq A$), and for any such choice, Bob is better of ignoring and not challenge the message. But, this is the strategy he already adopted in the case without challenge.

Chapter 3

Intelligent Virus Game

In this chapter, we use a game theoretic approach to study the interaction between an intelligent virus and an intrusion detection system (IDS). We consider a scenario where the virus is attempting to infect as many machines as possible in a network protected by a simple IDS. Figure 3.1 is a sketch of the model. The normal traffic going through the network is assumed to have a known rate α. When present, the virus, while trying to infect other parts of the network, generates some additional traffic with a rate β that the virus designer needs to set. The IDS is trying to detect the virus as early as possible while limiting false alarms. To detect intrusions, the IDS checks the amount of traffic going through the network against a chosen threshold. We use a zero-sum, Bayesian game model to analyze the situation.

The IDS *buffers* and counts the volume X_n of traffic in the interval $[(n-1)T, nT]$ for $n = 1, 2, 3, \ldots$ and decides that a virus is present the first time that $X_n > x$. T is a design parameter and is assumed to be known and x is a threshold to be chosen by the IDS. If the IDS decides that there is an infection, it then flushes the buffer to prevent the virus from infecting more machines. Buffering introduces some delay (T) in the traffic. We use a Markov chain model to compute the Nash equilibrium of the game and analyze it.

We compare the average attack cost of this IDS to the one studied in [76]. The main difference is that the one considered in this thesis *quarantines* the traffic while making its decision while the second lets the traffic through and decides later. For example, at time T, with the first IDS no traffic has not yet reached the other parts of the network, while with the second one, all the traffic

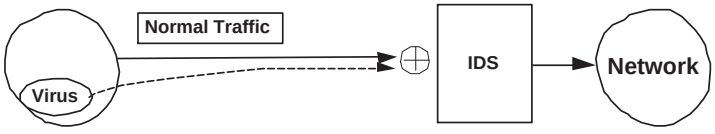


Figure 3.1: Intelligent virus game model.

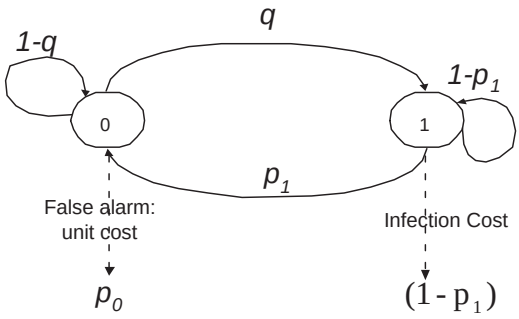


Figure 3.2: Markov chain model for computing the NE of the intelligent virus game with buffering IDS.

has already gone through. Because of the different mathematical tools used to analyze the games, a quantitative comparison of the two modes of operation is not possible. As of such, we will only qualitatively contrast them.

3.1 The Intelligent Virus Game

In this problem, we suppose that for detection purposes, the IDS delays the traffic by buffering it, and counts the volume X_n of traffic in the interval $[(n-1)T, nT]$ for $n = 1, 2, 3, \ldots$. T is a design parameter that is set in advance. The IDS decides that the network is infected the first time that $X_n > x$ and it then flushes the buffer to prevent the virus from infecting other machines. If the IDS finds that $X_n < x$, it transmits the X_n buffered bits of traffic. We are interested in the equilibrium of the game where the virus designer chooses β to maximize the damage that the virus causes and the IDS chooses the threshold x to minimize it. This damage is the expected sum of the inspection and infection costs. In practice, such IDS requires inserting a short delay into the system, say 40ms, which not all network designers might like.

We consider a SIS (Susceptible-Infected-Susceptible) virus spreading model ([77],[78]). In such model, a computer is first susceptible once a new virus appears. Then, eventually, the machine will get infected, and after cleaning it becomes again susceptible. A discrete version of such model is shown is figure 3.1. SIS models have been widely used in viral epidemiology [79], and computer virus propagation [78]. Although it is not always realistic (a more realistic model would be Susceptible-Infected-Recovered SIR), SIS models offer a nice analytical framework. Interested readers are referred to [77].

We evaluate the infection and the inspection costs of the model using a discrete time Markov chain model with time step T. The states are 0 and 1, where 0 means that the computer is not infected and 1 that it is. Let $p_0 = Pr[X_n > x|$ No Virus$]$ designate the probability that the IDS declares that the computer is infected when it is not (false alarm) and $p_1 = Pr[X_n > x|$ Virus$]$

x	p_0	p_1	$C(x,\beta)$
$0 \le x \le \beta$	$\frac{\alpha-x}{\alpha}$	1	$\frac{1-\frac{\alpha}{x}}{q+1}$
$\beta < x \le \alpha$	$\frac{\alpha-x}{\alpha}$	$\frac{\alpha+\beta-x}{\alpha}$	$C_3(x,\beta)$
$\alpha < x \le \alpha+\beta$	0	$\frac{\alpha+\beta-x}{\alpha}$	$\gamma\beta q\frac{x-\beta}{\alpha(q+1)+\beta-x}$

Table 3.1: Values of the false alarm p_0 and detection p_1 probabilities.

the probability that the IDS declares that the computer is infected when it actually is (correct detection). Finally, let q be the probability that the computer gets infected in one time step.

Figure 3.1 shows a diagram of the Markov chain. The transition probabilities of the Markov chain are then $P(0,1) = q$ and $P(1,0) = p_1$. Also, when the system is in state 0, it generates a false alarm with probability p_0 and this false alarm is assumed to have a unit cost. When the system is in state 1, it generates an average number of viruses equal to $\beta(1-p_1)$ and we assume that each released virus has an average cost equal to γ. Thus, γ measures the likelihood that a released virus is successful in infecting another computer. (More complex models are certainly plausible.)

With this model, the average cost per unit of time is

$$C(x,\beta) = p_0\pi_0 + \gamma\beta(1-p_1)\pi_1. \tag{3.1}$$

where π_0 (resp. π_1) is the stationary probability that the system is in state 0 (resp. 1).
The stationary distribution can be computed by solving the balance equations:

$$q\pi_0 = p_1\pi_1, \qquad \text{and} \qquad \pi_0 + \pi_1 = 1. \tag{3.2}$$

Consequently,

$$C(x,\beta) = \frac{p_0p_1 + \gamma\beta(1-p_1)q}{p_1+q}. \tag{3.3}$$

The virus designer chooses the infection rate β to maximize this cost while the IDS computes the best threshold to minimize it.

3.1.1 Deriving the NE

We analyze this model by assuming that the traffic X_n is uniform in $[0,\alpha]$ when the computer is not infected and uniform in $[\beta,\alpha+\beta]$ when the computer is infected (i.e., the virus is introducing additional traffic of constant rate β).

With this uniform assumptions, the probabilities p_0 and p_1 as functions of x and β are given in Table 3.1. Since an infection rate $\beta > \alpha$ will always result to 100% detection, we assume that β is less than α. The table also shows the corresponding values of the average cost $C(x,\beta)$.

The value $C_3(x,\beta)$ in the table is equal to

$$C_3(x,\beta) = \frac{\alpha+\beta\left(1-\gamma\beta q+\left(x^2+(\alpha\gamma\beta q-2\alpha-\beta)\right)\right)}{\alpha(q+1)+\beta-x}. \tag{3.4}$$

q	γ	β	x
0.01	0.02	150	980
0.01	0.05	220	900
0.01	0.10	250	750
0.10	0.02	250	580
0.10	0.05	150	310
0.10	0.10	90	180

Table 3.2: Nash equilibrium (β, x) as a function of the parameters (q, γ).

Since the game is zero-sum, the Nash equilibrium strategies correspond to the max-min strategies. Hence, to find the NE, we first minimize $C(x, \beta)$ over x in each of the intervals given above; for each region this gives a function that only depends on β. We then maximize over β in each region, and take the maximum value across the regions. Since the calculation is very involved, we use a small matlab script to compute the NE. The Nash equilibrium depends on the values of q, α, and γ. Table 3.2 shows some representative values when $\alpha = 1000$.

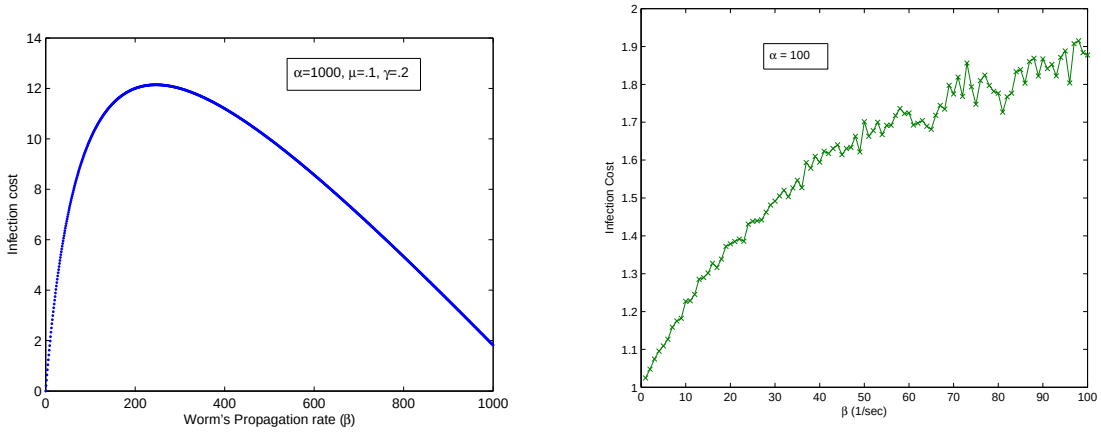
3.1.2 Qualitative Comparison

In this section, we analyze how the attack cost depends on the strategy (propagation rate) of the virus. In Figure 3.3(a) the attack cost of the IDS is shown as a function of the rate of propagation chosen by the virus. For each value of the infection rate (shown in the horizontal axis), we have computed the best response of the IDS which correspond to the optimal threshold for that rate. Then, we compute the cost as in (3.3) where x corresponds to the best response to β.

As can be seen in the figure, the cost of attack is small when the infection rate is small. This tells that a less aggressive virus will cause less damage to the system. The figure also shows a small cost of attack for high values of the infection rate. This means that a very aggressive virus will also cause small damage to the system. In fact, since the IDS buffers the traffic while making a decision, and flushes the buffer if it detects a virus, an aggressive attack will very likely be detected at the IDS level and will not make it into the system.

These two observations show that an intelligent virus will neither be too slow, nor too aggressive. Instead, it will choose an infection rate that balances between the potential reward of an aggressive attack and the risk of being quickly detected. This tradeoff is given by the Nash equilibrium attack rate. It is equal to the rate that maximizes the minimum attack cost that the IDS can impose.

For comparison purposes, we have plotted in 3.3(b), the average cost of an IDS based on the model in [76] as a function of the propagation rate of the virus. The model uses a sequential probability ratio test to compute the best threshold for a given infection cost and cost of false alarm (which in [76] are the false alarm and missed detection probabilities). In our simulation, we have assumed that the false alarm cost is fixed and the infection cost is a linear function of the infection rate and of the time to detection. Notice that in the model, the virus is assumed to be



(a) Virus attack cost as a function of the propagation rate. The cost is given by (3.3) where x is chosen to be the best response to β.

(b) Monte Carlo simulation of the cost of attack as a function of the virus propagation rate for the IDS model in [76]. The false alarm cost is normalized to 1. The infection cost is of the form $\gamma\beta t$, where $\gamma = 0.02$.

Figure 3.3: Cost of security as a function of the virus propagation rate. Different plots are shown for different values of the rate of normal traffic α.

present or not present at time 0 with known probabilities.

As can be seen in the figure, the attack cost is an increasing function of the infection rate. This tells that the more aggressive a virus is, the more damage it will cause to the system; the extreme case being a virus that can send at a rate equal to the maximum capacity the channel. Of course, there are some natural limitations to the rate of propagation that viruses cannot get around (e.g. finite channel bandwidth) . However, the figure suggests that as far as there is some room for increasing the infection rate, an intelligent virus will do so.

The phenomenon that we observe in Figure 3.3(b) is common to most IDS models because they essentially have an observation window, which delays the detection time. It is during this observation window when aggressive viruses can cause large damage by sending at arbitrary rate. The model we have considered, overcomes this by delaying the communication of (eventually) normal users. This might not be desirable for certain applications and is certainly a limitation of the model. However, using this model, one can guarantee a certain level of security independently to how aggressive the attackers are. By setting the threshold x equal to the Nash Equilibrium value x_{NE}, the infection cost $Cost(\beta, x_{NE})$ is always less than the NE cost $Cost(\beta_{NE}, x_{NE})$.

An interesting follow up question is how to balance between the delay introduced by the IDS (which might not be desirable for certain applications) and the reduction in cost compared to traditional IDS. In our study, we have assumed that the delay T is fixed and known. A more realistic game is one where both the delay T and the threshold x are chosen by the IDS. Such study will be the subject of future work.

Chapter 4

Blocking Games

In this chapter, we consider availability (or denial of service) attacks where resources that are (or might be) needed by a defender are the target of a cognitive attacker. More precisely, we consider that there is a finite set S and two collections $\mathcal{T}$ and $\mathcal{E}$ of subsets of S. The defender selects a subset $T \in \mathcal{T}$ to perform a *mission critical* task. Each subset $T \in \mathcal{T}$ needs some set of resources $e_{T_1}, e_{T_2}, \ldots, e_{T_p} \in \mathcal{E}$ in order to fulfill the task. To disrupt the mission, an attacker targets one resource $e \in \mathcal{E}$ to attack. For example, $\mathcal{T}$ could be the collection of spanning trees of a graph and the collection of resources $\mathcal{E}$ could be the edges of the graph (as we will see in our application examples).

When attacked, the cost (loss) to the defender is $\boldsymbol{\lambda}_{T,e}$, which models the damage caused by the attacker whenever subset $T \in \mathcal{T}$ and resource $e \in \mathcal{E}$ are selected.
The reward to the attacker is $\boldsymbol{\lambda}_{T,e} - \boldsymbol{\mu}(e)$, where $\boldsymbol{\mu}(e)$ is the cost of attacking the resource $e \in \mathcal{E}$. The attacker also has the option to not launch an attack. The goal of the defender is to minimize his loss, and the attacker is trying to maximize her net gain.

We model this situation as a *"quasi" zero-sum* game and consider mixed strategy Nash equilibria, where the defender chooses a distribution on the collection of subsets $\mathcal{T}$; and the attacker a distribution on the collection of resources $\mathcal{E}$. The details of the model are discussed in section 4.2. The results of the game are discussed as a theorem in section 4.3. A proof of the theorem is presented in section 4.5. Examples of applications of the model and results are discussed in section 4.4.

A certain number of observations are made from the analysis of the Nash equilibria.
First, there exists at least one set $E \subseteq \mathcal{E}$ of resources that is more vulnerable than the others – in the sense that it gives a higher expected net reward to the attacker. Such subsets of resources are called *critical*. We use the maximum achievable reward as the *vulnerability* metric for the defender's task. If the vulnerability is negative, the attacker will not launch an attack; and if it is positive, the attacker will always target critical subsets of resources. The defender will select subsets that

"minimally"* intersect the targeted critical subsets.

Second, if the defender's loss function can be written as a submodular function, then there exists a polynomial time algorithm to compute a critical subset of resources (and hence a Nash equilibrium) for the game. We derive this algorithm by relating the computation of the vulnerability of the task to the minimization of a submodular function. We then use the existing polynomial algorithms for submodular function minimization. The derivation of the algorithm is shown in section 4.6. It requires the notions of Matroid, Polymatroid, and submodular function minimization which we will briefly discuss in the same section.

The model is applied to two examples that are discussed in section 4.4.
Our first example considers a scenario where a network manager is choosing a spanning tree on the graph of a network to connect all nodes. An attacker is trying to cut the tree by attacking one link of the graph. One of our findings is that, in this scenario the minimum cutsets of the graph are not the most critical subsets of links.

The second example deals with a *supply-demand* network where a network manager is choosing a feasible flow to transport the maximum amount of goods from a set of sources to a set of destinations, and an attacker is trying to minimize this by attacking an arc of the network. In this case, we find that critical subsets are cutsets that maximize the minimum fraction of goods carried per link of the cutset. In most cases, these correspond to minimum cutsets of the graph.

Although in this thesis we focus our attention to examples on network topology, there is a broad range of applications that fit into the framework of this model.

For general games of this model, we determine the structure of a particular set of Nash equilibria. When the game is zero-sum, we characterize the set of all Nash equilibria. Section 4.5 shows the details of the proof. To compute the Nash equilibria, we make use of the combinatorial tools of *blocking pair of matrices*. Since these notions are central to our analysis, we start by reviewing them in section 4.1.

Notational Conventions:
In this chapter, we make the following notational conventions.
We use the *prime* sign (') for transpose of matrices and vectors. The column vector of ones of length $|\mathcal{T}| = N$ will be denoted by $\mathbf{1}_{\mathcal{T}}$, and $\mathbf{1}_{\mathcal{E}}$ denotes the column vector of ones of length $|\mathcal{E}| = m$. All vectors are denoted by *bold* lowercase letters and are assumed to be column vectors. Matrices will be denoted (mostly and not exclusively) by uppercase Greek letters Λ, Ω. The indicator function will be denoted 1_{cond}. It is equal to 1 if "*cond*" is satisfied and is equal to 0 otherwise. We will use $\boldsymbol{\alpha}$ to designate the mixed strategy of the defender, and for the attacker, we use $\boldsymbol{\beta}$. Ay other probability distribution will be denoted by γ.

*This notion of minimality depends on the cost functions of the games and will become more clear once we define those cost functions.

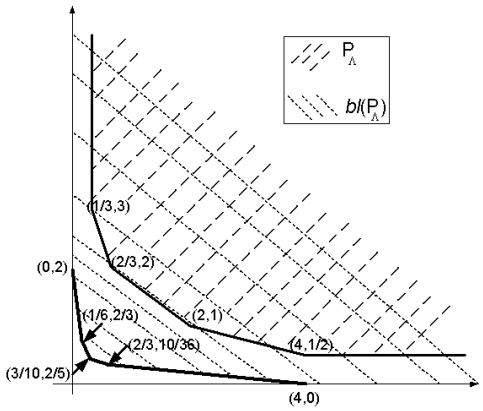


Figure 4.1: Example of polyhedron P_Λ defined by a nonnegative proper matrix Λ and its corresponding blocker $bl(P_\Lambda)$. The extreme points of the blocker define the nonnegative proper matrix Ω.

4.1 Blocking Pair of Matrices

The discussion in this section is mostly based on [80, pp. 99-101] and [81].
Let Λ be a $N \times m$ nonnegative matrix with non-zero rows. The polyhedron P_Λ associated with Λ is defined as the vector sum of the convex hull of its rows $(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)$ and the nonnegative orthant:

$$P_\Lambda = \text{conv.hull}\,(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N) + \mathbb{R}^m_+. \tag{4.1}$$

A row $\boldsymbol{\lambda}_i$ of Λ is said to be *inessential* if it dominates a convex combination of other rows of Λ, otherwise we say that $\boldsymbol{\lambda}_i$ is *essential*. If all the rows of Λ are essential, we say that Λ is *proper*. The set of essential rows corresponds to the set of *extreme* points of the polyhedron P_Λ. Since inessential rows are not relevant for the definition of P_Λ we will drop them and assume that Λ is proper.

The example in Figure 4.1 shows the associated polyhedron P_Λ to the nonnegative matrix Λ given below:

$$\Lambda = \begin{pmatrix} 1/3 & 3 \\ 2/3 & 2 \\ 2 & 1 \\ 4 & 1/2 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0 & 2 \\ 1/6 & 2/3 \\ 3/10 & 2/5 \\ 2/3 & 10/36 \\ 4 & 0 \end{pmatrix} \tag{4.2}$$

Given Λ and its associated polyhedron, we define the blocker of the polyhedron P_Λ as follow.

Definition 1 *The* blocker $\text{bl}(P_\Lambda)$ *of* P_Λ *is the polyhedron given as:*

$$\text{bl}(P_\Lambda) = \left\{\mathbf{y} \in \mathbb{R}^m_+ :\ \mathbf{y}'\mathbf{x} \geq 1, \quad \forall\ \mathbf{x} \in P_\Lambda\right\}, \tag{4.3}$$

where $\mathbf{y}'\mathbf{x}$ *is the inner product of* $\mathbf{y}$ *and* $\mathbf{x}$*. Recall that in this thesis we use the prime sign (') for vector and matrix transpose.*

We are interested in characterizing the polyhedron P_Λ and its blocker $bl(P_\Lambda)$. For that, we use the following theorem by Fulkerson [81]. It is based on the fact that there is a one-to-one correspondence between the essential rows of Λ and the extreme points of P_Λ.

Theorem 3 (Fulkerson, 1971) *Let the* N-by-m *matrix* Λ *be proper with rows* $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N$*, and let the polyhedron* P_Λ *be defined as in (4.1). Let* $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K$ *be the extreme points of* $\mathrm{bl}(P_\Lambda)$*, and let* Ω *be the matrix having those points as rows. Then,*

1. *The blocker* $\mathrm{bl}(P_\Lambda)$ *of* P_Λ *is given by* $\mathrm{bl}(P_\Lambda) = \left\{\mathbf{x} \in \mathbb{R}_+^m : \ \Lambda\mathbf{x} \geq \mathbf{1}_\mathcal{T}\right\}$.
2. Ω *is proper, and the polyhedron* P_Λ *can be described as* $P_\Lambda = \left\{\mathbf{x} \in \mathbb{R}_+^m : \ \Omega\mathbf{x} \geq \mathbf{1}_K\right\}$.
3. *The blocker of the blocker* $\mathrm{bl}(P_\Lambda)$ *verifies* $\mathrm{bl}(\mathrm{bl}(P_\Lambda)) = P_\Lambda$.

Λ *and* Ω *are said to form a blocking pair of matrices.*

Equations (4.2) show a blocking pair of matrices Λ and Ω, and the corresponding polyhedra are shown in Figure 4.1.

Blocking pairs of matrices play an important role in the combinatorial problem of *maximum packing* (see Fulkerson[81]). In this thesis, we use the theory of blocking pairs to provide an easy argument for the existence of a probability distribution that satisfies a certain number of constraints. For instance, consider the following linear program:

$$\begin{aligned} &\text{Maximize } \mathbf{1}_\mathcal{T}'\mathbf{x} \\ &\text{subject to } \Lambda'\mathbf{x} \leq \mathbf{w}, \quad \text{and} \quad \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{4.4}$$

where the constraints Λ form a nonnegative matrix, and $\mathbf{w}$ is a given nonnegative vector.
We are interested to knowing whether the value of the program is greater than 1 or not. If this is the case, one can easily derive a probability distribution by normalizing a feasible solution of the program. Indeed, since the normalizing factor is greater than 1, the constraints will still be satisfied. The following lemma gives an answer to our question.

Lemma 1 *The value of the LP in (4.4) is greater than 1 if and only if* $\mathbf{w}$ *belongs to the polyhedron* P_Λ *defined by* Λ*.*

Proof: The proof of the lemma is as follow.
First, notice that strong duality holds for this LP. In fact, Slater's condition [82] is satisfied for any nonnegative and nonzero $\mathbf{w}$. The dual of the LP is given as:

$$\begin{aligned} &\text{Minimize } \mathbf{w}'\mathbf{y} \\ &\text{subject to } \Lambda\mathbf{y} \geq \mathbf{1}_\mathcal{T}, \quad \text{and} \quad \mathbf{y} \geq \mathbf{0}. \end{aligned} \tag{4.5}$$

The constraints of the dual program (4.5) define the blocker $bl(P_\Lambda) = \left\{\mathbf{y} \in \mathbb{R}^m_+ : \ \Lambda\mathbf{y} \geq \mathbf{1}_\mathcal{T}\right\}$ of the polyhedron P_Λ. Now, if $\mathbf{w}$ belongs to P_Λ, then for all $\mathbf{y} \in bl(P_\Lambda)$, we have that $\mathbf{w}'\mathbf{y} \geq 1$.

Conversely, if $\mathbf{w}'\mathbf{y} \geq 1$ for all $\mathbf{y} \in bl(P_\Lambda)$, then $\mathbf{w}$ must be in the blocker of $bl(P_\Lambda)$, which by Fulkerson's theorem 3, is P_Λ. This implies that the value of the dual program is greater than 1. Combined with the strong duality property, we get that the value of the primal program is at least 1. ■

4.2 Game Model

This section presents the blocking game model and introduces some notations that we will need to characterize the Nash equilibria of the game.

We consider that there is a nonempty, finite set S and two nonempty collections $\mathcal{T} = \{T_1, \ldots, T_N\}$ and $\mathcal{E} = \{e_1, \ldots, e_m\}$ of nonempty subsets of S. We call $\mathcal{E}$ the collection of resources. The defender selects a subset $T \in \mathcal{T}$ to perform a *mission critical* task. Each subset $T \in \mathcal{T}$ needs some set of resources $e_{T_1}, e_{T_2}, \ldots, e_{T_p} \in \mathcal{E}$ in order to fulfill the task. To disrupt the mission, an attacker targets one resource $e \in \mathcal{E}$ to attack. Each resource $e \in \mathcal{E}$ has a cost of attack $\boldsymbol{\mu}(e)$ that is the amount of effort that the attacker needs to spend to successfully launch the attack. The attacker also has the option of not attacking ("No Attack"); which we materialize by the choice of $e_\emptyset$. This choice results to zero loss for the defender and zero reward for the attacker.

Whenever the defender chooses subset T and resource e is attacked, he loses some value $\boldsymbol{\lambda}_{T,e}$. This loss goes to the attacker. It is conceivable that $\boldsymbol{\lambda}_{T,e} = 0$ if subset T does not need resource e. Hence, when the pair (T, e) is selected, the defender's loss is $\boldsymbol{\lambda}_{T,e}$ and the attacker's net reward is equal to $\boldsymbol{\lambda}_{T,e} - \boldsymbol{\mu}(e)$.

This scenario can be modeled as a two-player matrix game where the players (the defender and the attacker) choose their pure strategies in the nonempty and finite sets $\mathcal{T}$ and $\mathcal{E} \cup \{e_\emptyset\}$, respectively (with $|\mathcal{E}| = m$ and $|\mathcal{T}| = N$).
The defender and attacker's respective payoff matrices are given by

$$\tilde{\Lambda} = [\Lambda|0], \quad \text{and} \quad \tilde{\Pi} = [\Pi|0], \tag{4.6}$$

where Λ and Π are N-by-m matrices; $(\Lambda, \ [\Lambda]_{T,e} = \boldsymbol{\lambda}_{T,e})$ is a nonnegative matrix with no zero rows[†],

[†]A row with all zeros would lead to a trivial game because it means that there exists a subset $T \in \mathcal{T}$ that will always result to zero loss. The defender would then always select such strategy and the game ends.

and such that there is no column of Λ that dominates all other columns[‡]. Π is given by,

$$\Pi = \Lambda - \begin{pmatrix} \boldsymbol{\mu}(1) & \boldsymbol{\mu}(2) & \dots & \boldsymbol{\mu}(m) \\ \boldsymbol{\mu}(1) & \boldsymbol{\mu}(2) & \dots & \boldsymbol{\mu}(m) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mu}(1) & \boldsymbol{\mu}(2) & \dots & \boldsymbol{\mu}(m) \end{pmatrix}, \tag{4.7}$$

The "last" all-zero column in the definition of $\tilde{\Lambda}$ and $\tilde{\Gamma}$ captures the zero-loss for the defender ($\boldsymbol{\lambda}_T(e_\emptyset) = 0$) and zero-reward for the attacker ($\boldsymbol{\mu}(e_\emptyset) = 0$) when this latter chooses the *"No Attack"* strategy ($e_\emptyset$). For notational simplicity, we will only mention Λ and Γ.

We consider mixed strategies of this game where the defender chooses a distribution $(\boldsymbol{\alpha}_T,\ T \in \mathcal{T})$ on $\mathcal{T}$ and the attacker chooses a distribution $(\boldsymbol{\beta}(e),\ e \in \mathcal{E} \cup \{e_\emptyset\})$ on $\mathcal{E} \cup \{e_\emptyset\}$. The goal of the defender is to minimize the expected loss

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\alpha}' \tilde{\Lambda}[\boldsymbol{\beta}; \boldsymbol{\beta}(e_\emptyset)] = \sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \boldsymbol{\lambda}_{T,e} \right), \tag{4.8}$$

while the attacker is trying to maximize the expected reward

$$R(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\alpha}' \tilde{\Pi}[\boldsymbol{\beta}; \boldsymbol{\beta}(e_\emptyset)] = \sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \left(\sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \boldsymbol{\lambda}_{T,e} - \boldsymbol{\mu}(e) \right). \tag{4.9}$$

In our notation, $[\boldsymbol{\beta}; \boldsymbol{\beta}(e_\emptyset)]$ is the column vector obtained by appending the additional row (entry) $\boldsymbol{\beta}(e_\emptyset)$ to the column vector $\boldsymbol{\beta}$.

Let P_Λ be the polyhedron associated with Λ given in (4.6), and let $bl(P_\Lambda)$ denote its blocker. From the discussion in the previous section and from Fulkerson's theorem, the blocker $bl(P_\Lambda) \subseteq \mathbb{R}^m_+$ is the polyhedron associated with the nonnegative matrix Ω whose rows are the vertices of $bl(P_\Lambda)$. It is also known that $bl(P_\Lambda)$ is the vector sum of the convex hull of rows of Ω with the positive orthant $\mathbb{R}^m_+$, and that its blocking polyhedron is P_Λ (see [81] and [80, pp. 99-101]). Also, Theorem 3 gives that

$$P_\Lambda = \left\{ \mathbf{x} \in \mathbb{R}^m_+,\ \text{s.t.}\ \ \Omega \mathbf{x} \geq \mathbf{1}_\mathcal{T} \right\} \tag{4.10}$$

Now, for $\boldsymbol{\omega}$ row of Ω[§], which we denote as $\boldsymbol{\omega} \in \Omega$, we write $\boldsymbol{\omega} = (\boldsymbol{\omega}(e), e \in \mathcal{E})$, and let $\boldsymbol{\omega}(\mathcal{E}) := \sum_{e \in \mathcal{E}} \boldsymbol{\omega}(e)$. Note that $\boldsymbol{\omega}(e) \geq 0$ for all $e \in \mathcal{E}$ and $\boldsymbol{\omega}(\mathcal{E}) > 0$[¶]; so that $(\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}, e \in \mathcal{E})$ is a probability distribution on $\mathcal{E}$. We call it the probability distribution induced by $\boldsymbol{\omega}$.

[‡]This would also lead to a trivial game if the attack cost is not too high. In fact, it means that there is a resource that will always give higher attack gain, independently of the subset chosen by the defender. The attacker would always target such a resource.

[§]Notice that this is an abuse of language because $\boldsymbol{\omega}$ is a column vector.

[¶]This is because the blocker $bl(P_\Lambda)$ is not empty (Λ is not a one-rowed zero matrix), and does not contain the all-zero vector (the origin) (this could not give an inner product with rows of Λ that is greater than 1).

Also define, for each $\boldsymbol{\omega} \in \Omega$, the quantities

$$\lambda(\boldsymbol{\omega}) := \min_{T \in \mathcal{T}} \left(\sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\lambda}_T(e) \right); \tag{4.11}$$

and

$$\theta(\boldsymbol{\omega}) := \lambda(\boldsymbol{\omega}) - \sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\mu}(e). \tag{4.12}$$

$\lambda(\boldsymbol{\omega})$ is the minimum loss seen by the defender if the attacker were to choose a target according to the distribution $(\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}, e \in \mathcal{E})$ induced by $\boldsymbol{\omega}$.
The expression of $\theta(\boldsymbol{\omega})$ is composed with two terms. If the attacker were to choose a resource to attack according to the distribution $\boldsymbol{\omega} = (\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}, e \in \mathcal{E})$, then the first term would have been the loss seen by the defender, which, as we have assumed, goes to the attacker. The second term is the average cost of attack corresponding to the distribution induced by $\boldsymbol{\omega}$. Hence, $\theta(\boldsymbol{\omega})$ can be seen as the expected attack reward associated $\boldsymbol{\omega}$.

We call the vertex $\boldsymbol{\omega}$ of the blocking polyhedron a *critical* vertex if

$$\theta(\boldsymbol{\omega}) = \max_{\tilde{\boldsymbol{\omega}} \in \Omega} \theta(\tilde{\boldsymbol{\omega}}) \ . \tag{4.13}$$

A critical vertex is one whose induced probability distribution gives a maximum rewards to the attacker (considering the defender's response).
We define $\theta := \max_{\tilde{\boldsymbol{\omega}} \in \Omega} \theta(\tilde{\boldsymbol{\omega}})$ to be the maximum achievable value in the preceding expression, and we let Ω_{max} denote the matrix having as rows the critical vertices of $bl(P_\Lambda)$.
We use θ as the *vulnerability* for the defender's task.

4.3 Nash Equilibrium Theorem

This section presents the main results of the two-player matrix game defined in section 4.2. We claim that:

Theorem 4 *For the game defined above, the following always hold.*

1. *If $\theta \leq 0$, then "No Attack" (i.e. $\boldsymbol{\beta}(e_\emptyset) = 1$) is always an optimal strategy for the attacker. In this case, the equilibrium strategy $(\boldsymbol{\alpha}_T,\ T \in \mathcal{T})$ for the defender is such that*

$$\bar{\lambda}_{\boldsymbol{\alpha}}(e) := \sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \boldsymbol{\lambda}_T(e) \leq \boldsymbol{\mu}(e), \quad \forall e \in \mathcal{E}. \tag{4.14}$$

   The corresponding payoff is 0 for both players.

2. *If $\theta \geq 0$, then for every probability distribution $(\gamma_{\boldsymbol{\omega}}, \boldsymbol{\omega} \in \Omega_{max})$, the attacker's strategy $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ defined by*

$$\boldsymbol{\beta}(e) = \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \tag{4.15}$$

is in Nash equilibrium with any strategy $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ of the defender that satisfies the following properties:

$$\begin{cases} \bar{\lambda}_{\boldsymbol{\alpha}}(e) - \boldsymbol{\mu}(e) = \theta \quad \text{for all} \quad e \in \mathcal{E} \quad \text{such that} \quad \boldsymbol{\beta}(e) > 0. \\ \bar{\lambda}_{\boldsymbol{\alpha}}(e) - \boldsymbol{\mu}(e) \leq \theta \quad \text{for all} \quad e \in \mathcal{E}. \end{cases} \tag{4.16}$$

Further, there exists at least one such strategy $\boldsymbol{\alpha}$.

The corresponding payoffs are θ for the attacker, and $r(\gamma)$ for the defender, where

$$r(\gamma) := \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \lambda(\boldsymbol{\omega}). \tag{4.17}$$

3. *If $\boldsymbol{\mu} = 0$, then every Nash equilibrium pair of strategies for the game is of this type.*

Note: $\theta = 0$ is a particular case where both cases (1) and (2) can occur. In all cases, the maximum achievable attack reward is equal to 0. There exist equilibria where the attacker decides to not attack. In those cases, the defender has to choose $\boldsymbol{\alpha}$ according to (4.14). There might also exist equilibria where the attack launches an attack but gets a expected reward of 0. In such equilibrium, $\boldsymbol{\alpha}$ has to satisfy (4.16).
In our experiments we did not find any equilibrium where the attacker would mix between $e_\emptyset$ and some resources in $\mathcal{E}$.

4.4 Examples

In this section, we illustrate the game model and the NE theorem using two examples.
The first example considers a network given as a connected undirected graph. The aim is to study the *strategic* interaction between a network manager whose goal is to choose a spanning tree of the network as communication infrastructure, and an attacker who tries to disrupt the communication tree by attacking one link in the network. In the second example, we analyze games on a *supply-demand* network where a defender selects a feasible flow to carry a maximum amount of goods from a set of sources to a set of destinations. An attacker is trying to disrupt the transport by attacking one arc of the network. We will focus our study on networks with infinite capacity on arcs. Capacitated networks can also be studied by using the same tools as un-capacitated networks. We leave this study for future work.

4.4.1 The spanning tree – link game

The network topology is given by a connected undirected graph $G = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{E}| = m$ links and $|\mathcal{V}| = n$ nodes. The collection of subsets $\mathcal{T}$ is the set of spanning trees; we let $N = |\mathcal{T}|$.

To get all nodes connected in a cycle-free way, the network manager chooses a spanning tree $T \in \mathcal{T}$ of the graph. The attacker simultaneously selects an edge $e \in \mathcal{E}$ to attack. Each edge $e \in \mathcal{E}$ is associated with some cost $\boldsymbol{\mu}(e)$ that an attacker needs to spend to launch a successful attack on e. All trees T have the same cost for the network manager that we assume to be equal to 1. The attacker wins if the attacked link belongs to the chosen spanning tree, otherwise the network wins. More precisely, for a choice pair (T, e) of tree and edge, the attack loss is $1_{e\in T}$ for the network, while the net attack reward is equal to $1_{e\in T} - \boldsymbol{\mu}(e)$ for the attacker. $1_{e\in T}$ is the indicator variable that is equal to 1 of $e \in T$, and 0 otherwise. It is also assumed that the attacker has the option of not attacking.

This scenario fits well to the model discussed in the previous section. The payoff matrix Λ in this case, corresponds to the *tree-link* incidence matrix of the graph, and the associated polyhedron P_Λ is called the *spanning tree polyhedron*. Next, we characterize P_Λ and its blocker $bl(P_\Lambda)$.

The spanning tree polyhedron P_Λ and its blocker $bl(P_\Lambda)$

Recall (4.1) that the polyhedron P_Λ is defined as the vector sum of the convex hull of its rows $(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)$ and the nonnegative orthant:

$$P_\Lambda = \text{conv.hull}\,(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N) + \mathbb{R}^m_+. \tag{4.18}$$

We are interested in characterizing P_Λ as well as its blocker polyhedron $bl(P_\Lambda)$. For that, Chopra [83] proposed the following idea that is based on partitioning the nodes of the graph.

For a connected graph $G = (\mathcal{V}, \mathcal{E})$, a minimum cut partitions the node set $\mathcal{V}$ into two subsets $\mathcal{V}_1$ and $\mathcal{V}_2$, and includes all the edges having one end point in $\mathcal{V}_1$ and the other one in $\mathcal{V}_2$. Furthermore, each of the subgraphs, $G_i = (\mathcal{V}_i, \mathcal{E}(\mathcal{V}_i)),\ i = 1, 2$ is connected. This notion can be generalized to a partition $P = (\mathcal{V}_1, \ldots, \mathcal{V}_{|P|})$ of the nodes of G such that each subgraph $G_i = (\mathcal{V}_i, E(\mathcal{V}_i)),\ i = 1, \ldots, |P|$ is connected. Such a partition is said to be *feasible*. $|P|$ is the size of the partition (number of partition elements).

Now, consider a feasible partition $P = (\mathcal{V}_1, \ldots, \mathcal{V}_{|P|})$ and any spanning tree T. Since the graph is connected, the links used by T must connect all elements $\mathcal{V}_i$ of P. This is done by using only links that go from one element $\mathcal{V}_i$ of the partition to another $\mathcal{V}_j,\ j \neq i$. We let $\mathcal{E}(P)$ be the set of edges that go across the elements of the partition. At least $|P| - 1$ of those links are needed to connect the $\mathcal{V}_i,\ i = 1, \ldots, |P|$. Thus, the spanning tree T contains at least $|P| - 1$ links in $\mathcal{E}(P)$. The rows of Λ (denoted by $\boldsymbol{\lambda}_T$) are the incidence vector corresponding to the spanning trees T. From this we deduce that the $\boldsymbol{\lambda}_T$'s must verify

$$\sum_{e\in\mathcal{E}(P)} \boldsymbol{\lambda}_{T,e} \geq |P| - 1 \text{ for all feasible partitions, } P \text{ and all } T \in \mathcal{T}. \tag{4.19}$$

Recall that in this example, $\boldsymbol{\lambda}_{T,e} = 1_{e\in T}$.

If a vector $\mathbf{x}$ belongs to the spanning tree polyhedron P_Λ, then it is the sum of a convex combination of the vertices of Λ plus some positive vector. As a consequence, $\mathbf{x}$ must also satisfy

$\sum_{e\in\mathcal{E}(P)} \mathbf{x}(e) \geq |P| - 1$. Using this idea, we give the following characterization (by Chopra [83]) of the spanning tree polyhedron.

Proposition 1 *The spanning tree polyhedron of the graph G corresponds to the set*

$$P_\Lambda = \left\{ \mathbf{x} \in R_+^m \;\; | \;\; \sum_{e\in\mathcal{E}(P)} \mathbf{x}(e) \geq |P| - 1, \;\; \forall \; P \text{ feasible partitions} \right\}, \tag{4.20}$$

where $\mathcal{E}(P)$ denotes the set of edges that go between vertices in distinct elements of the partition P.

We have already argued that any vector in the spanning tree polyhedron belongs to the set defined in the RHS of the equation above. To see the inverse, assume that $\mathbf{x}$ verifies the characterization in the RHS of (4.20) and assume that $\mathbf{x} \notin P_\Lambda$. This latter assumption implies that $\mathbf{x}$ is dominated by any linear combination of the $\boldsymbol{\lambda}_T$'s, i.e.

$$\mathbf{x} < \sum_{T\in\mathcal{T}} \gamma_T \boldsymbol{\lambda}_T, \tag{4.21}$$

for any $(\gamma_T, \; T \in \mathcal{T})$ such that $\sum_{T\in\mathcal{T}} \gamma_T = 1$.
For any feasible partition P, there exists a spanning tree T_P that contains no more than $|P| - 1$ edges of $\mathcal{E}(P)$. For such T_P, $\sum_{e\in\mathcal{E}(P)} \boldsymbol{\lambda}_{T,e} = |P| - 1$. Now, let $(\gamma_T, \; T \in \mathcal{T})$ be such that $\gamma_{T_P} = 1$, then (4.21) implies that

$$\sum_{e\in\mathcal{E}(P)} \mathbf{x}(e) < \sum_{e\in\mathcal{E}(P)} \sum_{T\in\mathcal{T}} \gamma_T \boldsymbol{\lambda}_{T,e} = \sum_{e\in\mathcal{E}(P)} \boldsymbol{\lambda}_{T,e} = \sum_{e\in\mathcal{E}(P)} 1_{e\in T} = |P| - 1. \tag{4.22}$$

This contradicts the hypothesis that $\mathbf{x}$ satisfies the characterization in the RHS of (4.20). As a consequence, $\mathbf{x} \in P_\Lambda$, which completes the argument.

Next, we need to characterize the blocker $bl(P_\Lambda)$ of the spanning tree polyhedron. Recall that $bl(P_\Lambda)$ is defined as:

$$bl(P_\Lambda) = \left\{ \mathbf{y} \in \mathbb{R}_+^m : \; \mathbf{y}'\mathbf{x} \geq 1, \;\; \forall \; \mathbf{x} \in P_\Lambda \right\}. \tag{4.23}$$

This is the set of nonnegative vectors $\mathbf{y}$ that have an inner product greater to 1 with any $\mathbf{x} \in P_\Lambda$. Since any $\mathbf{x} \in P_\Lambda$ dominates a convex combination of the vertices of P_Λ, we can focus on inner products $\mathbf{y}'\mathbf{x}$, for $\mathbf{x}$ vertex of the spanning tree polyhedron. We know that those vertices correspond to the rows of Λ which are the incidence vector of the spanning trees $(\boldsymbol{\lambda}_T)$. Hence, we are looking for nonnegative vectors $\mathbf{y}$ that have an inner product equal to at least 1 with any incidence vector $\boldsymbol{\lambda}_T$.

We start by constructing, for any given feasible partition $P = (\mathcal{V}_1, \ldots, \mathcal{V}_{|P|})$, the vector $\boldsymbol{\omega}_P$ that satisfies $(\boldsymbol{\omega}_P)' \boldsymbol{\lambda}_T \geq 1$ for all $\boldsymbol{\lambda}_T$. The construction goes as follows. For any feasible partition P, form $\boldsymbol{\omega}_P$ whose entries correspond to the edges of the graph. Set entry $\boldsymbol{\omega}_P(e)$ equal to $\frac{1}{|P|-1}$ if edge $e \in \mathcal{E}$ goes from one element $\mathcal{V}_i$ of the partition to another $\mathcal{V}_j$ (i.e. if $e \in \mathcal{E}(P)$). Otherwise, set $\boldsymbol{\omega}_P(e) = 0$.

The claim is that the inner product $(\boldsymbol{\omega}_P)' \boldsymbol{\lambda}_T \geq 1$, for any $\boldsymbol{\lambda}_T$. This is the case because any spanning tree contains at least $|P| - 1$ edges in $\mathcal{E}(P)$. The inner product is equal to 1 if the spanning tree contains exactly $|P| - 1$ edges in $\mathcal{E}(P)$.

From this and Fulkerson's theorem (Theorem 3), we deduce that the vectors $\boldsymbol{\omega}_P$ belongs to the blocker $bl(P_\Lambda)$ of the spanning tree polyhedron. Indeed, they have an inner product equal to at least 1 with any $\mathbf{x} \in P_\Lambda$ because of the fact that such $\mathbf{x}$ dominates a convex combination of the $\boldsymbol{\lambda}_T$'s.

Also, for a feasible partition P, the *essential*[‖] vectors of the construction above are *minimal* in the sense that any vector $\mathbf{y}$ that is strictly dominated by $\boldsymbol{\omega}_P$ (essential) will have an inner product $\mathbf{y}'\boldsymbol{\lambda}_T$ strictly less than 1 with some $\boldsymbol{\lambda}_T$ (for instance the spanning tree that crosses $\mathcal{E}(P)$ in exactly $|P| - 1$ edges). Thus, the (essential) vectors $\boldsymbol{\omega}_P$ must be vertices of the blocker $bl(P_\Lambda)$ of the spanning tree polyhedron. To show that they completely characterize $bl(P_\Lambda)$, we define the following polyhedron (corresponding to the sum of the convex hall of those vectors plus any positive vector),

$$P_\Omega = \text{conv.hull}\,(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K) + \mathbb{R}^m_+. \tag{4.24}$$

K here is the number of feasible partitions.

Since any vector $\boldsymbol{\omega}_P$ has an inner product equal to at least 1 with any $\mathbf{x} \in P_\Lambda$, any vector $\mathbf{y} \in P_\Omega$ verifies $\mathbf{y}'\mathbf{x} \geq 1$ for any $\mathbf{x} \in P_\Lambda$. Thus, $P_\Omega \subseteq bl(P_\Lambda)$. Now we need to show that $bl(P_\Lambda) \subseteq P_\Omega$. For that, assume that $\mathbf{y} \in bl(P_\Lambda)$. If $\mathbf{y} \notin P_\Omega$, then $\mathbf{y}$ is dominated by any linear combination of the essential vectors $\boldsymbol{\omega}_P$, i.e.:

$$\mathbf{y} < \sum_P \gamma_P \boldsymbol{\omega}_P, \tag{4.25}$$

for any probability distribution γ_P on the set of feasible partitions P for which $\boldsymbol{\omega}_P$ is essential. We also know that for any feasible partition P_o, there exists a spanning tree T_o such that $\sum_{e \in \mathcal{E}(P_o)} \boldsymbol{\lambda}_{T_o} = |P_o| - 1$.

Now, let the distribution γ be such that $\gamma_{P_o} = 1$ Then, (4.25) implies that

$$\begin{aligned}
\mathbf{y}'\boldsymbol{\lambda}_{T_o} < \left(\boldsymbol{\omega}'_{P_o}\boldsymbol{\lambda}_{T_o}\right) &= \sum_{e \in \mathcal{E}} \boldsymbol{\omega}_{P_o}(e)\boldsymbol{\lambda}_{T_o,e} && (4.26)\\
&= \sum_{e \in \mathcal{E}(P_o)} \frac{1}{|P_o| - 1} 1_{e \in T_o} && (4.27)\\
&= \frac{1}{|P_o| - 1} \sum_{e \in \mathcal{E}(P_o)} 1_{e \in T_o} && (4.28)\\
&= \frac{1}{|P_o| - 1} (|P_o| - 1) && (4.29)\\
&= 1 && (4.30)
\end{aligned}$$

[‖] We need the vectors to be essential because a priori, a feasible partition P can yield to a $\boldsymbol{\omega}_P$ that is a convex combination of the others.

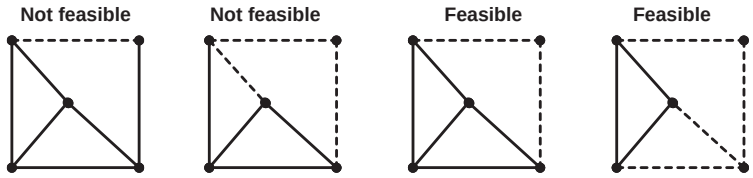


Figure 4.2: Examples of feasible and not feasible subsets of links. The chosen subset is shown in dashed line. The two subsets in the left are not feasible, the two subsets in the right are feasible.

This contradicts the hypothesis that $\mathbf{y} \in bl(P_\Lambda)$. Thus, $\mathbf{y} \in P_\Omega$ and as a consequence $bl(P_\Lambda) \subseteq P_\Omega$**.

This gives a characterization of $bl(P_\Lambda)$ by its set of vertices which are the essential vectors of the family defined as follow: *for any feasible partition $P = (\mathcal{V}_1, \ldots, \mathcal{V}_{|P|})$, form $\boldsymbol{\omega}_P$ whose entries correspond to the edges of the graph. Set entry $\boldsymbol{\omega}_P(e)$ equal to $\frac{1}{|P|-1}$ if edge $e \in \mathcal{E}$ goes from one element $\mathcal{V}_i$ of the partition to another $\mathcal{V}_j$ (i.e. if $e \in \mathcal{E}(P)$). Otherwise, set $\boldsymbol{\omega}_P(e) = 0$. More precisely, for feasible partition P and $e \in \mathcal{E}$,*

$$\boldsymbol{\omega}_P(e) = \frac{1_{e \in \mathcal{E}(P)}}{|P| - 1}. \tag{4.31}$$

From feasible partitions to feasible subsets of edges

Before applying the theory and results of the previous section to the spanning tree-link game example, let's give one more result that will enable us to switch the discussion about partition to a discussion on edges of the graph.

First, observe that every feasible partition P *uniquely* defines the set $\mathcal{E}(P)$ of edges that go across the elements of P. Also, each such edge has its two ends in two different connected components of the graph $G_{\bar{\mathcal{E}}(P)}$ obtained by removing from G, the edges in $\mathcal{E}(P)$. This means that adding any edge $e \in \mathcal{E}(P)$ to $G_{\bar{\mathcal{E}}(P)}$ will decrease the number of its connected components by 1. Let's call a subset of edges $E \subseteq \mathcal{E}$ *feasible* if *E is such that, for every edge $e \in E$, adding e to $G_{\bar{E}}$ (the graph obtained by removing from G, the edges in E) decreases its number of connected components by 1.* Figure 4.2 shows examples of feasible and not feasible subsets of a graph. The following lemma establishes the correspondence between feasible partitions and feasible subsets.

Lemma 2 *1. For every feasible partition P, the set $\mathcal{E}(P)$ of edges that go between vertices in distinct elements of P is a feasible subset.*

2. Every feasible subset is the set of edges going across the elements of some feasible partition.

Proof: The first part of the lemma follows from the discussion in the previous paragraph.
To show the second part, let E be a feasible subset and let $P_{\bar{E}}$ be the partition induced by E defined

**In fact, from Fulkerson's Theorem, we know that Ω is the matrix having as rows the essential vectors $\boldsymbol{\omega}_P$.

as follow: two vertices of the graph are in the same partition if they are connected by a path that involves only edges that are not in E. This partition is also the one obtained by removing edges in E from the graph, and considering vertices in the same connected component as being in the same partition element. We would like to show that $\mathcal{E}(P_{\bar{E}}) = E$.

First, notice that if $e = (u, v)$ is an edge having its two ends u and v in distinct elements of the partition, then it must be that $e \in E$. Otherwise, by definition, u and v would be in the same partition. Thus, we have that $\mathcal{E}(P_{\bar{E}}) \subseteq E$.

Now, assume that $e = (u, v) \in E$. Since E is feasible, u and v must be in distinct elements of the partition. Otherwise, adding e to $G_{\bar{E}}$ would not decrease its number of connected components. As a consequence, $e \in \mathcal{E}(P_{\bar{E}})$. Since e is arbitrary, this implies that $E \subseteq \mathcal{E}(P_{\bar{E}})$. Thus, $\mathcal{E}(P_{\bar{E}}) = E$. ■

One consequence of the theorem is that the number of connected components of $G_{\bar{E}}$ is equal to the size of the partition $P_{\bar{E}}$. We will denote it $Q(G_{\bar{E}}) = |P_{\bar{E}}|$.

In the remaining parts of this subsection, we will use feasible subsets (E) in place of feasible partitions (P). For instance, the vertices of the blocker $bl(P_\Lambda)$ are now denoted $\boldsymbol{\omega}_E$, for E feasible subset. Later, since our goal will be to compute critical subsets of links, we will see that we can even drop the *feasibility* requirement and consider arbitrary subset of links E.

Applying the model

We are now ready to define $\lambda(\boldsymbol{\omega}_E)$ in (4.11), $\theta(\boldsymbol{\omega}_E)$ in (4.12), and θ. For ease of notation, we write $\lambda(E) := \lambda(\boldsymbol{\omega}_E)$, and $\theta(E) := \theta(\boldsymbol{\omega}_E)$.

First, from the characterization of the vertices of the blocker of P_Λ (4.31), we have that, for feasible subset E[††]

$$\boldsymbol{\omega}_E(e) = \frac{1_{e\in E}}{Q(G_{\bar{E}}) - 1}, \quad \text{and} \quad \sum_{e\in\mathcal{E}} \boldsymbol{\omega}_E(e) = \frac{|E|}{Q(G_{\bar{E}}) - 1}, \tag{4.32}$$

so that

$$\frac{\boldsymbol{\omega}_E(e)}{\sum_{e\in\mathcal{E}} \boldsymbol{\omega}_E(e)} = \frac{1_{e\in E}}{|E|}. \tag{4.33}$$

[††]Again $\boldsymbol{\omega}_E$ needs to be essential to be a vertex of $bl(P_\Lambda)$.

Next, we have for a given feasible subset E that,

$$\begin{aligned}
\lambda(E) &= \min_{T\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} \frac{1_{e\in E}}{|E|} 1_{e\in T} \right) && (4.34)\\
&= \frac{\min_{T\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} 1_{e\in E} 1_{e\in T} \right)}{|E|} && (4.35)\\
&= \frac{\min_{T\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} 1_{e\in E\cap T} \right)}{|E|} && (4.36)\\
&= \frac{\min_{T\in\mathcal{T}} \left(|E\cap T| \right)}{|E|} && (4.37)\\
&= \frac{\mathcal{M}(E)}{|E|}, && (4.38)
\end{aligned}$$

where in the last equality we have defined $\mathcal{M}(E)$ as

$$\mathcal{M}(E) = \min_{T\in\mathcal{T}} \left(|E\cap T| \right). \tag{4.39}$$

$\theta(E)$ is given by

$$\begin{aligned}
\theta(E) &= \lambda(E) - \sum_{e\in\mathcal{E}} \frac{1_{e\in E}}{|E|} \boldsymbol{\mu}(e) && (4.40)\\
&= \frac{\mathcal{M}(E)}{|E|} - \frac{\boldsymbol{\mu}(E)}{|E|} && (4.41)
\end{aligned}$$

Finally, from (4.32), the distribution induced by a feasible subset E is given by

$$\beta_E(e) = \frac{1_{e\in E}}{|E|}, \quad \forall e \in \mathcal{E}. \tag{4.42}$$

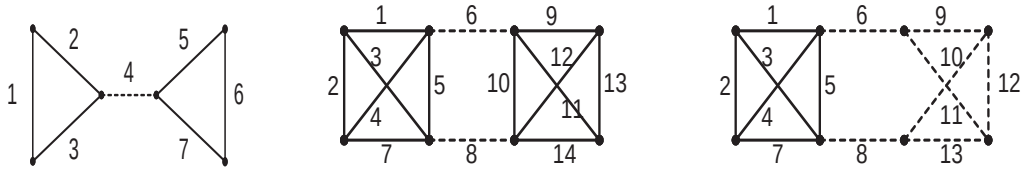
This distribution is uniform in E.

A feasible subset E of links is said to be *critical* if

$$\theta(E) = \max_{\tilde{E}:\text{feasible}} \left(\frac{\mathcal{M}(\tilde{E}) - \boldsymbol{\mu}(\tilde{E})}{|\tilde{E}|} \right). \tag{4.43}$$

We let θ denote the LHS of the above equation and use it as the *vulnerability* metric of the graph. We also define $\mathcal{C}$ to be the family of critical subsets of links.

Remark:
As was hinted earlier, in the definition of critical subset we do not need the *feasibility* requirement of E. In fact, we can show that (see appendix A.2) a non-feasible subset can never be critical. As of such, in the sequel, we will drop the term "feasible" and consider any arbitrary subset of edges $E \subseteq \mathcal{E}$.



(a) Graph with bridge (b) Network where minimum cutset is critical (c) Minimum cutset is not critical

Figure 4.3: Illustrative network examples where the attack cost $\boldsymbol{\mu} = 0$. Example 4.3(a) is a network that contains a bridge. A bridge is always a critical set. The network in 4.3(b) is an example of graph where the minimum cutset (links 6,8) corresponds to a critical subset. Example 4.3(c) shows a graph where the minimum cutset is not critical.

Some examples of graph

The examples shown in Figure 4.3 illustrate the definitions discussed above. In these figures, we have assumed that $\boldsymbol{\mu} = 0$ so that a subset of edges is critical if the ratio $\frac{\mathcal{M}(E)}{|E|}$ is maximal.

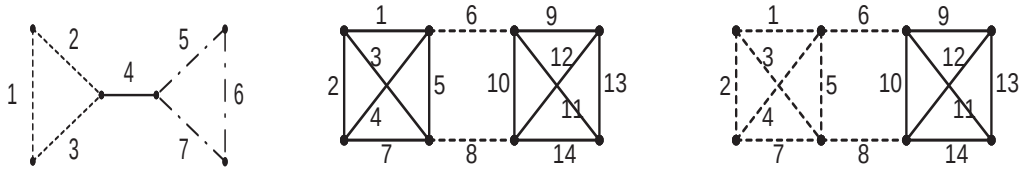
For the network in Figure 4.3(a), all spanning trees must go through the middle link, so that $\theta(E) = 1$ if E is the set with only that link. That set is critical. In general, an edge that must be part of every spanning tree is called a *bridge*. Also, it is not difficult to verify that if $\boldsymbol{\mu} = 0$ the vulnerability of a subset E is equal to the maximum value of 1 if and only if E is only composed of bridges.

The graph in Figure 4.3(b) contains 8 nodes and 14 links. It has one minimum cutset composed of the links 6 and 8. If $E = \{6, 8\}$, then any spanning tree contains at least one link in E. Thus, $|T \cap E| \geq 1$ for any tree T. Furthermore, there exists T such that $T \cap E = \{6\}$. Thus, $\mathcal{M}(E) = 1$, giving a vulnerability of $\theta(E) = 1/2$. This is the maximum vulnerability of this graph, which implies that $E = \{6, 8\}$ is a critical subset.

The set $\{6, 8\}$ is not the only critical subset. For instance, if we consider the set of all links $E = \mathcal{E}$, then $|T \cap E| = n - 1 = 7$ for any tree T because any spanning tree contains $n - 1$ links. This set is also critical because $\theta(E) = \frac{7}{14} = 1/2$. Another critical subset is $E = \{1, 2, 3, 4, 5, 6, 7, 8\}$. If $E = \{1, 2, 4\}$, choosing $T = \{3, 6, 7, 8, 9, 13, 14\}$ gives $T \cap E = \emptyset$. Hence, $\mathcal{M}(E) = 0$.

The minimum cutset of a graph is not always critical. In Figure 4.3(c) if $E = \{6, 8\}$ then $\theta(E) = 1/2$. However, choosing $E = \{6, 8, 9, 10, 11, 12, 13\}$ gives $\theta(E) = 4/7 > 1/2$. One can verify that 4/7 is the maximum achievable ratio $\frac{\mathcal{M}(E)}{|E|}$ implying that $E = \{6, 8, 9, 10, 11, 12, 13\}$ is critical and $E = \{6, 8\}$ is not.

Figure 4.4 shows examples where the attack costs are positive. The network in Figure 4.4(a) has a vector of attack cost $\boldsymbol{\mu} = [0.5, 0.5, 0.5, 2, 0.5, 0.5, 0.5]$. It contains a bridge that has a relatively high cost of attack ($\boldsymbol{\mu}(4) = 2$). As a consequence it is not critical. There are two critical subsets $E_1 = \{1, 2, 3\}$ and $E_2 = \{5, 6, 7\}$. This example illustrates the impact of the attack cost. When a



(a) Graph with bridge (b) Network where minimum cutset is critical (c) Minimum cutset is not critical

Figure 4.4: Illustrative network examples for $\boldsymbol{\mu} > 0$. Example 4.4(a) is a network that contains a bridge. The vector of attack costs is $\boldsymbol{\mu} = [0.5, 0.5, 0.5, 2, 0.5, 0.5, 0.5]$. There are 2 critical sets: $E_1 = \{1, 2, 3\}$ and $E_2 = \{5, 6, 7\}$. The bridge is neither critical nor it does belong to a critical subset. In the network in 4.4(b) the vector attack costs is $\boldsymbol{\mu} = [5, 3, 3, 5, 5, 4, 3, 3, 5, 5, 4, 5, 5, 3]/14$. In this case the minimum cutset (links 6,8) corresponds to a critical subset. Example 4.4(c) shows a graph where the minimum cutset is not critical. In this case $\boldsymbol{\mu} = [2, 5, 1, 2, 1, 1, 6, 5, 3, 7, 1, 4, 3, 6]/21$.

link is too costly to attack, it becomes less critical.

Figures 4.4(b) and 4.4(c) show the same network topology with different costs of attack. In the first one, the attack costs are $\boldsymbol{\mu} = [5, 3, 3, 5, 5, 4, 3, 3, 5, 5, 4, 5, 5, 3]/14$. For these values of the costs of attack, the minimum cutset of the graph (links 6 and 8) is critical. When the attack costs are equal to $\boldsymbol{\mu} = [2, 5, 1, 2, 1, 1, 6, 5, 3, 7, 1, 4, 3, 6]/21$, then the minimum cutset is no longer critical. It has vulnerability $\theta(6, 8) = \frac{1-(4+3)/14}{2} = 1/4$. One critical subset of the graph is given by the set $E = \{1, 2, 3, 4, 5, 6, 7, 8, \}$. Its vulnerability is $\theta(E) = 0.3631$.

Nash equilibrium theorem applied to the spanning tree game

For this example, the Nash equilibrium theorem 4 can be written as follow.

1. If $\theta = \max_{E \subseteq \mathcal{E}} \left(\frac{\mathcal{M}(E) - \boldsymbol{\mu}(E)}{|E|} \right) \leq 0$, then the attacker will opt to not launch an attack. The equilibrium strategy $(\boldsymbol{\alpha}_T,\ T \in \mathcal{T})$ for the defender is such that

$$\sum_{T \ni e} \boldsymbol{\alpha}_T \leq \boldsymbol{\mu}(e), \quad \forall e \in \mathcal{E}. \tag{4.44}$$

   The corresponding payoff is 0 for both players.

2. If $\theta \geq 0$, then for every probability distribution $(\gamma_E, E \in \mathcal{C})$ on the set $\mathcal{C}$ of critical subsets, the attacker's strategy $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ defined by

$$\boldsymbol{\beta}(e) = \sum_{E \in \mathcal{C}} \gamma_E \frac{1_{e \in E}}{|E|} \tag{4.45}$$

is in Nash equilibrium with any strategy $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ of the defender that satisfies the following properties:

$$\begin{cases} \sum_{T \ni e} \boldsymbol{\alpha}_T - \boldsymbol{\mu}(e) = \theta \ \text{ for all } \ e \in \mathcal{E} \ \text{ such that } \ \boldsymbol{\beta}(e) > 0. \\ \sum_{T \ni e} \boldsymbol{\alpha}_T - \boldsymbol{\mu}(e) \leq \theta \ \text{ for all } \ e \in \mathcal{E}. \end{cases} \tag{4.46}$$

Furthermore, there exists at least one such strategy $\boldsymbol{\alpha}$.

The corresponding payoffs are θ for the attacker, and $r(\gamma)$ for the defender, where

$$r(\gamma) := \sum_{E \in \mathcal{C}} \gamma_E \frac{\mathcal{M}(E)}{|E|}. \tag{4.47}$$

3. If $\boldsymbol{\mu} = 0$, then every Nash equilibrium pair of strategies for the game is of this type.

Next, we analyze the NE theorem. We will separately discuss the case $\boldsymbol{\mu} = 0$ and the case $\boldsymbol{\mu} > 0$.

Analyzing the NE: case $\boldsymbol{\mu} = 0$

Notice that if $\boldsymbol{\mu} = 0$, we have $\theta > 0$ and as a consequence the attacker will always launch an attack. In this case, one can refine the defender's payoff.

$$\begin{cases} \sum_{T \ni e} \boldsymbol{\alpha}_T = \theta \ \text{ for all } \ e \in \mathcal{E} \ \text{ such that } \ \boldsymbol{\beta}(e) > 0. \\ \sum_{T \ni e} \boldsymbol{\alpha}_T \leq \theta \ \text{ for all } \ e \in \mathcal{E}. \end{cases} \tag{4.48}$$

where, $\theta = \max_{E \subseteq \mathcal{E}} \left(\frac{\mathcal{M}(E)}{|E|} \right)$.

- The attack only focuses on critical subsets by taking convex combination of uniform strategies on critical subsets. All edges in a given critical subsets are attacked with the same probability. This uniformity of attack comes from the geometry of the spanning tree polyhedron and its blocker as we have seen earlier.
- If $\gamma_{E_c} = 1$ for a some critical subset E_c, we have that the corresponding attack is to target uniformly links in E_c. The defense strategy should verify $\sum_{T \ni e} \boldsymbol{\alpha}_T \leq \frac{\mathcal{M}(E_c)}{|E_c|}$ for all $e \in \mathcal{E}$, and equality holds for each $e \in E_c$. Also, by the Nash equilibria conditions it must be that for any spanning tree T

$$\sum_{e \in T} \boldsymbol{\beta}(e) = \sum_{e \in T} \frac{1_{e \in E_c}}{|E_c|} = \frac{|E_c \cap T|}{|E_c|} \geq \frac{\mathcal{M}(E_c)}{|E_c|}. \tag{4.49}$$

The minimum value in the equation above is achieved for each T for which $\boldsymbol{\alpha}_T > 0$.
Since the value of the game is equal to $\frac{\mathcal{M}(E_c)}{|E_c|}$, we have that $\mathcal{M}(E_c) = \min_T (|E \cap T|) = |E_c \cap T|$. In other words, the defender will select only spanning trees that cross the critical

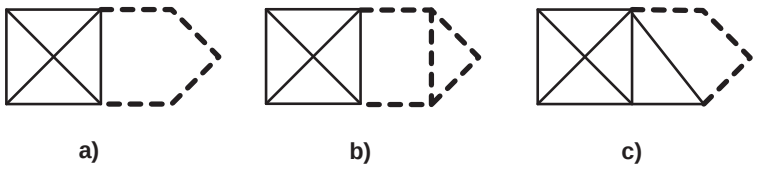


Figure 4.5: Critical subset and topology design. Graphs (b) and (c) are two different ways of adding a link to graph (a) which have a vulnerability of 3/4. If it is added as in (b), then the vulnerability is $\frac{3}{5}$. If it is done as in (c), the vulnerability is $\frac{2}{3} > \frac{3}{5}$, which is leads to a less robust network.

subset in the minimum number of links. Furthermore, the sum ($\sum_{T \ni e} \boldsymbol{\alpha}_T$) of the probability assigned to the trees crossing each link $e \in E_c$ is the same for all links in the critical subset. This sum is equal to θ, the vulnerability of the subset E_c. For any other link, this sum should be less than θ.

- From the second part of the theorem we see that every critical subset supports some Nash equilibrium (for instance the critical subset attack equilibrium). Conversely, from the third part of the theorem, if $\boldsymbol{\mu} = 0$ each support of NE for the attacker is a union of some critical subsets.

- Knowing the critical subsets (the weakest points of the network) is important for the network manager. The example in Figure 4.5 is an illustration. Consider the network in Figure 4.5.(a) whose vulnerability is equal to $\frac{3}{4}$. In all these figures, the critical subset is represented by the dashed edges. Suppose that the network manager has an extra link to add to this network and would like to know the optimal way to add this link. If the additional link is put in the position as in Figure 4.5.(b), then the vulnerability of the graph becomes $\frac{3}{5} < \frac{3}{4}$ (the graph is always less vulnerable with an additional link). If instead the link is added as in Figure 4.5.(c), the vulnerability of the graph is $\frac{2}{3} > \frac{3}{5}$ leading to a less robust network.

- The notion of graph vulnerability introduced in this study has been previously (with some differences) defined in a related but slightly different context. In [84], Gusfield discussed the consequences of Tutte [85] and Nash-Williams' [86] theorem and was particularly interested in the maximum number (M) of edge-disjoint spanning trees of a graph G. Two spanning trees of G are called disjoint if they have no edge in common.

  Gusfield showed that

$$M = \min_{E \subseteq \mathcal{E}} \lfloor \frac{|E|}{Q(G_{\bar{E}}) - 1} \rfloor \ , \tag{4.50}$$

  where $G_{\bar{E}}$ is the graph resulting from deleting the edges in E from G, and $Q(G_{\bar{E}})$ is the number of connected components in $G_{\bar{E}}$. $\bar{E}$ denotes the complement of E in $\mathcal{E}$.

  The quantity $\sigma(G) = \min_{E \subseteq \mathcal{E}} \left(\frac{|E|}{Q(G_{\bar{E}}) - 1} \right)$ was then used as a measure of the *invulnerability* of the graph, i.e. the smaller this is the more vulnerable the graph is, in the sense of Gusfield. In that paper, any minimizing set for this quantity was interpreted as *a set of edges whose*

removal from G maximizes the number of additional components created, per edge removed. The main question that was asked then was whether there exists a polynomial time algorithm to compute $\sigma(G)$.

Cunningham provided such an algorithm in [87]. Considering $\sigma(G)$ as the *strength* of G, he defined (in a non-game theoretic setting) an *optimal attack* problem as well as a *network reinforcement* problem. The optimal attack problem consists of computing the strength of G and determining a minimizing set. Cunningham considered edge-weighted graphs, with edge j having strength s_j; the strength of the graph is defined as $\sigma(G) = \min_{E \subseteq \mathcal{E}} \left(\frac{\sum_{j \in E} s_j}{Q(G_{\bar{E}})-1} \right)$, which corresponds to the invulnerability defined by Gusfield when $s_j = 1$ for all $j \in \mathcal{E}$. The network reinforcement problem of [87] is related to minimizing the cost of increasing the strengths of individual edges in order to achieve a target strength for the graph. For details, see [87].

Using *polymatroid theory* and *network flow analysis*, Cunningham provided polynomial time algorithmic solutions to both problems. In section 4.6, we discuss this algorithm in the context of the spanning tree game.

Recently, a paper by Catlin *et al.* [88] generalizes Gusfield's notion of invulnerability by imposing bounds on the number of connected components, $Q(G_{\bar{E}})$.

In our study, the critical subsets have been found to correspond to (basis of) Nash equilibria of a *quasi* zero-sum game. It is to be noticed that our definition of vulnerability verifies $\theta = \sigma(G)^{-1}$. To see that, one needs to show that,

Lemma 3 *For any $E \subseteq \mathcal{E}$,*

$$\mathcal{M}(E) = Q(G_{\bar{E}}) - 1. \tag{4.51}$$

Proof Sketch: The ideas in the proof are as follows. Consider the different connected components of the graph when the edges in E are removed. Any spanning tree of the original graph has to connected those components, and this connection is done by only using edges in E. Since there are $Q(G_{\bar{E}})$ connected components, one needs *exactly* $Q(G_{\bar{E}})$-1 to connect them in a cycle-free way. ■

It is interesting to note that, despite the fact that this metric ($\theta = \max_E \left(\frac{\mathcal{M}(E)}{|\bar{E}|} \right)$) is more refined than the *edge connectivity* (i.e. size of minimum cutset), it has largely not been used in the graph theory community. One reason suggested by Gusfield is the complexity of its computation. As was stated earlier, Cunningham [87] has subsequently provided a polynomial time algorithm to compute θ as well as a minimizing subset.

Our result shows that, in a environment where the adversary is cognitive, θ is indeed the appropriate metric of graph vulnerability.

Analyzing the NE: case $\boldsymbol{\mu} > 0$

We discuss the NE theorem for $\boldsymbol{\mu} > 0$ by considering a game on the graph shown in Figure 4.6. Table 4.1 shows the parameters and results of the game. The first column shows different values

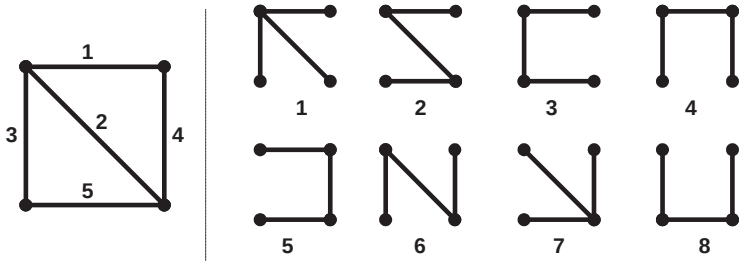


Figure 4.6: Example of graph and its spanning trees. The left figure is the original graph with the 5 edges labeled with their number. The right figures are the 8 spanning trees of the graph also labeled with their numbers. In the payoff column, L is the defender's loss while R is the attacker's reward.

Table 4.1: Game with positive attack cost played for different values of the cost of attack $\boldsymbol{\mu}$.

Attack Cost $\boldsymbol{\mu}$	Critical Set E_c	Vulnerability θ	Nash Equilibria $\boldsymbol{\alpha}$	Nash Equilibria $\boldsymbol{\beta}$	Payoffs (L,R)
[3 3 5 3 4]/5	(1,4)	-0.1	[0 0 2 0 0 1 0 2]/5	No Attack	$(0,0)$
			[0 0 2 1 0 0 0 2]/5	No Attack	$(0,0)$
[5 2 3 2 3]/5	(1-5)&(2-4)	0	[0 2 1 2 0 0 0 0]/5	[1 1 1 1 1]/5	$(0.6,0)$
			[0 2 1 2 0 0 0 0]/5	[0 1 1 1 1]/4	$(0.5,0)$
			[0 2 1 2 0 0 0 0]/5	No Attack	$(0,0)$
[5 4 2 4 2]/8	(3,5)	0.25	[0 1 0 0 1 2 0 0]/4	[0 0 1 0 1]/2	$(0.5,0.25)$
			[0 1 0 1 0 1 1 0]/4	[0 0 1 0 1]/2	$(0.5,0.25)$
[4 3 2 4 3]/8	(1-5)	0.2	[0 11 1 5 11 12 0 0]/40	[1 1 1 1 1]/5	$(0.6,0.2)$
			[0 11 1 16 0 1 11 0]/40	[1 1 1 1 1]/5	$(0.6,0.2)$
			[1 10 1 16 0 0 12 0]/40	[1 1 1 1 1]/5	$(0.6,0.2)$

of the attack costs $\boldsymbol{\mu}$ and the second column shows the corresponding critical subset(s). The third column displays the vulnerability of the graph. For each vector of attack costs, we compute the Nash equilibria of the game. The next two columns of the table show the Nash equilibrium strategies, respectively $\boldsymbol{\alpha}$ for the network manager, and $\boldsymbol{\beta}$ for the attacker. The equilibrium payoffs are displayed in the last column. In all equilibria, we have chosen the distribution γ_{E_c} to only focus on a particular critical subset (the ones shown on the table). Note that we have not shown all Nash equilibria.

- The first game considers a case where $\boldsymbol{\mu} = [3\ 3\ 5\ 3\ 4]/5$. Here, edge 3 has a particularly high cost (equal to the cost of a tree). In this case, the vulnerability of the graph ($\theta = -0.1$) is negative and the attacker does not make any gain by attacking. Her best strategy is to "*not attack*" and the network manager chooses a tree according to a distribution $\boldsymbol{\alpha}$ that satisfies (4.44). There exist many such distributions $\boldsymbol{\alpha}$; two of which are shown in the table. Since there is no attack, each player gets a payoff of zero.

  This game models scenarios where attacking requires so much investment from the attacker that it is not worth doing it. The network manager needs to randomize his choice of trees to deter the attacker from attacking. In fact, if the network were to pick a fixed tree, then the attacker could get a positive reward by attacking the cheapest link (of cost 3/5) of that tree. In other word, the randomization is necessary for the NE to hold.

- In the next game (second row of the table), the cost of attack is $\boldsymbol{\mu} = [5\ 2\ 3\ 2\ 3]/5$. In this case, the maximum attack reward is exactly equal to zero, and it can be achieved by several attack strategies as can be seen in the table (column 5). Although the attacker cannot gain by launching an attack, the damage she can cause to the network varies depending on the attack she launches.

  This game illustrates the importance of knowing the *type/nature* of an opponent. For example, if the attacker is a competitor who also wants to maximize the loss to the network, then, she will likely attack a link at random with the same probability (which gives a loss of 0.6). However, if the attacker is just interested in her own payoff, then she will probably not launch an attack.

- From these two examples and the first part of the theorem, one can infer that if the network manager is able to influence the attack costs $\boldsymbol{\mu}$, for example making the links harder to attack by investing on security (physical protection, Firewalls, Intrusion Prevention Systems (IPS) to avoid Denial of Service (DoS), etc...), then he can deter the attacker from attacking. This can be done by investing on the links to the point that $\mathcal{M}(E) \leq \boldsymbol{\mu}(E)$ for all subsets of edges $E \subseteq \mathcal{E}$. One can compute the optimal investment by solving an *optimal reinforcement* like problem. The network reinforcement problem of [87] is related to minimizing the price of increasing the cost of attack of individual edges in order to achieve a target vulnerability (here 0) for the graph. For details, see [87]. If the cost of attack can be estimated by some means, this can be a very good candidate for preventive security.

- The last two games are examples where the maximum attack reward is strictly positive. In the first one, the attacker only targets the links that are less costly which turn out to be the

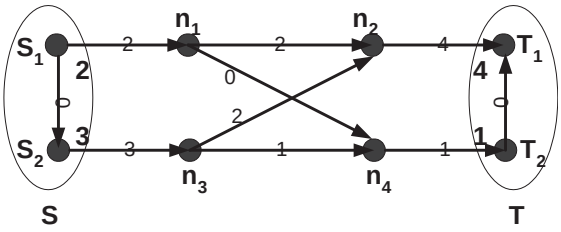


Figure 4.7: A demand-supply network. Sources $S = \{S_1, S_2\}$ produces 5 units of goods. Destinations $T = \{T_1, T_2\}$ needs 5 units of good. The amount of goods produced by the sources (S_1, S_2) is $(2, 3)$. The destination nodes (T_1, T_2) need $(4, 1)$. An example of feasible flow is shown in the figure with the amount of goods carried by each arc is the value marked on the arcs.

minimum cutset of the graph (seen by the attacker). In the second example, the minimum cut seen by the attacker corresponds to links 3 and 5. However, the attack's reward is maximized by targeting all the links with the same probability as it is shown in the table.

- The algorithm mentioned in the case $\boldsymbol{\mu} = 0$ can also be used to compute the vulnerability as well as a critical subset for the case of positive cost of attack. We will discuss this in section 4.6.

4.4.2 Examples 2: Un-capacitated supply demand networks

Most of the results discussed in these example are based on Fulkerson [89].
Let $(\mathcal{V}, \mathcal{A})$ be an arbitrary network with node-set $\mathcal{V}$ and arc-set $\mathcal{A}$. Arcs are considered to be directed edges and able to carry infinite amount of goods (un-capacitated network). Let some nonempty subset S of $\mathcal{V}$ be considered "source" nodes, and some nonempty subset T of $\mathcal{V}$ be considered "destinations" nodes, where $S \cap T = \emptyset$. With each node $x \in S$ we associate a nonnegative number $s(x)$, the "supply" at x, and with each node $x \in T$ we associate a nonnegative number $d(x)$, the "demand" at x. Throughout this examples, we assume that the demand is exactly equal to the supply, i.e.

$$\sum_{x \in S} s(x) = \sum_{x \in T} d(x). \tag{4.52}$$

Figure 4.7 shows an example of demand supply network.

Definition 2 *A* feasible flow *for this system is a function* $f : \mathcal{A} \to \mathbb{R}_+$ *such that*

$$f(x, \mathcal{V}) - f(\mathcal{V}, x) = s(x) \qquad \textit{for all} \quad x \in S \tag{4.53}$$
$$f(\mathcal{V}, x) - f(x, \mathcal{V}) = d(x) \qquad \textit{for all} \quad x \in T \tag{4.54}$$
$$f(x, \mathcal{V}) - f(\mathcal{V}, x) = 0 \qquad \textit{for all} \quad x \notin S \cup T \tag{4.55}$$

We let $\mathcal{F}$ denotes the set of feasible flows.

In other terms, a feasible flow is an assignment of values to the arcs that satisfies the conservation of flow. In Figure 4.7, a feasible flow is shown on the arcs of the graph.
In the equations above, $f(x, \mathcal{V})$ denotes

$$f(x, \mathcal{V}) = \sum_{\{y \in \mathcal{V} | (x,y) \in \mathcal{A}\}} f(x, y) \tag{4.56}$$

We will also use the following notations for arbitrary $X \subseteq \mathcal{V}$, $Y \subseteq \mathcal{V}$.

$$(X, Y) = \{(x, y) \in \mathcal{A} \mid x \in X,\ y \in Y\}, \tag{4.57}$$

and if g is any real-valued function defined on the arcs,

$$g(X, Y) = \sum_{(x,y) \in (X,Y)} g(x, y). \tag{4.58}$$

All data (i.e. supplies and demands) are assumed to be integers and we only consider integral feasible flows.

To put this example in the context of the blocking game, we assume that a network manager (defender) is choosing a set of arcs on which to carry positive amount of goods. To determine the amount of goods on each selected arc, we add the constraints of conservation of flows so that the defender's choice is restricted to the set of feasible flows. The defender would like to select a set of arcs (feasible flow) to maximize the amount of goods transferred from a set of sources to a set of destinations. We let $f(a)$ be the quantity of goods that flow f carries on arc a.

An attacker is selecting an arc to attack in order to minimize the amount of goods transferred. The attacker pays a cost $\boldsymbol{\mu}(a)$ to successfully disrupt the transport on arc a. She also has the option of not attacking. If flow f is selected and arc a is targeted, the defender loses $f(a)$ and the attacker gets $f(a) - \boldsymbol{\mu}(a)$.

We model this interaction as a game and consider mixed strategy Nash equilibria where the attacker chooses a distribution $(\boldsymbol{\alpha}_f,\ f \in \mathcal{F})$ on the set $\mathcal{F}$ of feasible flows, and the attacker, a distribution $(\boldsymbol{\beta}_a,\ a \in \mathcal{A})$ on the $\mathcal{A}$ of arcs. The model and results in sections 4.2 and 4.3 can be used to analyze the game.

Remark:
Since flows are mostly seen as functions (rather than vectors), we have adopted the 'function notation' in these examples. We will also keep the usual network flow notation: f to designate a flow and a for an *arc*. For instance, instead of the *math-bold* $(\boldsymbol{\lambda}_{T,e})$ to designate the payoff of the choice pair (T, e), we will use $f(a)$ to refer to the payoff of the game when the defender chooses the feasible flow f and the attacker the arc a. We have also preferred $\mathcal{F}$ and $\mathcal{A}$ to $\mathcal{T}$ and $\mathcal{E}$ to designate the sets of flows and arcs. Also, we use F to denote the payoff matrix (which used to be Λ).

The Flow polyhedron P_F and its blocker $bl(P_F)$

Let F be the matrix whose columns are indexed by the arcs of the network and whose rows are indexed by the integral feasible flows, with entry $f(a)$ representing the amount that flow f assigned to arc a[‡‡]. Let P_F be the polyhedron associated with F. We are interested in characterizing P_F and its blocker $bl(P_F)$.

The following theorem by Fulkerson and Weinberger [89] describes the polyhedron P_F and gives the generator matrix of its blocker $bl(P_F)$.

Theorem 5 *Let F be the matrix of integral feasible flows in an un-capacitated supply-demand network $(\mathcal{V}, \mathcal{A})$ with integral-valued supply and demand functions, $s(\cdot)$ and $d(\cdot)$, respectively. Then the polyhedron P_F is described by*

$$P_F = \left\{ \mathbf{x} \in \mathbb{R}_+^{|\mathcal{A}|} \mid \sum_{a \in (X, \bar{X})} \mathbf{x}_a \geq d(\bar{X}) - s(\bar{X}), \quad \forall X \subseteq \mathcal{V} \text{ s.t. } d(\bar{X}) - s(\bar{X}) \geq 1 \right\}. \tag{4.59}$$

Here, $d(\bar{X})$ is the total demand of the nodes in $\bar{X}$, and $s(\bar{X})$ is the corresponding total supply; $d(\bar{X}) - s(\bar{X})$ is the *excess* demand in $\bar{X}$.

This description tells that P_F is the set of nonnegative vector $\mathbf{x}$ such that the sum of the entries of $\mathbf{x}$ corresponding to links going from $X \subseteq \mathcal{V}$ to its complement $\bar{X}$ is greater than the excess of demand in $\bar{X}$, for any X for which the excess demand in $\bar{X}$ is greater than 1 unit. In other words, a feasible flow should be able to compensate any excess of demand.

From the characterization above, Fulkerson and Weinberger [89] describe the blocker $bl(P_F)$ as follow. Let H be the matrix whose columns represent the arcs of the network, having a row $\boldsymbol{\omega}_X$ for each $X \subseteq \mathcal{V}$ such that $d(\bar{X}) - s(\bar{X}) \geq 1$. Set $\boldsymbol{\omega}_X(a) = 1/(d(\bar{X}) - s(\bar{X}))$ for each arc $a \in (X, \bar{X})$ and zero for other entries. Then, the essential rows of F and H form a blocking pair of matrices. Hence, the vertices of the blocking polyhedron are given by

$$\boldsymbol{\omega}_X(a) = \frac{1}{d(\bar{X}) - s(\bar{X})} 1_{a \in (X, \bar{X})}, \quad \forall a \in \mathcal{A}, \tag{4.60}$$

for all $X \subseteq \mathcal{V}$ verifying $d(\bar{X}) - s(\bar{X}) \geq 1$. This implies that

$$\boldsymbol{\omega}_X(\mathcal{A}) = \sum_{a \in \mathcal{A}} \boldsymbol{\omega}_X(a) = \frac{|(X, \bar{X})|}{d(\bar{X}) - s(\bar{X})}, \tag{4.61}$$

where $|(X, \bar{X})|$ is the number of arcs going from X to $\bar{X}$. Hence, we have that,

$$\frac{\boldsymbol{\omega}_X(a)}{\boldsymbol{\omega}_X(\mathcal{A})} = \frac{1_{a \in (X, \bar{X})}}{|(X, \bar{X})|} \quad \forall a \in \mathcal{A}. \tag{4.62}$$

[‡‡]Note that F may not be proper; also note that F may have no rows (if a feasible flow does not exist).

Applying the model

We can now write the expressions of $\lambda(\boldsymbol{\omega}_X)$ in (4.11), $\theta(\boldsymbol{\omega}_X)$ in (4.12), and θ for this example. We simplify the notation and write $\lambda(X) := \lambda(\boldsymbol{\omega}_X)$, and $\theta(X) := \theta(\boldsymbol{\omega}_X)$.

The quantity $\lambda(X)$ can be written as

$$\begin{aligned}
\lambda(X) &= \min_f \left(\sum_{a \in (X,\bar{X})} \frac{f(a)}{|(X,\bar{X})|} \right) && (4.63)\\
&= \frac{\min_f \left(\sum_{a \in (X,\bar{X})} f(a) \right)}{|(X,\bar{X})|} && (4.64)\\
&= \frac{\min_f \left(f(X,\bar{X}) \right)}{|(X,\bar{X})|} && (4.65)\\
&= \frac{\mathcal{M}(X)}{|(X,\bar{X})|}, && (4.66)
\end{aligned}$$

where in the last equation, we have defined $\mathcal{M}(X) = \min_f \left(f(X,\bar{X}) \right)$. It is the minimum amount of goods that any feasible flow will carry on the arcs going from X to its complement $\bar{X}$. From this, we see that $\lambda(X)$ is the minimum average flow per link going from X to $\bar{X}$.

The vulnerability $\theta(X)$ of $X \subseteq \mathcal{V}$ is given by

$$\begin{aligned}
\theta(X) &= \frac{\mathcal{M}(X)}{|(X,\bar{X})|} - \sum_{a \in (X,\bar{X})} \frac{\boldsymbol{\mu}_a}{|(X,\bar{X})|} && (4.67)\\
&= \frac{\mathcal{M}(X)}{|(X,\bar{X})|} - \frac{\sum_{a \in (X,\bar{X})} \boldsymbol{\mu}_a}{|(X,\bar{X})|} && (4.68)\\
&= \frac{\mathcal{M}(X)}{|(X,\bar{X})|} - \frac{\boldsymbol{\mu}(X,\bar{X})}{|(X,\bar{X})|} && (4.69)
\end{aligned}$$

It is is the minimum average flow per link going from X to $\bar{X}$ minus the average cost of attacking links in $(X,\bar{X})$.

A subset of nodes X_c is critical if

$$\theta(X_c) = \max_{X \subseteq \mathcal{V} : d(\bar{X}) - s(\bar{X}) \geq 1} \left(\theta(X) \right). \tag{4.70}$$

We let θ (the vulnerability of the network) be the maximum value in the above equation, and we define $\mathcal{X}_c$ to be the set of all critical subsets of nodes.

Finally, the distribution $\boldsymbol{\beta}_X(a)$ induced by $(X,\bar{X})$ is given by

$$\boldsymbol{\beta}_X(a) = \frac{1_{a \in (X,\bar{X})}}{|(X,\bar{X})|}, \quad \forall a \in \mathcal{A} \tag{4.71}$$

Next, we illustrate the model and the results by some examples of demand-supply networks. We will separately consider the case where there is no cost of attack, and the case where the attacker "spends" a positive amount of effort to disrupt the network.

Nash equilibrium theorem for the supply-demand network game

Recall that

$$\theta = \max_{X \subseteq \mathcal{V}: d(\bar{X}) - s(\bar{X}) \geq 1} \left(\frac{\mathcal{M}(X) - \boldsymbol{\mu}(X, \bar{X})}{|(X, \bar{X})|} \right). \tag{4.72}$$

Applying Theorem 4 to the un-capacitated demand-supply game, we get the following.

1. If $\theta \leq 0$, then the attacker will opt to not launch an attack. The equilibrium strategy $(\boldsymbol{\alpha}_f,\ f \in \mathcal{F})$ for the defender is such that

$$\sum_{f \in \mathcal{F}} \boldsymbol{\alpha}_f f(a) \leq \boldsymbol{\mu}(a), \quad \forall a \in \mathcal{A}. \tag{4.73}$$

   The corresponding payoff is 0 for both players.

2. If $\theta \geq 0$, then for every probability distribution $(\gamma_X, X \in \mathcal{X}_c)$ on the set $\mathcal{X}_c$ of critical subsets of nodes, the attacker's strategy $(\boldsymbol{\beta}(a), a \in \mathcal{A})$ defined by

$$\boldsymbol{\beta}(a) = \sum_{X \in \mathcal{X}_c} \gamma_X \frac{1_{a \in (X, \bar{X})}}{|(X, \bar{X})|} \tag{4.74}$$

   is in Nash equilibrium with any strategy $(\boldsymbol{\alpha}_f, f \in \mathcal{F})$ of the defender that satisfies the following properties:

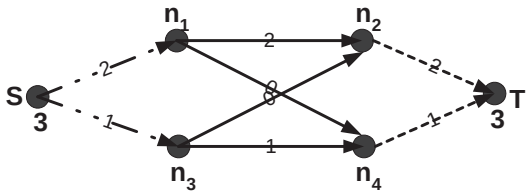
$$\begin{cases} \sum_{f \in \mathcal{F}} \boldsymbol{\alpha}_f f(a) - \boldsymbol{\mu}(a) = \theta \ \text{ for all } \ a \in \mathcal{A} \ \text{ such that } \ \boldsymbol{\beta}(a) > 0. \\ \sum_{f \in \mathcal{F}} \boldsymbol{\alpha}_f f(a) - \boldsymbol{\mu}(a) \leq \theta \ \text{ for all } \ a \in \mathcal{A}. \end{cases} \tag{4.75}$$

   Furthermore, there exists at least one such strategy $\boldsymbol{\alpha}$.

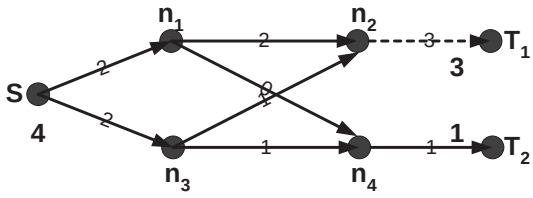
   The corresponding payoffs are θ for the attacker, and $r(\gamma)$ for the defender, where

$$r(\gamma) := \sum_{X \in \mathcal{C}} \gamma_X \frac{\mathcal{M}(X)}{|(X, \bar{X})|}. \tag{4.76}$$

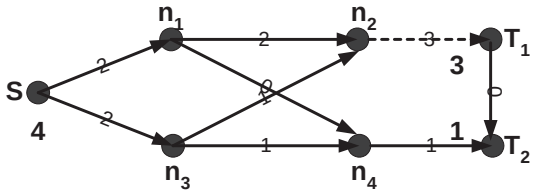
3. If $\boldsymbol{\mu} = 0$, then every Nash equilibrium pair of strategies for the game is of this type.



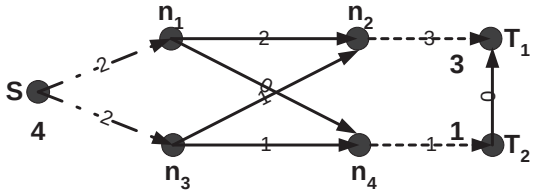
(a) A demand-supply network with one source and one destination. The total demand (supply) is 3. There are two critical subsets $X = \{S\}$ and $X = \{S, n_1, n_2, n_3, n_4\}$ which correspond to the 2 minimal cutsets of the network. The vulnerability of the network is 2.



(b) A demand-supply network with 2 bridges. Total demand (supply) is equal to 4. The critical subset is $X = (S, n_1, n_2, n_3, n_4, T_2)$ which corresponds to the bridge (n_1, T_1). The vulnerability of the network is equal to 3.



(c) Demand-Supply network with one additional route from T_1 to T_2. Total demand (supply) is equal to 4. There is only one bridge (n_2, T_1), whose corresponding set $X = \{S, n_1, n_2, n_3, n_4, T_2\}$ is critical. The vulnerability is equal to 3.



(d) Demand-Supply chain with one additional route from T_2 to T_1. Total demand (supply) is equal to 4. There is only one bridge (n_4, T_1), corresponding to subset $X = \{S, n_1, n_2, n_3, n_4, T_1\}$. This set is not critical. There are two critical subsets: $X = S$ and $X = \{S, n_1, n_2, n_3, n_4\}$. The vulnerability of the network is 2.

Figure 4.8: Illustrative demand-supply network examples. For each graph, a feasible flow is given by the values on the arcs. The arcs $(X, \bar{X})$ corresponding to critical subset X, are the dashed (dash-dotted) lines.

Examples and discussions: case case $\boldsymbol{\mu} = 0$

In these subsection, we assume that the attack cost $\boldsymbol{\mu}$ is equal to zero so that a subset of vertices $X \subseteq \mathcal{V}$ is critical if it maximizes the ratio $\frac{\mathcal{M}(X)}{|(X,\bar{X})|}$, where again $\mathcal{M}(X) = \min_f \big(f(X,\bar{X})\big)$.

- Let's first consider a network flow problem with one source (S) and one destination (T), and where the supply and demand functions are equal to 1: $s(S) = d(T) = 1$. An example is the graph in 4.8(a) with unit demand and supply.

  In this case, any $X \subseteq \mathcal{V}$ that verifies $d(\bar{X}) - s(\bar{X}) \geq 1$ must be an S-T cut. Indeed, if the source is in X, then the destination cannot be in X (otherwise the difference $d(\bar{X}) - s(\bar{X})$ would have been zero), and vice-versa. Inversely, any S-T cut groups the nodes into two disjoint sets; one containing S (say X) and the other containing T ($\bar{X}$). Also, since only one unit of good is transferred from S to T, any feasible flow has to satisfy $f(X,\bar{X}) = 1$ (only one unit is going across X and $\bar{X}$). This implies that a subset X_c is *critical* if and only if

$$\theta(X_c) = \max_{X\subseteq\mathcal{V}:d(\bar{X})-s(\bar{X})\geq 1} \left(\frac{1}{|(X,\bar{X})|}\right) = \min_{X\subseteq\mathcal{V}:d(\bar{X})-s(\bar{X})\geq 1} \big(|(X,\bar{X})|\big)\,. \tag{4.77}$$

  Thus, in this case, the critical subsets of nodes correspond to the S-T cuts of minimum size i.e. the *minimum cutsets*.

  In the NE of this game, the attacker's mixed strategy equilibria are the convex combinations of uniform attacks on the edges of the minimum cutsets. The defender's strategy should satisfy $\sum_f \boldsymbol{\alpha}_f 1_{f(a)>0} \leq \max_X \left\{\frac{1}{|(X,\bar{X})|},\ d(\bar{X}) - s(\bar{X}) \geq 1\right\}$, for all arcs $a \in \mathcal{A}$; and equality holds for each arc a that is targeted with positive probability.

- Figure 4.8(a) shows a demand-supply network with one source and one destination. The total demand is equal to the total supply and is equal to 3. The critical subsets of the graph $X_1 = \{S\}$ and $X_2 = \{S, n_1, n_2, n_3, n_4\}$ correspond to its 2 minimum cutsets (shown in dashed and dash-dotted lines). The vulnerability of the graph is equal to $\theta = \frac{\mathcal{M}(X)}{|(X,\bar{X})|} = 3/2$.

  The attacker's mixed strategy Nash equilibria have the form $\gamma\frac{1}{2}(X_1,\bar{X}_1) + (1-\gamma)\frac{1}{2}(X_2,\bar{X}_2)^*$. For example, $\gamma = 1$ gives a NE where the attacker targets each arc of the first minimum cutset with probability 1/2. The defender chooses a feasible flow with probability $\boldsymbol{\alpha}$ such that the sum $\sum_f \boldsymbol{\alpha}_f f(a)$ is equal to $\frac{3}{2}$ for each arc a that is attacked with positive probability; and for any other arc, the sum is less than $\frac{3}{2}$.

- Consider the graphs shown in Figures 4.8(b), 4.8(c), and 4.8(d). In all these networks, the source S produces 4 units of goods and destination T_1 needs 3 units while T_2 needs 1 unit.

  The network in Figure 4.8(b) contains two bridges: destination T_1 is reachable only via the arc (n_2, T_1), while access to T_2 is only via arc (n_4, T_2). Since a bridge belongs to each path

*In this compact notation, $\gamma\frac{1}{2}(X_1,\bar{X}_1)$ means that each of the two links in $(X_1,\bar{X}_1)$ is attacked with probability $\gamma\frac{1}{2}$.

that reaches its corresponding destination, all bridges are a priori vulnerable links of the graph. For instance (n_2, T_1) must carry any traffic going from $X = \{S, n_1, n_2, n_3, n_4, T_2\}$ to $\bar{X} = \{T_1\}$. Since node T_1 needs 3 units of goods, any flow will assign 3 units to arc (n_2, T_1). This implyies that $\mathcal{M}(X) = 3$ and $\theta(X) = 3$. This is the maximum vulnerability of this network implying that $X = \{S, n_1, n_2, n_3, n_4, T_2\}$ is the critical subset of the network.

In all NE, the attacker attack arc (n_2, T_1) with probability 1, and the defender chooses tree according to a distribution $\boldsymbol{\alpha}$ that satisfies: $\sum_f \boldsymbol{\alpha}_f f(a) \leq 3$ for all arcs $a \in \mathcal{A}$. Equality is satisfied on arc (n_2, T_1).

Figure 4.8(c) shows the same network as in the previous with one additional link from T_1 to T_2. The network now contains only one bridge (n_2, T_1). The same arguments used in the previous example give that subset $X = \{S, n_1, n_2, n_3, n_4, T_2\}$, corresponding to the only bridge of the network, is the critical subset of the network. Its vulnerability is 3.

Compared to the previous case, we see that having an additional link (added in the way given here) does not yield to a stronger network. This suggests that if the defender wants to benefit from adding a new link to the network, he has to determine a proper way to do it. An arbitrary addition of link might not provide any benefit.

In the graph shown in Figure 4.8(d) the additional links goes from T_2 to T_1. The network still contains a bridge because the only way to reach T_2 is via (n_4, T_2). However, that bridge is not a critical link. The most vulnerable arcs are the ones going from the critical subset $X = S$ to $\bar{X}$, and those going from the critical subset $X = \{S, n_1, n_2, n_3, n_4\}$ to $\bar{X}$.

In fact, since bridge (n_4, T_2) (corresponding to $X_1 = \{S, n_1, n_2, n_3, n_4, T_1\}$) is the only way to reach T_2, any feasible flow will assign positive load to that arc. The minimum amount of traffic carried over (n_4, T_2) on any feasible flow is equal to 1. Thus, $\mathcal{M}(X_1) = 1$ and $\theta(X_1) = 1$. On the other hand, if $X = S$, any feasible flow will carry the total amount of goods on the arcs $(X, \bar{X})$. As a consequence, $\mathcal{M}(X) = 4$ and the vulnerability $\theta(X) = 4/2 = 2$. The is the maximum vulnerability implying that X is critical. Similarly, subset $X = \{S, n_1, n_2, n_3, n_4\}$ is critical.

Adding the extra link in the way described here, strictly decreases the vulnerability of the network (as opposed to the previous case where the additional link was not optimally added). This phenomenon was already observe in the spanning tree game. It illustrates the importance of knowing the critical sets when designing a network topology.

- As was mentioned earlier, $\mathcal{M}(X)$ is the minimum amount of goods that any feasible flow will carry from X to its complement. For any target subset of nodes X, the defender is trying to find a feasible flow that achieves this minimum. The attacker is choosing a subset X that maximizes the fraction $\frac{\mathcal{M}(X)}{|(X,\bar{X})|}$. In other words, the attacker chooses a subset X that maximizes the minimum average traffic per link from X to $\bar{X}$. This maximum can be achieved by targeting set of edges that separate all sources to all destinations. This is the case in Figures 4.8(a) and 4.8(d) where for example the edges going from S to the rest of the network are critical. The maximum can also be achieved by attacking edges that separate a

given set of destination nodes to the rest of the network as is the case in examples 4.8(b) and 4.8(c).

Examples and discussions: case case $\boldsymbol{\mu} > 0$

Now we assume that the attack cost $\boldsymbol{\mu}$ is a positive vector. The networks in Figures 4.9 illustrate this case. In all networks, the values of the attack costs are placed close to the corresponding arcs.

In Figure 4.9(a) the costs of attack are relatively high. The vulnerability θ of the network is equal to $-1/2$ and is achieved by subsets $X = \{S\}$ (dash-dotted lines) and $X = \{S, n_1, n_2, n_3, n_4\}$ (dashed lines). Since $\theta < 0$, the attacker will not launch an attack. The defender needs to choose feasible flows according to a distribution $\boldsymbol{\alpha}$ that satisfies

$$\sum_f \boldsymbol{\alpha}_f f(a) \leq \boldsymbol{\mu}(a), \quad \forall\, a \in \mathcal{A}. \tag{4.78}$$

We know from the theorem that such distribution exists.

It is worth noting that although the network contains a bridge (link (n_4, T_2)), the attacker will not attack (even not the bridge). In fact, in this example, the cost of successfully attacking the bridge (which is 2) is greater than the reward that the attacker expects from such attack (which is equal to 1).

Figure 4.9(b) is an example of network with vulnerability zero. This maximum vulnerability is achieved at $X = \{S\}$ (dash-dotted lines) and $X = \{S, n_1, n_2, n_3, n_4, T_2\}$ (dashed lines). A certain number of interesting phenomenon happen in this example.

First, the bridge (n_4, T_2) is so costly that it is not part of the edges induced by any critical subset of nodes. The are two critical subsets and each will lead to an zero average attack reward for the attacker. An attack that focuses on the links going from $X = \{S\}$ to $\bar{X}$ is preventing the traffic to reach the destinations, without any distinction of the destinations. It results to an average attack loss of $\frac{\mathcal{M}(X)}{|(X,\bar{X})|} = \frac{4}{2} = 2$ for the defender, and in average 2 units of goods will reach the destinations. An attack that focuses on the edges entering T_1 (dashed lines) is preventing traffic to reach T_1. The average loss of such attack is $\frac{\mathcal{M}(X)}{|(X,\bar{X})|} = \frac{3}{2}$, and an average of 2.5 units of goods will always reach the destinations; 1 unit of it will always go to node T_2.

This phenomenon was already seen in the spanning tree game. Although the attacks give the same payoff to the attacker, their impacts on the defender are different. One interesting related question is which equilibrium will be played, or how can the defender force the second equilibrium which is clearly better for him. Our model does not permit the study of equilibrium selection. Other inputs and parameters need to be considered to answer to those questions.

Figure 4.9(c) is an example of network with positive vulnerability. The network also contains a single critical subset $X = \{S, n_1, n_2, n_3, n_4\}$ and its vulnerability is equal to 1. The attacker will attack the two links in $(X, \bar{X})$ each with probability $1/2$. Her average net reward is 1. The defender

will choose feasible flows according to a distribution $\boldsymbol{\alpha}$ that satisfies

$$\sum_f \boldsymbol{\alpha}_f f(a) - \boldsymbol{\mu}(a) = \theta, \ \forall\ a \in (X, \bar{X}) \tag{4.79}$$

$$\sum_f \boldsymbol{\alpha}_f f(a) - \boldsymbol{\mu}(a) \leq \theta, \ \forall\ a \in \mathcal{A}. \tag{4.80}$$

The average attack loss is equal to 2 and in average 2 units of goods will reach the destinations.

4.5 Proof of the NE Theorem

In this section we provide a proof of the Nash equilibrium theorem presented in section 4.3.

4.5.1 "No Attack" Option

We start by showing that if $\theta \leq 0$ "*No Attack*" and $\boldsymbol{\alpha}$ verifying (4.14) are best responses to each other.

Best Responses

First, notice that if the attacker chooses "No Attack", then any $\boldsymbol{\alpha}$ will result to the minimum loss of zero for the defender (in particular the one given in the theorem). Now, assume that $\boldsymbol{\alpha}$ satisfies (4.14). Then, "No Attack" is a *dominant* strategy for the attacker. In fact, the expected attack reward is

$$R(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{e \in E} \boldsymbol{\beta}(e) \left(\bar{\lambda}_{\boldsymbol{\alpha}}(e) - \boldsymbol{\mu}(e) \right), \tag{4.81}$$

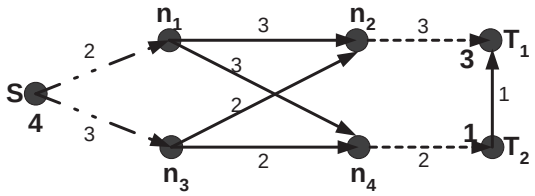
which is less than zero if $\bar{\lambda}_{\boldsymbol{\alpha}}(e) - \boldsymbol{\mu}(e) \leq 0, \ \forall\ e \in \mathcal{E}$, and for any $\boldsymbol{\beta}$. On the other hand, zero reward can always be achieved by playing $e_\emptyset$. As a consequence, not attacking is a best response to $\boldsymbol{\alpha}$ satisfying (4.14).

Existence of the Equilibrium Distribution $\boldsymbol{\alpha}$

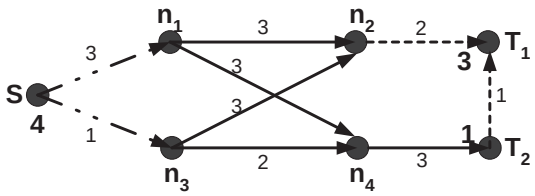
Now, we need to prove the existence of a distribution $\boldsymbol{\alpha}$ that satisfies (4.14) whenever $\theta \leq 0$. To summarize, we are looking for $\boldsymbol{\alpha}$ verifying:

$$\boldsymbol{\alpha} : \begin{cases} \boldsymbol{\alpha} \geq 0 \\ \mathbf{1}_{\mathcal{I}}' \boldsymbol{\alpha} = 1 \\ \Lambda' \boldsymbol{\alpha} \leq \boldsymbol{\mu}, \end{cases} \tag{4.82}$$

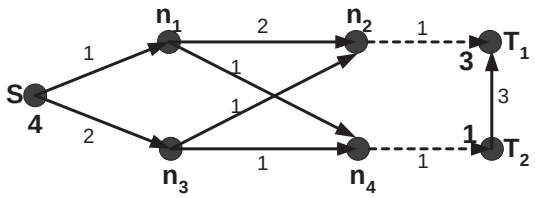
We first show the following lemma.



(a) Example of demand/supply network with high attack costs. The vulnerability of the network is $-1/2$ and the attacker does not launch an attack. Two subsets achieve the maximum vulnerability of $-1/2$. The corresponding edges are shown in dashed and dash-dotted.



(b) Example of demand/supply network with attack costs that give zero payoff to the attacker. Vulnerability is 0, and corresponding critical set are shown in dashed and dash-dotted.



(c) Example of demand/supply with positive cost of attack and positive vulnerability. There is only one critical subset shown by the dashed line. The vulnerability of the network is $\frac{4-2}{2} = 1$.

Figure 4.9: Illustrative demand-supply network examples for $\boldsymbol{\mu} > 0$. For each graph, the value of the attack cost is marked close to the arc. The arcs $(X, \bar{X})$ corresponding to critical subset X, are the dashed (dash-dotted) lines.

Lemma 4

$$\theta \leq 0 \quad \Rightarrow \quad \boldsymbol{\mu} \in P_\Lambda. \tag{4.83}$$

Proof: For this we show that if $\theta \leq 0$, then $\boldsymbol{\mu}'\boldsymbol{\omega} \geq 1$, for any vertex $\boldsymbol{\omega}$ of $bl(P_\Lambda)$. As a result of this, $\boldsymbol{\mu}$ belongs to the blocker of $bl(P_\Lambda)$ which is P_Λ.

If $\theta \leq 0$, then we have that, for all $\boldsymbol{\omega} \in \Omega$,

$$\sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}\boldsymbol{\mu}(e) \geq \min_{T\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}\lambda_{T,e} \right). \tag{4.84}$$

Or, equivalently

$$\boldsymbol{\mu}'\boldsymbol{\omega} \geq \min_{T\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} \boldsymbol{\omega}(e)\lambda_{T,e} \right), \quad \text{for all} \quad \boldsymbol{\omega} \in \Omega. \tag{4.85}$$

Now, since $\boldsymbol{\omega} \in bl(P_\Lambda)$ we have that

$$\sum_{e\in\mathcal{E}} \boldsymbol{\omega}(e)\lambda_{T,e} \geq 1, \tag{4.86}$$

which implies that $\boldsymbol{\mu} \cdot \boldsymbol{\omega} \geq 1, \ \forall\ \boldsymbol{\omega} \in \Omega$, or equivalently $\boldsymbol{\mu} \in P_\Lambda$.
Using Lemmas 1 and 4, we conclude that the value of the following LP is greater than 1.

$$\begin{aligned} &\text{Maximize } \mathbf{1}_{\mathcal{T}}\mathbf{x} \\ &\text{subject to } \Lambda'\mathbf{x} \leq \boldsymbol{\mu}, \quad \text{and} \quad \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{4.87}$$

Construct $\boldsymbol{\alpha}$ satisfying (4.14) by normalizing any solution of this LP. ■

4.5.2 The "Always Attack" option

As in the previous section, we will first argue that the strategies given in the theorem are best responses to each other, then we show the existence of a distribution $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ that satisfies (4.16). We start by the following lemma.

Lemma 5 *If $\theta > 0$, then "No Attack" is a* strictly dominated *strategy for the attacker.*

Note that if $\theta \geq 0$, all the steps of the proof still hold. However, "No Attack" will only be weakly dominated, and as seen in the previous case, there will exist equilibrium for which the attacker will opt to not launch an attack. In other words, if $\theta = 0$ there exist equilibria for which $\boldsymbol{\beta}_{e_\emptyset} = 1$, as well as equilibria for which $\boldsymbol{\beta}_{e_\emptyset} = 0$.

Proof: Suppose that $\boldsymbol{\omega} \in \Omega_{max}$ is a critical vertex of $bl(P_\Lambda)$ and let $\boldsymbol{\beta} = (\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}, e \in \mathcal{E})$. We will show that the attacker can achieve positive reward by playing $\boldsymbol{\beta}$ (independently of $\boldsymbol{\alpha}$). To see this,

first notice that since $\theta(\boldsymbol{\omega}) = \theta > 0$ for all $\boldsymbol{\omega} \in \Omega_{max}$, we have that

$$\lambda(\boldsymbol{\omega}) = \min_{T \in \mathcal{T}} \left(\sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) > \sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\mu}(e) = \frac{\boldsymbol{\mu}'\boldsymbol{\omega}}{\boldsymbol{\omega}(\mathcal{E})} \tag{4.88}$$

Playing the strategy $\boldsymbol{\beta}$, the attacker's expected reward against any defense strategy $\boldsymbol{\alpha}$ is given by

$$\begin{aligned}
R(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \left(\sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \lambda_{T,e} - \boldsymbol{\mu}(e) \right) \\
&= \sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} - \sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\mu}(e) \right) && (4.89) \\
&\geq \sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \left(\lambda(\boldsymbol{\omega}) - \sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\mu}(e) \right) && (4.90) \\
&= \sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \left(\lambda(\boldsymbol{\omega}) - \frac{\boldsymbol{\mu}'\boldsymbol{\omega}}{\boldsymbol{\omega}(\mathcal{E})} \right) && (4.91) \\
&= \lambda(\boldsymbol{\omega}) - \frac{\boldsymbol{\mu}'\boldsymbol{\omega}}{\boldsymbol{\omega}(\mathcal{E})} && (4.92) \\
&> 0, && (4.93)
\end{aligned}$$

where in (4.90) we use the definition of $\lambda(\boldsymbol{\omega})$; (4.93) is implied by (4.88). ■

As a consequence of this lemma, we conclude that if $\theta > 0$, then the attacker will never play the "No Attack" (i.e. $e_\emptyset$) strategy.

Best Responses

Given the set of critical vertices Ω_{max} and $\boldsymbol{\alpha}$ satisfying (4.16), any distribution $\boldsymbol{\beta}$ of the form $\boldsymbol{\beta}(e) = \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}$ for some distribution $\gamma = (\gamma_{\boldsymbol{\omega}},\ \boldsymbol{\omega} \in \Omega_{max})$, achieves a reward of θ. This is the maximum possible reward that the attacker can get. To see this, observe that for any $\boldsymbol{\beta}$,

$$R(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \left(\bar{\lambda}_{\boldsymbol{\alpha}}(e) - \boldsymbol{\mu}(e) \right) \leq \sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \theta \leq \theta. \tag{4.94}$$

The upper bound of θ is achieved by any $\tilde{\boldsymbol{\beta}} = (\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}, e \in \mathcal{E})$ and $\boldsymbol{\omega} \in \Omega_{max}$, because for any such $\tilde{\boldsymbol{\beta}}$,

$$\begin{aligned}
R(\boldsymbol{\alpha}, \tilde{\boldsymbol{\beta}}) &= \sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \left(\sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \lambda_{T,e} - \boldsymbol{\mu}(e) \right) && (4.95)\\
&= \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} - \sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\mu}(e) \right) && (4.96)\\
&\geq \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\lambda(\boldsymbol{\omega}) - \sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\mu}(e) \right) && (4.97)\\
&= \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \theta && (4.98)\\
&= \theta, && (4.99)
\end{aligned}$$

where (4.97) uses the definition of $\lambda(\boldsymbol{\omega})$, and in (4.98) we use the fact that $\boldsymbol{\omega} \in \Omega_{max}$.
As a consequence, any distribution of the form $(\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}, e \in \mathcal{E})$ for $\boldsymbol{\omega} \in \Omega_{max}$ is a best response and any convex combination of those distributions is also a best response.

Now assume that $\boldsymbol{\beta}$ is given as in (4.15) for some distribution $(\gamma_{\boldsymbol{\omega}},\ \boldsymbol{\omega} \in \Omega_{max})$. Then, the distribution $(\boldsymbol{\alpha}_T,\ T \in \mathcal{T})$ in (4.16) achieves a loss of $r(\gamma) = \sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \lambda(\boldsymbol{\omega})$. This is the minimum possible loss. To see this, note that, for any $\boldsymbol{\alpha}$, the expected loss seen by the defender is given by

$$\begin{aligned}
L(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e) \lambda_{T,e} \right) && (4.100)\\
&= \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e\in\mathcal{E}} \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \right) \lambda_{T,e} \right) && (4.101)\\
&= \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \left(\sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) \right) && (4.102)\\
&\geq \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \lambda(\boldsymbol{\omega}) \right) && (4.103)\\
&= \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T r(\gamma) && (4.104)\\
&= r(\gamma). && (4.105)
\end{aligned}$$

The lower bound $r(\gamma)$ can be achieved by choosing $\boldsymbol{\alpha}$ such that $\sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \lambda_{T,e} = \theta + \boldsymbol{\mu}(e)$ for each $e \in \mathcal{E}$ such that $\boldsymbol{\beta}(e) > 0$ (the existence of such $\boldsymbol{\alpha}$ is shown in the second part of the theorem).

This can be seen by rewriting $L(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \left(\sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \lambda_{T,e} \right) && (4.106) \\ &= \sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \left(\theta + \boldsymbol{\mu}(e) \right) && (4.107) \\ &= \theta + \boldsymbol{\beta}' \boldsymbol{\mu} = r(\gamma). && (4.108) \end{aligned}$$

The last equality above is justified by

$$\begin{aligned} \theta + \boldsymbol{\beta}' \boldsymbol{\mu} &= \theta + \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \left(\frac{1}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\omega}' \boldsymbol{\mu} \right) - r(\gamma) + r(\gamma) && (4.109) \\ &= \theta + \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \left(\frac{1}{\boldsymbol{\omega}(\mathcal{E})} \boldsymbol{\omega}' \boldsymbol{\mu} - \lambda(\boldsymbol{\omega}) \right) + r(\gamma) && (4.110) \\ &= \theta + \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \left(-\theta \right) + r(\gamma) && (4.111) \\ &= \theta - \theta + r(\gamma) && (4.112) \\ &= r(\gamma), && (4.113) \end{aligned}$$

where in (4.109) and (4.110) we use the definitions of $\boldsymbol{\beta}$ and $r(\gamma)$ respectively and write the summation over $\mathcal{E}$ as a product of a row vector and a column vector. In 4.110 we have also used the definition of $\theta(\boldsymbol{\omega})$ for a critical vertex $\boldsymbol{\omega} \in \Omega_{max}$.

From this analysis, we see that for the set Ω_{max} of critical vertices, the distributions given in the theorem are best responses to each other, as a consequence, they form a Nash equilibria under the assumption that $\boldsymbol{\alpha}$ exists. Such existence is shown in the next section.

Existence of the Equilibrium Distribution $\boldsymbol{\alpha}$

We claim that for any $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ of the form in the statement (4.15) of Theorem 4, we can find an associated $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ of the form (4.16).

Theorem 6 *Assume that $\theta \geq 0$ and let Ω_{max} be the set of critical vertices. Let $\mathbf{x}^*$ be the solution of the following LP:*

$$\begin{aligned} &\textit{Maximize } \mathbf{1}_{\mathcal{T}}' \mathbf{x} \\ &\textit{subject to } A' \mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{4.114}$$

where $\mathbf{b} = \theta(E) \mathbf{1}_{\mathcal{E}} + \boldsymbol{\mu}$. Then,

a) $\mathbf{1}_{\mathcal{T}}' \mathbf{x}^* \leq 1$;

b) $\mathbf{1}'_{\mathcal{T}}\mathbf{x}^* \geq 1$*;*

c) $A'\mathbf{x}^*(e) = \mathbf{b}(e), \forall e \in \mathcal{E}$ *for which* $\boldsymbol{\beta}(e) > 0$.

As a consequence, $\mathbf{x}^*$ *satisfies (4.16) and implies the existence of* $\boldsymbol{\alpha}$.

Proof: a) To prove that $\mathbf{1}'_{\mathcal{T}}\mathbf{x}^* \leq 1$, we first observe that

$$\begin{aligned}
\boldsymbol{\beta}'\Lambda'\mathbf{x} &= \sum_{T\in\mathcal{T}} \mathbf{x}_T \left(\sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e)\lambda_{T,e} \right) && (4.115)\\
&= \sum_{T\in\mathcal{T}} \mathbf{x}_T \left(\sum_{e\in\mathcal{E}} \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \right) \lambda_{T,e} \right) && (4.116)\\
&= \sum_{T\in\mathcal{T}} \mathbf{x}_T \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \left(\sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) \right) && (4.117)\\
&\geq \sum_{T\in\mathcal{T}} \mathbf{x}_T \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \lambda(\boldsymbol{\omega}) \right) && (4.118)\\
&= \sum_{T\in\mathcal{T}} \mathbf{x}_T r(\gamma) && (4.119)\\
&= r(\gamma)\mathbf{1}'_{\mathcal{T}}\mathbf{x} && (4.120)
\end{aligned}$$

On the other hand, from the constraints $\Lambda'\mathbf{x} \leq \mathbf{b} = \theta\mathbf{1}_{\mathcal{E}} + \boldsymbol{\mu}$ and from (4.113), we have that

$$\boldsymbol{\beta}'\Lambda'\mathbf{x} \leq \boldsymbol{\beta}' \left(\theta\mathbf{1}_{\mathcal{E}} + \boldsymbol{\mu}\right) = \theta + \boldsymbol{\beta}'\boldsymbol{\mu} = r(\gamma) \tag{4.121}$$

Combining (4.120) and (4.113) it follows that,

$$r(\gamma)\mathbf{1}'_{\mathcal{T}}\mathbf{x} \leq \boldsymbol{\beta}'\Lambda'\mathbf{x} \leq r(\gamma) \tag{4.122}$$

Thus $\mathbf{1}'_{\mathcal{T}}\mathbf{x} \leq 1$ for all feasible $\mathbf{x}$, i.e. the value of the program is at most 1.

b) To prove that $1_{\mathcal{T}}^T \mathbf{x}^* \geq 1$, we use Lemma 1 above to claim that it suffices to verify that the vector $\mathbf{b}$ belongs to the polyhedron P_Λ. For that, we will show that $\mathbf{b}'\boldsymbol{\omega} \geq 1$ for all $\boldsymbol{\omega} \in \Omega$. In fact*,

$$\begin{aligned}
\boldsymbol{\omega}'\mathbf{b} &= \theta\boldsymbol{\omega}'\mathbf{1}_{\mathcal{E}} + \boldsymbol{\omega}'\boldsymbol{\mu} && (4.123)\\
&= \boldsymbol{\omega}(\mathcal{E})\left(\theta + \frac{1}{\boldsymbol{\omega}(\mathcal{E})}\boldsymbol{\omega}'\boldsymbol{\mu}\right) && (4.124)\\
&\geq \boldsymbol{\omega}(\mathcal{E})\left(\lambda(\boldsymbol{\omega}) - \frac{1}{\boldsymbol{\omega}(\mathcal{E})}\boldsymbol{\omega}'\boldsymbol{\mu} + \frac{1}{\boldsymbol{\omega}(\mathcal{E})}\boldsymbol{\omega}'\boldsymbol{\mu}\right) && (4.125)\\
&= \boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega}) && (4.126)\\
&= \boldsymbol{\omega}(\mathcal{E})\min_{T\in\mathcal{T}}\left(\sum_{e\in\mathcal{E}}\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}\lambda_{T,e}\right) && (4.127)\\
&= \min_{T\in\mathcal{T}}\left(\sum_{e\in\mathcal{E}}\boldsymbol{\omega}(e)\lambda_{T,e}\right) && (4.128)\\
&\geq 1 && (4.129)
\end{aligned}$$

where (4.129) follows from the fact that $\boldsymbol{\omega} \in bl(P_\Lambda)$. Indeed, if $\boldsymbol{\omega} \in bl(P_\Lambda)$, then by definition of the blocker polyhedron, $\sum_{e\in\mathcal{E}}\boldsymbol{\omega}(e)\lambda_{T,e} \geq 1$ for all $T \in \mathcal{T}$. Thus, we have that $\Omega\mathbf{b} \geq 1$, which implies that $\mathbf{b}$ belongs to the blocker of $bl(P_\Lambda)$ which is equal to P_Λ.

Now, using Lemma 1, we conclude that the value of the LP is greater than 1.
This, together with the previous part a) imply that the value of the LP is equal to 1; hence, any solution $\mathbf{x}^*$ is a probability distribution on $\mathcal{T}$.

c) We first observe from (4.120) and (4.113) that

$$\boldsymbol{\beta}'\Lambda'\mathbf{x}^* = r(\gamma)\mathbf{1}_{\mathcal{T}}\mathbf{x}^* = r(\gamma). \tag{4.130}$$

Also, $\Lambda'\mathbf{x}^* \leq \theta\mathbf{1}_{\mathcal{E}} + \boldsymbol{\mu}$ by the constraints of the primal LP above.
Now, assume that $\Lambda'\mathbf{x}^*(e) < \theta + \boldsymbol{\mu}(e)$ for some $e \in \mathcal{E}$ with $\boldsymbol{\beta}(e) > 0$. Then,

$$\begin{aligned}
\boldsymbol{\beta}'\Lambda'\mathbf{x}^* &= \sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)\Lambda'\mathbf{x}^*(e) && (4.131)\\
&< \sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)(\theta + \boldsymbol{\mu}(e)) && (4.132)\\
&= \theta + \sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)\boldsymbol{\mu}(e) && (4.133)\\
&= r(\gamma), && (4.134)
\end{aligned}$$

where the last equality is obtained by using the same arguments as in (4.109)-(4.113). This contradicts observation (4.130). As a consequence, $\Lambda'\mathbf{x}^*(e) = \theta + \boldsymbol{\mu}(e)$ for all $e \in \mathcal{E}$ with $\boldsymbol{\beta}(e) > 0$.

This ends the proof of the theorem and establishes the existence of an $\boldsymbol{\alpha}$ satisfying (4.16) for any $\boldsymbol{\beta}$ defined as in (4.15). ■

*Again we use vector product for summation.

4.5.3 Enumerating all Nash Equilibria

In this section, we consider the zero-sum game where $\boldsymbol{\mu} = 0$. In this case, since there is no cost of attack $\theta > 0$. We will show that for any strategy pair $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ and $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ that are in Nash equilibrium, it must be the case that $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ is given by

$$\boldsymbol{\beta}(e) = \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} , \tag{4.135}$$

for some probability distribution $(\gamma_{\boldsymbol{\omega}}, \boldsymbol{\omega} \in \Omega_{max})$.
As a consequence of this, we will conclude that $\boldsymbol{\alpha}$ must be in the form given in the Nash equilibrium theorem.

First, notice that since $\boldsymbol{\mu} = 0$, we have that $\lambda(\boldsymbol{\omega}) = \theta(\boldsymbol{\omega})$ (see the definitions in (4.12) and (4.11)). Next, we use the zero-sum structure of the game to observe that it has a well-defined value, which, by the second part of Theorem 4, is equal to,

$$\theta = \max_{\boldsymbol{\omega} \in \Omega} (\lambda(\boldsymbol{\omega})) = \max_{\boldsymbol{\omega} \in \Omega} \left(\min_{T \in \mathcal{T}} \left(\sum_{e \in \mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) \right). \tag{4.136}$$

Thus, we must have that, for any Nash equilibrium pair $(\boldsymbol{\alpha}, \boldsymbol{\beta})$,

$$\theta = \sum_{T \in \mathcal{T}} \sum_{e \in \mathcal{E}} \boldsymbol{\alpha}_T \boldsymbol{\beta}(e) \lambda_{T,e} = \sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \lambda_{T,e} \right) > 0. \tag{4.137}$$

From the above equation we can argue that there exists a scaling factor $\tilde{\kappa} > 0$ such that $(\tilde{\kappa}\boldsymbol{\beta}(e),\ e \in \mathcal{E})$ belongs to the blocker $bl(P_\Lambda)$, or equivalently,

$$\sum_{e \in \mathcal{E}} \tilde{\kappa} \boldsymbol{\beta}(e) \lambda_{T,e} \geq 1, \quad \text{for all} \quad T \in \mathcal{T}. \tag{4.138}$$

In fact, by letting $\boldsymbol{\alpha}_{max}$ be the maximum $\boldsymbol{\alpha}_T$, and defining $\tilde{\kappa} = \frac{N \boldsymbol{\alpha}_{max}}{\theta}$, then $(\tilde{\kappa}\boldsymbol{\beta}(e),\ e \in \mathcal{E})$ verifies (4.138). We let κ denote the smallest such scaling that works among all scalings $\tilde{\kappa} > 0$.

Also, observe that since κ is the smallest nonnegative scaling of $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ such that $(\kappa\boldsymbol{\beta}(e), e \in \mathcal{E})$ belongs to $bl(P_\Lambda)$, there must exist some $T_o \in \mathcal{T}$ for which $\sum_{e \in \mathcal{E}} \kappa \boldsymbol{\beta}(e) \lambda_{T_o,e} = 1$.
Indeed this is the case because since $\kappa\boldsymbol{\beta} \in bl(P_\Lambda)$, we have that $\sum_{e \in \mathcal{E}} \kappa \boldsymbol{\beta}(e) \lambda_{T,e} \geq 1$ for all $T \in \mathcal{T}$. If this inequality were strict for all $T \in \mathcal{T}$, then considering the strategy T_{min} that minimizes the sum $\sum_{e \in \mathcal{E}} \kappa \boldsymbol{\beta}(e) \lambda_{T,e}$ over all $\mathcal{T}$, we can construct

$$\tilde{\kappa} = \frac{\kappa}{\sum_{e \in \mathcal{E}} \boldsymbol{\beta}(e) \lambda_{T_{min},e}}. \tag{4.139}$$

$\tilde{\kappa}$ verifies $\tilde{\kappa} < \kappa$ and $\tilde{\kappa}\boldsymbol{\beta} \in bl(P_\Lambda)$. This contradicts the assumption that κ was the smallest such $\tilde{\kappa}$.

Now, we claim that:

Lemma 6 *If* $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ *and* $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ *form a NE of the game, then we can write*

$$\kappa\boldsymbol{\beta}(e) = \sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})}, \tag{4.140}$$

for some probability distribution $(\gamma_{\boldsymbol{\omega}},\ \boldsymbol{\omega} \in \Omega)$.

We delay the proof of this lemma for later.

Using this expression of $\kappa\boldsymbol{\beta}$, we can write the value of the game θ as:

$$\begin{aligned}
\theta &= \frac{1}{\kappa} \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e\in\mathcal{E}} \kappa\boldsymbol{\beta}(e)\lambda_{T,e} \right) && (4.141)\\
&= \frac{1}{\kappa} \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e\in\mathcal{E}} \left(\sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})} \right) \lambda_{T,e} \right) && (4.142)\\
&= \frac{1}{\kappa} \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \left(\frac{1}{\lambda(\boldsymbol{\omega})} \sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) \right) && (4.143)\\
&\geq \frac{1}{\kappa} \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \right) && (4.144)\\
&= \frac{1}{\kappa}, && (4.145)
\end{aligned}$$

where in (4.144) we use the fact that $\lambda(\boldsymbol{\omega}) \leq \sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}\lambda_{T,e}$ for all $T \in \mathcal{T}$.

Now, since $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is a Nash equilibrium pair, the expression $\sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e)\lambda_{T,e}$ is minimal for each $T \in \mathcal{T}$ for which $\boldsymbol{\alpha}_T > 0$. Furthermore, this minimum value is equal to θ. From this, we get that,

$$\begin{aligned}
\theta &= \min_{T\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e)\lambda_{T,e} \right) && (4.146)\\
&= \frac{1}{\kappa} \min_{T\in\mathcal{T}} \left(\sum_{e\in\mathcal{E}} \kappa\boldsymbol{\beta}(e)\lambda_{T,e} \right) && (4.147)\\
&\leq \frac{1}{\kappa}, && (4.148)
\end{aligned}$$

where in (4.148) we use the fact that the minimum in (4.147) is less than $\sum_{e\in\mathcal{E}} \kappa\boldsymbol{\beta}(e)\lambda_{T_o,e}$, which, by definition, is equal to 1. Thus, $\theta = \frac{1}{\kappa}$.

This, combined with (4.140) that we sum over $e \in \mathcal{E}$, imply:

$$\frac{1}{\theta} = \kappa = \sum_{e\in\mathcal{E}} \kappa\boldsymbol{\beta}(e) \tag{4.149}$$

$$= \sum_{e\in\mathcal{E}} \sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})} \tag{4.150}$$

$$= \sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \frac{1}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})} \sum_{e\in\mathcal{E}} \boldsymbol{\omega}(e) \tag{4.151}$$

$$= \sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \frac{1}{\lambda(\boldsymbol{\omega})}. \tag{4.152}$$

Now, recalling (4.136) that $\theta = \max_{\boldsymbol{\omega}\in\Omega}(\lambda(\boldsymbol{\omega}))$, we conclude that $\gamma_{\boldsymbol{\omega}}$ can be nonzero only for $\boldsymbol{\omega} \in \Omega$ that satisfies $\lambda(\boldsymbol{\omega}) = \max_{\tilde{\boldsymbol{\omega}}\in\Omega}(\lambda(\tilde{\boldsymbol{\omega}}))$. In other terms, $\gamma_{\boldsymbol{\omega}} > 0$ only for $\boldsymbol{\omega} \in \Omega_{max}$.
Hence, we can write

$$\boldsymbol{\beta}(e) = \sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}. \tag{4.153}$$

The last thing that remains to be shown to complete the proof of the theorem is that if $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ and $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ are in Nash equilibrium and $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ is of the form (4.15) in the statement Theorem 4, then $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ must also be of the form in the statement (4.16) of the theorem. We have already shown that for $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ of the form (4.15), we must have for *every* strategy $(\tilde{\boldsymbol{\alpha}}_T, T \in \mathcal{T})$

$$\sum_{T\in\mathcal{T}} \tilde{\boldsymbol{\alpha}}_T \sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e)\lambda_{T,e} = \sum_{T\in\mathcal{T}} \tilde{\boldsymbol{\alpha}}_T \left(\sum_{e\in\mathcal{E}} \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) \right) \tag{4.154}$$

$$= \sum_{T\in\mathcal{T}} \tilde{\boldsymbol{\alpha}}_T \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}} \left(\sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) \right) \tag{4.155}$$

$$\geq \sum_{T\in\mathcal{T}} \tilde{\boldsymbol{\alpha}}_T \left(\sum_{\boldsymbol{\omega}\in\Omega_{max}} \gamma_{\boldsymbol{\omega}}\theta \right) \tag{4.156}$$

$$= \theta, \tag{4.157}$$

where (4.156) follows from (4.136). The minimum value of θ can be achieved by choosing $\boldsymbol{\alpha}$ such that $\sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \lambda_{T,e} = \theta$ whenever $\boldsymbol{\beta}(e) > 0$. To see that, rewrite the summation as

$$\sum_{T\in\mathcal{T}} \tilde{\boldsymbol{\alpha}}_T \sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e)\lambda_{T,e} = \sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e) \left(\sum_{T\in\mathcal{T}} \tilde{\boldsymbol{\alpha}}_T \lambda_{T,e} \right), \tag{4.158}$$

and observe the claim.

The existence of such $\boldsymbol{\alpha}$ has been shown in the previous section. It also has been shown that such $\boldsymbol{\alpha}$ is a best response to $(\boldsymbol{\beta}(e), e \in \mathcal{E})$. For $\boldsymbol{\beta}$ to be a best response to $\boldsymbol{\alpha}$ (hence, the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ pair to be in Nash equilibrium), $\boldsymbol{\alpha}$ must also satisfy

$$\sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \lambda_{T,e} \leq \theta \text{ for all } e \in \mathcal{E}. \tag{4.159}$$

Suppose, on the contrary, that this is not the case (i.e. there is some $e \in \mathcal{E}$ for which $\sum_{T \in \mathcal{T}} \boldsymbol{\alpha}_T \lambda_{T,e} > \theta$). Then the attacker will prefer to switch to playing strategy e with probability 1 and receive higher reward. This violates the assumption that $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ and $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ are in Nash equilibrium. Thus, $\boldsymbol{\alpha}$ satisfies (4.159). This completes the proof of the theorem, provided that the claim in Lemma 6 can be justified. We give a proof of the lemma in the next section.

Proof of Lemma 6

The claim is that if $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ and $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ are in Nash equilibrium, and if $\kappa > 0$ denotes the smallest $\tilde{\kappa} > 0$ for which $(\tilde{\kappa}\boldsymbol{\beta}(e), e \in \mathcal{E}) \in bl(P_\Lambda)$, then we must have

$$\kappa \boldsymbol{\beta}(e) = \sum_{\boldsymbol{\omega} \in \Omega} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})}, \tag{4.160}$$

for some probability distribution $(\gamma_{\boldsymbol{\omega}}, \boldsymbol{\omega} \in \Omega)$.

Indeed, this needs a proof, because a priori we only know that we can write

$$\kappa \boldsymbol{\beta}(e) = \sum_{\boldsymbol{\omega} \in \Omega} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})} + v(e), \tag{4.161}$$

for some probability distribution $(\gamma_{\boldsymbol{\omega}}, \boldsymbol{\omega} \in \Omega)$ and some $(v(e), e \in \mathcal{E})$ such that $v(e) \geq 0$ for all $e \in \mathcal{E}$.

We now provide the proof that works as follow. We consider the expression of this form for $(\kappa\boldsymbol{\beta}(e), e \in \mathcal{E})$ for which $v(\mathcal{E}) := \sum_{e \in \mathcal{E}} v(e)$ is as small as possible. We will assume that $v(\mathcal{E}) > 0$ for this expression, and arrive at a contradiction.
Note that we must have $v(\mathcal{E}) < \kappa$ for this expression, i.e. there has to be a nontrivial "convex hull part".

First observe that for all $T \in \mathcal{T}$, we have

$$\begin{aligned}
\sum_{e\in\mathcal{E}} \boldsymbol{\beta}(e)\lambda_{T,e} &= \frac{1}{\kappa}\sum_{e\in\mathcal{E}} \kappa\boldsymbol{\beta}(e)\lambda_{T,e} && (4.162)\\
&= \frac{1}{\kappa}\sum_{e\in\mathcal{E}}\left(\sum_{\boldsymbol{\omega}\in\Omega}\gamma_{\boldsymbol{\omega}}\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})} + v(e)\right)\lambda_{T,e} && (4.163)\\
&= \frac{1}{\kappa}\sum_{\boldsymbol{\omega}\in\Omega}\gamma_{\boldsymbol{\omega}}\left(\frac{1}{\lambda(\boldsymbol{\omega})}\sum_{e\in\mathcal{E}}\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})}\lambda_{T,e}\right) + \frac{1}{\kappa}\sum_{e\in\mathcal{E}} v(e)\lambda_{T,e} && (4.164)\\
&\geq \frac{1}{\kappa}\sum_{\boldsymbol{\omega}\in\Omega}\gamma_{\boldsymbol{\omega}} + \frac{1}{\kappa}v_\lambda(T) && (4.165)\\
&= \frac{1}{\kappa} + \frac{v_\lambda(T)}{\kappa}, && (4.166)
\end{aligned}$$

where in (4.165) we use the definition of $\lambda(\boldsymbol{\omega})$, and define $v_\lambda(T) := \sum_{e\in\mathcal{E}} v(e)\lambda_{T,e}$. As a consequence, we have that

$$\min_{T\in\mathcal{T}}\left(\sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)\lambda_{T,e}\right) \geq \frac{1}{\kappa} + \min_{T\in\mathcal{T}}\left(\frac{v_\lambda(T)}{\kappa}\right) \tag{4.167}$$

Next, observe that since $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$ and $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ are in Nash equilibrium, it must be the case that $\sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)\lambda_{T,e}$ is the same for every $T \in \mathcal{T}$ such that $\boldsymbol{\alpha}_T > 0$. Also, by the same reasoning as in (4.148), we have that for all such T

$$\sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)\lambda_{T,e} = \min_{\tilde{T}\in\mathcal{T}}\left(\sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)\lambda_{\tilde{T},e}\right) \leq \frac{1}{\kappa}. \tag{4.168}$$

This, combined with (4.159), implies that $\min_{T\in\mathcal{T}}\left(\sum_{e\in\mathcal{E}}\boldsymbol{\beta}(e)\lambda_{T,e}\right) = \frac{1}{\kappa}$ and that $v_\lambda(T) = 0$ for all $T \in \mathcal{T}$ such that $\boldsymbol{\alpha}_T > 0$.

Now, let

$$\tilde{\boldsymbol{\beta}}(e) := \frac{1}{\kappa - v(\mathcal{E})}\sum_{\boldsymbol{\omega}\in\Omega}\gamma_{\boldsymbol{\omega}}\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})}\ . \tag{4.169}$$

This quantity satisfies $\sum(e)\tilde{\boldsymbol{\beta}}(e) = 1$ and $\tilde{\boldsymbol{\beta}}(e) \geq 0$ for all $e \in \mathcal{E}$. Thus, $(\tilde{\boldsymbol{\beta}}(e), e \in \mathcal{E})$ is a probability distribution on $\mathcal{E}$, and can be used as a strategy by the attacker. This can be verified by summing both sides of (4.161) over $e \in \mathcal{E}$ to get,

$$\begin{aligned}
\kappa = \sum_{e\in\mathcal{E}}\kappa\boldsymbol{\beta}(e) &= \sum_{e\in\mathcal{E}}\left(\sum_{\boldsymbol{\omega}\in B}\gamma_{\boldsymbol{\omega}}\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})} + v(e)\right) && (4.170)\\
&= \sum_{e\in\mathcal{E}}\left(\sum_{\boldsymbol{\omega}\in B}\gamma_{\boldsymbol{\omega}}\frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})}\right) + v(\mathcal{E}) && (4.171)\\
&= (\kappa - v(\mathcal{E}))\sum_{e\in\mathcal{E}}\tilde{\boldsymbol{\beta}}(e) + v(\mathcal{E}). && (4.172)
\end{aligned}$$

This last equation implies that

$$\sum_{e\in\mathcal{E}} \tilde{\boldsymbol{\beta}}(e) = \frac{\kappa - v(\mathcal{E})}{\kappa - v(\mathcal{E})} = 1. \tag{4.173}$$

For this strategy $(\tilde{\boldsymbol{\beta}})$, in response to $(\boldsymbol{\alpha}_T, T \in \mathcal{T})$, the attack reward is at least $\frac{1}{\kappa - v(\mathcal{E})}$. In fact,

$$\begin{aligned}
\sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \sum_{e\in\mathcal{E}} \tilde{\boldsymbol{\beta}}(e)\lambda_{T,e} &= \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{e\in\mathcal{E}} \left(\frac{1}{\kappa - v(\mathcal{E})} \sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})\lambda(\boldsymbol{\omega})} \right) \lambda_{T,e} \right) && (4.174)\\
&= \frac{1}{\kappa - v(\mathcal{E})} \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \left(\sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} \frac{1}{\lambda(\boldsymbol{\omega})} \left(\sum_{e\in\mathcal{E}} \frac{\boldsymbol{\omega}(e)}{\boldsymbol{\omega}(\mathcal{E})} \lambda_{T,e} \right) \right) && (4.175)\\
&\geq \frac{1}{\kappa - v(\mathcal{E})} \sum_{T\in\mathcal{T}} \boldsymbol{\alpha}_T \sum_{\boldsymbol{\omega}\in\Omega} \gamma_{\boldsymbol{\omega}} && (4.176)\\
&= \frac{1}{\kappa - v(\mathcal{E})}. && (4.177)
\end{aligned}$$

However, because of the fact that $v(T) = 0$ for all $T \in \mathcal{T}$ for which $\boldsymbol{\alpha}_T > 0$, the benefit obtained by the attacker by playing the NE strategy $(\boldsymbol{\beta}(e), e \in \mathcal{E})$ is only $\frac{1}{\kappa}$, which is strictly smaller than $\frac{1}{\kappa - v(\mathcal{E})}$ under the standing assumption that $v(\mathcal{E}) > 0$. As a consequence, if $v(\mathcal{E}) > 0$, the attacker can be better off by changing her strategy to $\tilde{\boldsymbol{\beta}}$. But this contradicts the assumption that $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ form a NE. Thus, $v(\mathcal{E}) = 0$ implying that $v(e) = 0$ for all $e \in \mathcal{E}$ as we wanted to show.

4.6 Algorithm to Compute θ and a Critical Vertex

In this section, we show how one can derive an algorithm to compute the vulnerability θ of the game, as well as a maximizing (hence critical) vertex.We will consider that the cost $\boldsymbol{\mu}$ is equal to 0. The case of non-zero $\boldsymbol{\mu}$ can be analyzed using the same techniques.

Recall that if $\boldsymbol{\mu} = 0$,

$$\theta = \max_{\boldsymbol{\omega}\in\Omega} \left(\frac{\mathcal{M}(\boldsymbol{\omega})}{\boldsymbol{\omega}(\mathcal{E})} \right), \tag{4.178}$$

where Ω is the matrix having as rows the vertices $\boldsymbol{\omega}$ of blocking polyhedron $bl(P_\Lambda)$. The function $\mathcal{M}(\cdot)$ is defined as

$$\mathcal{M}(\boldsymbol{\omega}) = \min_{T\in\mathcal{T}} \sum_{e\in\mathcal{E}} \boldsymbol{\omega}_e \boldsymbol{\lambda}_{T,e}. \tag{4.179}$$

The algorithm is inspired by Cunningham [87] who used it in a none-game theoretic context. For the sake of completeness, we discuss the algorithm here, and adapt it to the context of the games considered in the chapter. To derive the algorithm, we need the the notions of *polymatroid*, and which we present in section 4.6.2.

Table 4.2: Pseudocode: algorithm computing the value θ of the game and a critical vertex. The algorithm *CunninghamMin* is used to perform the minimization in 4.180. It returns θ and a minimizing subset $\boldsymbol{\omega}_o$

IterativeApprox
Input: set $\mathcal{E}$, vertices $\Omega = \{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_K\}$
Output: θ, $\boldsymbol{\omega}_c$: critical

```
1 begin
2   σ = M(ω_1)/ω_1(E)
3   while (1)
4      (ρ, ω) = CunninghamMin(E, Ω, σ)
5      if  ρ ≤ 0
6          return (ρ, ω)
7      else
8         σ := ρ
9       endif
10   endwhile
11 end
```

4.6.1 Deriving an algorithm

We start by noticing that:

Lemma 7 *For a fixed real number σ,*

$$\theta \leq \sigma \Leftrightarrow 0 \leq \min_{\boldsymbol{\omega} \in \Omega} \left(\sigma \boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega}) \right). \tag{4.180}$$

Proof: This can be easily verified by the following:

$$\theta \leq \sigma \Leftrightarrow \max_{\boldsymbol{\omega} \in \Omega} \left(\frac{\mathcal{M}(\boldsymbol{\omega})}{\boldsymbol{\omega}(\mathcal{E})} \right) \leq \sigma \quad \Leftrightarrow \quad \frac{\mathcal{M}(\boldsymbol{\omega})}{\boldsymbol{\omega}(\mathcal{E})} \leq \sigma, \quad \forall \ \boldsymbol{\omega} \in \Omega \tag{4.181}$$

$$\Leftrightarrow \quad 0 \leq \sigma \boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega}), \quad \forall \ \boldsymbol{\omega} \in \Omega \tag{4.182}$$

$$\Leftrightarrow \quad 0 \leq \min_{\boldsymbol{\omega} \in \Omega} \left(\sigma \boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega}) \right). \tag{4.183}$$

■

Now, we make the following observation made by Cunningham [87]. Suppose that σ is such that $\theta \geq \sigma$. Then, if $0 \leq \min_{\boldsymbol{\omega} \in \Omega} \left(\sigma \boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega}) \right)$ then $\theta \leq \sigma$, implying $\theta = \sigma$. Also, the minimizing $\boldsymbol{\omega}$ is a critical vertex. If $0 > \min_{\boldsymbol{\omega} \in \Omega} \left(\sigma \boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega}) \right)$, let $\boldsymbol{\omega}_o$ be a minimizing vertex and define a new $\sigma_o = \frac{\mathcal{M}(\boldsymbol{\omega}_o)}{\boldsymbol{\omega}_o(\mathcal{E})}$. Then, since the minimum is strictly negative, we have $\sigma_o > \sigma$. Now, one can replace σ by σ_o and repeat the same procedure. This gives an algorithm to compute θ as long as the minimization in 4.180 can be computed. The algorithm is shown in Table 4.2.

The following lemma shows how the algorithm converges.

Lemma 8 *Suppose that $\sigma \leq \theta$, and let $\boldsymbol{\omega}_o$ minimizes $\sigma\boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega})$. Let $\sigma_o = \frac{\mathcal{M}(\boldsymbol{\omega}_o)}{\boldsymbol{\omega}_o(\mathcal{E})} > \sigma$ and let $\boldsymbol{\omega}_1$ such that $\sigma_o\boldsymbol{\omega}_1(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega}_1) < 0$. Then $\mathcal{M}(\boldsymbol{\omega}_o) < \mathcal{M}(\boldsymbol{\omega}_1)$.*

To understand how the algorithm converges, first, notice that the updates of σ are driven by the minimizers of $\sigma\boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega})$. The lemma tells that at each iteration either the algorithm has found θ and a critical vertex $\boldsymbol{\omega}_o$, or the value of $\mathcal{M}(\boldsymbol{\omega}_1)$ corresponding to the new σ is *strictly* larger than $\mathcal{M}(\boldsymbol{\omega}_o)$. But, we know that, for a fixed game, $\mathcal{M}(\boldsymbol{\omega})$ can only take a finite set of values (given by the vertices of the blocker polyhedron). Thus, we have only a finite number of iterations.

The worst case happens when there is a different value of $\mathcal{M}(\boldsymbol{\omega})$ for each $\boldsymbol{\omega}$. In practice, we have much better than that. For instance, in the case of the spanning tree game studied above, we have established an equivalence between the vertices $\boldsymbol{\omega}$ and the subsets $E \subseteq \mathcal{E}$ of edges of the graph. We have also seen that $\mathcal{M}(E) = \min_T(|E \cap T|)$ which is at most $|\mathcal{V}| - 1$; $|\mathcal{V}|$ is the number of nodes in the graph. Thus in this case, the algorithm has at most $|\mathcal{V}| - 1$ iterations. For the demand-supply network game, the vertices $\boldsymbol{\omega}$ are given by the subsets of nodes $X \subseteq \mathcal{N}$ verifying $d(X) - s(X) \geq 1$. In this case, $\mathcal{M}(X) = \min_f \sum_f f(X, \bar{X})$. This is the minimum amount of goods that any feasible flow will carry along the edges in $(X, \bar{X})$. It is at most equal to the total demand $d(T)$ of the network, which hence is an upper bound to the number of iterations.

Proof:[of Lemma 8] We have that,

$$
\begin{aligned}
0 &> \sigma_o\boldsymbol{\omega}_1(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega}_1) && (4.184)\\
&= \sigma_o\left(\boldsymbol{\omega}_1(\mathcal{E}) - \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma_o} - \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma} + \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma}\right) && (4.185)\\
&= \sigma_o\left(\boldsymbol{\omega}_1(\mathcal{E}) - \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma} - \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma_o} + \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma}\right) && (4.186)\\
&\geq \sigma_o\left(\boldsymbol{\omega}_o(\mathcal{E}) - \frac{\mathcal{M}(\boldsymbol{\omega}_o)}{\sigma} - \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma_o} + \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma}\right) && (4.187)\\
&= \sigma_o\left(\frac{\mathcal{M}(\boldsymbol{\omega}_o)}{\sigma_o} - \frac{\mathcal{M}(\boldsymbol{\omega}_o)}{\sigma} - \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma_o} + \frac{\mathcal{M}(\boldsymbol{\omega}_1)}{\sigma}\right) && (4.188)\\
&= \sigma_o\left(\frac{1}{\sigma} - \frac{1}{\sigma_o}\right)\left(\mathcal{M}(\boldsymbol{\omega}_1) - \mathcal{M}(\boldsymbol{\omega}_o)\right). && (4.189)
\end{aligned}
$$

The inequality in (4.187) follows from the fact that $\boldsymbol{\omega}_o$ minimizes $\sigma\boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega})$, and (4.188) is obtained from $\sigma_o = \frac{\mathcal{M}(\boldsymbol{\omega}_o)}{\boldsymbol{\omega}_o(\mathcal{E})}$. Since $\sigma_o > \sigma$, it must be the case that $\mathcal{M}(\boldsymbol{\omega}_1) > \mathcal{M}(\boldsymbol{\omega}_o)$. ■

In general computing the minimization in (4.180) is difficult. However, if the function $\mathcal{M}(\boldsymbol{\omega})$ can be related to a *submodular* function on the subsets $E \subseteq \mathcal{E}$ of resources, one can use the techniques of submodular function minimization to derive a polynomial-time algorithm. We illustrate this in the next section for the spanning tree game.

4.6.2 Computing critical subset of edges of a graph

Recall that in the case of the spanning tree-link game, , if $\boldsymbol{\mu}$ is assumed to be zero, then

$$\theta = \max_{e \in \mathcal{E}} \left(\frac{\mathcal{M}(E)}{|E|} \right), \tag{4.190}$$

where $\mathcal{M}(E) = \min_{T \in \mathcal{T}} (|E \cap T|)$.
Also, by Lemma 7, we have that for fixed values of p and q ($\sigma := \frac{p}{q}$),

$$\theta \leq \sigma \quad \Leftrightarrow \quad 0 \leq \min_{E \subseteq \mathcal{E}} (\sigma |E| - \mathcal{M}(E)) \ . \tag{4.191}$$

We have seen in the previous section that if an oracle exists that can efficiently compute the minimization above, the algorithm in Table 4.2 finds θ and a critical subset by making at most $|\mathcal{V}| - 1$ calls to the oracle. Since in this example θ is always a ratio of the form $\frac{p}{q}$ for $0 \leq p \leq |\mathcal{V} - 1|$ and $1 \leq q \leq |\mathcal{E}|$, one can also implement a binary search algorithm, provided that the minimization above is easy to compute. We present the binary search algorithm in appendix A.1.

As stated earlier, computing the minimization in (4.191) can be very challenging. However, we will show that $\mathcal{M}(E)$ can be written in term of a submodular function $f(E)$ on the subset of resources $\mathcal{E}$. This will allow us to use a theorem of polymatroid theory that says that the minimum value of $\sigma\boldsymbol{\omega}(\mathcal{E}) - \mathcal{M}(\boldsymbol{\omega})$ is achieved at a P_f-basis of $\sigma\boldsymbol{\omega}(\mathcal{E})$ where P_f is the polymatroid associated to $f(\cdot)$ (we discuss those notions next). Finding a P_f-basis of a polymatroid can be done in polynomial time by using (in general) a matroid partition algorithm [72]. In the case of the spanning tree game discussed in this chapter, Cunningham [87] has proposed a more efficient algorithm that is based on network flow methods.
Next, we define the notions of polymatroid and discuss Cunningham's algorithm.

Polymatroids

Definition 3 *A real-valued function $f(\cdot)$, defined on subsets of $\mathcal{E}$, is called a* polymatroid function *if it verifies*

P.0: $f(\emptyset) = 0$,

P.1: If $E \subseteq F \subseteq \mathcal{E}$, then $f(E) \leq f(F)$ (i.e. $f(\cdot)$ is non-decreasing),

P.2: $f(E \cup \{e_2, e_1\}) - f(E \cup \{e_2\}) \leq f(E \cup \{e_1\}) - f(E)$ *for all* $E \subseteq \mathcal{E}$, $e_1, e_2 \in \mathcal{E}$, $e_1, e_2 \notin E$*.

*Actually, the most common definition of submodularity is as follow:
P.2(bis): If $E, F \subseteq \mathcal{E}$, then $f(E \cup F) + f(E \cap F) \leq f(E) + f(F)$.
For convenience, we have used the equivalent definition given above.

Given a polymatroid function $f(\cdot)$, the following polytope is called the polymatroid *associated to f.*

$$P_f = \left\{ \mathbf{x} \in R_+^{|\mathcal{E}|},\ \sum_{e \in E} \mathbf{x}(e) \leq f(E),\ \forall\ E \subseteq \mathcal{E} \right\}. \tag{4.192}$$

For any $\mathbf{y} \in R_+^{|\mathcal{E}|}$, $\mathbf{x} \in P_f$ is called a P_f-basis of $\mathbf{y}$ if $\mathbf{x}$ is a componentwise maximal vector of the set $\{\mathbf{x},\ \mathbf{x} \in P_f$ and $\mathbf{x} \leq \mathbf{y}\}$.

We have earlier seen that, to compute a critical subset a graph and its vulnerability θ, one needs an oracle that solves

$$\min_{E \subseteq \mathcal{E}} \left(\mathbf{y}_0(E) - \mathcal{M}(E) \right)\ , \tag{4.193}$$

where $\mathbf{y}_0 = \frac{p}{q}\mathbf{1}_{\mathcal{E}}$ for p and q given by the search algorithm, so that $\mathbf{y}_0(E) = \frac{p}{q}|E|$ for any subset of edges $E \subseteq \mathcal{E}$ of the graph.

Now, define the function $f(E) := |\mathcal{V}| - 1 - \mathcal{M}(\bar{E})$, where $\bar{E}$ is the complement of E in $\mathcal{E}$. Then,

Lemma 9 *$f(\cdot)$ is a polymatroid function.*

Proof: To see this, we show that $f(\cdot)$ satisfies the conditions of definition 3.

First, notice that, by Lemma 3, $\mathcal{M}(\bar{E}) = Q(G_E) - 1$ where $Q(G_E)$ is the number of connected component of the graph $G_E = G(\mathcal{V}, E)$ (G with only the edges in E). Thus, $f(E) = |\mathcal{V}| - Q(G_E)$. Also, notice that $Q(G_\emptyset) = |\mathcal{V}|$. Indeed, the graph G without edges is the set of nodes without any connection between them. Thus, the number of connected components is equal to the number of nodes. Hence, $f(\emptyset) = 0$.

Now, suppose that $E \subseteq \mathcal{E}$. Then adding a link to E can only decrease (or not change) the number of connected components in the graph. Hence, if $E \subseteq F$, we have that $Q(G_E) \geq Q(G_F)$, implying that $f(E) \leq f(F)$.

To verify the last condition of the definition, we just need to show that $Q(G_{E\cup\{e_2\}}) - Q(G_{E\cup\{e_2,e_1\}}) \leq Q(G_E) - Q(G_{E\cup\{e_1\}})$. But observe that,

$$0 \leq Q(G_{E\cup\{e_2\}}) - Q(G_{E\cup\{e_2,e_1\}}) \leq 1, \quad \text{and} \quad 0 \leq Q(G_E) - Q(G_{E\cup\{e_1\}}) \leq 1. \tag{4.194}$$

Thus, to verify $P.2$, it is enough to show that if $Q(G_{E\cup\{e_2\}}) - Q(G_{E\cup\{e_2,e_1\}}) = 1$, then it must be the case that $Q(G_E) - Q(G_{E\cup\{e_1\}}) = 1$. But, if $Q(G_{E\cup\{e_2\}}) - Q(G_{E\cup\{e_2,e_1\}}) = 1$, then e_1 must have connected two different connected components of the graph G with only the edges in $E \cup \{e_2\}$. If this is the case, then e_1 must also connect two different connected components of G_E, implying that $Q(G_E) - Q(G_{E\cup\{e_1\}}) = 1$. This ends the proof. ■

Using the definition of $f(\cdot)$, we can rewrite 4.191 as

$$\theta \leq \sigma \quad \Leftrightarrow \quad |\mathcal{V}| - 1 \leq \min_{E \subseteq \mathcal{E}} \left(\sigma|E| + f(\bar{E}) \right)\ . \tag{4.195}$$

The following (max-min) theorem relates the minimization above to finding P_f-basis of $\mathbf{y}_0$. The proof of the theorem can be found in [87].

Theorem 7 *Let $f(\cdot)$ be a polymatroid function on subsets of $\mathcal{E}$. Then, for any $\mathbf{y} \in R_+^{|\mathcal{E}|}$ and any P_f-basis $\mathbf{x}$ of $\mathbf{y}$, we have*

$$\mathbf{x}(\mathcal{E}) = \min\left(\mathbf{y}(E) + f(\bar{E}),\ E \subseteq \mathcal{E}\right) \ . \tag{4.196}$$

From this theorem, we see that an oracle that computes a P_f-basis of $\mathbf{y}_0 = \sigma \mathbf{1}_{\mathcal{E}}$ suffices for the minimization in (4.195). Let's see how such an oracle can be built.

The definition of P_f-basis implies a very simple method for finding a P_f-basis of any $\mathbf{y} \in R_+^{|\mathcal{E}|}$. Namely,
start with $\mathbf{x} = 0$ and successively increase each component of $\mathbf{x}$ as much as possible while still satisfying $\mathbf{x} \leq \mathbf{y}$, and $\mathbf{x} \in P_f$.

Implementing this simple and greedy algorithm might, however, not be so simple. In fact, it requires one to be able to compute, for a given $\mathbf{x} \in P_f$ and any $e \in \mathcal{E}$, the quantity

$$\epsilon_{max}(e) = \max(\epsilon \ : \ \mathbf{x} + \epsilon \mathbf{1}_e \in P_f) \ , \tag{4.197}$$

where $\mathbf{1}_e$ is the incidence vector of subset $\{e\}$. $\epsilon_{max}(e)$ is the maximum amount by which component e of $\mathbf{x}$ can be increased while keeping $\mathbf{x}$ in P_f.

Verifying that a vector $\mathbf{x}$ belongs to the polymatroid can be done using the following idea: if $\mathbf{x} \notin P_f$, then one can find a subset E for which $\mathbf{x}(E) \leq f(E)$ is violated. If $\mathbf{x} \in P_f$ and $E \in \mathcal{E}$, then any ϵ such that $\epsilon > \min_{E \subseteq \mathcal{E}} (f(E) - \mathbf{x}(E),\ e \in E)$ will send $\mathbf{x} + \epsilon \mathbf{1}_e$ out of P_f.
Also, if $\mathbf{x}$ is a P_f-basis of $\mathbf{y}$, then for any $e \in \mathcal{E}$, either $\mathbf{x}(e) = \mathbf{y}(e)$ or $\mathbf{x}(E) = f(E)$ for some subset E containing e. In fact, for all $e \in \mathcal{E}$

$$\epsilon_{max}(e) = \min\left\{\mathbf{y}(e) - \mathbf{x}(e), \min_E (f(e) - \mathbf{x}(E),\ e \in E \subseteq \mathcal{E})\right\} \ . \tag{4.198}$$

If the minimum is achieved at $\mathbf{y}(e) - \mathbf{x}(e)$, then $\mathbf{x} \leftarrow \mathbf{x} + \epsilon_{max}(e)\mathbf{1}_e$ will satisfy $\mathbf{x}(e) = \mathbf{y}(e)$. Otherwise, there exists some $E_e \ni e$, such that $\mathbf{x}(E_e) = f(E_e)$ (E_e is said to be *tight*). Letting $\bar{E}_U = \bigcup_e E_e$, and $\mathbf{x}$ being the P_f-basis obtained after running the greedy algorithm, it can be shown (see [87]) that $f(\bar{E}_U) = \mathbf{x}(\bar{E}_U)$ (union of tight set is tight). For such $\bar{E}_U$, we have that

$$\mathbf{x}(\mathcal{E}) = \mathbf{x}(E_U) + \mathbf{x}(\bar{E}_U) = \mathbf{y}(E_U) + f(\bar{E}_U) \ . \tag{4.199}$$

This is because $\mathbf{x}(\bar{E}_U) = f(\bar{E}_U)$ and if $e \notin \bar{E}_U$, $\mathbf{x}(e) = \mathbf{y}(e)$.

Cunningham's algorithm

Based on these observations, Cunningham [87] proposed a modified version of the greedy algorithm to compute a P_f-basis, as well as a minimizing subset for the minimization in (4.196). The algorithm is presented in Table 4.3.

It starts with $\mathbf{x} = 0$ and $\bar{E}_U = \emptyset$. For each $e \in \mathcal{E}$, the component $\mathbf{x}(e)$ is increased as much as possible: $\mathbf{x} \leftarrow \mathbf{x} + \epsilon_{max}(e)\mathbf{1}_e$. If the minimum in (4.198) is achieved at $\min_E (f(E) - \mathbf{x}(E),\ e \in E)$,

Table 4.3: Pseudocode of the oracle *CunninghamMin* that solves the minimization (4.200).

Cunningham
Input: Polymatroid function f, $\mathbf{y} \in R_+^{|\mathcal{E}|}$
Output: minimum ϵ, minimizer E_U

```
1 begin
2  x = 0
3  E_U := {}
4  for e ∈ ℰ
5     ε := min(f(E) − x(E) : e ∈ E)
6     E_e := a minimizer
7     if ε <= y(e) − x(e) then E_U := E_U ∪ E_e
8     else ε := y(e) − x(e)
9     end        //if
10    x = x + ε * 1_e
11  end   //for
12 end    //begin
```

then update $\bar{E}_U \leftarrow \bar{E}_U \cup E$ where E is a minimizer. At the end of the algorithm, $\bar{E}_U$ is a tight set and $\mathbf{x}$ is maximal. Also, it satisfies $\mathbf{x} \in P_f$ and $\mathbf{x} \leq \mathbf{y}$, with $\mathbf{x}(\mathcal{E}) = \mathbf{y}(E_U) + f(\bar{E}_U)$.

To find a P_f-basis, Cunningham's algorithm performs $|\mathcal{E}|$ computations of the the minimization below:

$$\min_E \left(f(E) - \mathbf{x}(E),\ e \in E \subseteq \mathcal{E}\right) \ . \tag{4.200}$$

Now, all that remains is to find an algorithm that computes the minimization in polynomial time. This is the subject of the next section.

Network Flow

Let $G = (\mathcal{V}, \mathcal{E})$ be a connected graph and let $f(\cdot)$ be defined as $f(E) = |\mathcal{V}| - Q(G_E)$ ($f(\cdot)$ the rank function of the graphic matroid that is associated to G [90]). Let P_f be the polymatroid associated with $f(\cdot)$. An equivalent description of P_f is given as follows (see [87]):

$$P_f = \left\{\mathbf{x} \in R_+^{|\mathcal{E}|},\ \mathbf{x}(\gamma(B)) \leq |B| - 1 \text{ for all } B,\ \emptyset \neq B \subseteq \mathcal{V}\right\} \ , \tag{4.201}$$

where $\gamma(B)$ denotes the set of edges with both ends in B.

Recall that our goal is, for a given e, to find a subset E, $e \in E \subseteq \mathcal{E}$ that minimizes $f(E) - \mathbf{x}(E)$. This is equivalent to finding B that minimizes $|B| - 1 - \mathbf{x}(\gamma(B))$, with $e \in \gamma(B)$.

To find the minimizing subset of nodes, B, we define the following graph G' for a given polymatroid function $f(\cdot)$, $\mathbf{x} \in P_f$, and edge $e \in \mathcal{E}$. The vertices of G' are $\mathcal{V} \cup (r, s)$ for new vertices r and s. Each $e \in \mathcal{E}$ is an edge of G', having the same ends and having capacity $\frac{1}{2}\mathbf{x}_e$. There

is an edge joining v to s for each $v \in \mathcal{V}$, it has capacity 1. There is an edge joining r to v for each $v \in \mathcal{V}$. It has capacity ∞ if v is an end of j, and otherwise it has capacity $\frac{1}{2}\mathbf{x}(\delta(v))$. (Here $\delta(B) = \{e \in \mathcal{E},\ e \text{ has exactly one end in } B \subseteq \mathcal{V}\}$, $\delta(v)$ is shorthand for $\delta(\{v\})$. This construction is illustrated in Figure 4.10(a). Its motivation is to ensure that $e \in \gamma(B)$ as can be seen next.

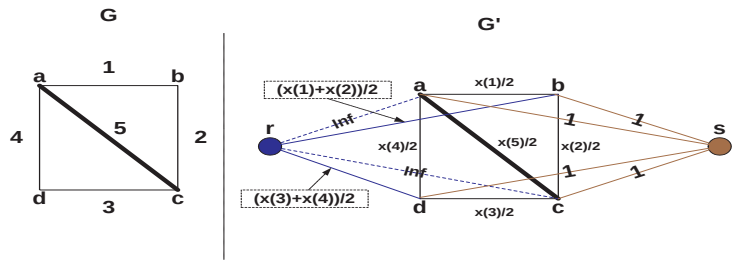
Consider a cut in G' induced by the set $B \cup \{r\}$, where $e \in B \subseteq \mathcal{V}$. It is the set of links that have one end in $B \cup \{r\}$ and the other end in the complement of $B \cup \{r\}$. The capacity of such cut is (see an illustration in Figures 4.10(b))

$$|B| + \frac{1}{2}\mathbf{x}\left(\delta(B)\right) + \mathbf{x}\left(\gamma(\bar{B})\right) + \frac{1}{2}\mathbf{x}\left(\delta(B)\right) = |B| + \mathbf{x}(\mathcal{E}) - \mathbf{x}(\gamma(B)) \tag{4.202}$$

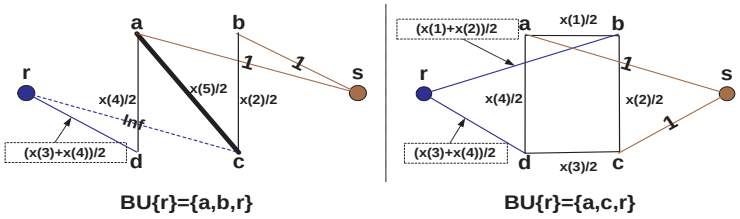
$$= |B| - 1 - \mathbf{x}(\gamma(B)) + (\mathbf{x}(\mathcal{E}) + 1). \tag{4.203}$$

The first term in the LHS of equation (4.202) corresponds to edges going from nodes in B to the sink s. There are $|B|$ of them, each having capacity 1. The next term corresponds to edges going from a node in B to a node in $\bar{B}$. The last two terms correspond to edges going from the root r to nodes in $\bar{B}$. For each such edge (r, u), the capacity is defined as $\frac{1}{2}\delta(\{u\})$. Let $e = (u, v) \in \delta(\{u\})$. Then, if $v \in B$ (i.e. $e \in \delta(B)$), then $\mathbf{x}(e)$ appears only in the capacity of (r, u); implying the term $\frac{1}{2}\mathbf{x}\left(\delta(B)\right)$. If, on the other hand, $v \notin B$ (i.e. $e \in \gamma(\bar{B})$), then $\mathbf{x}(e)$ appears both in the capacity of (r, u), and in that of (r, v), thus the term $\mathbf{x}\left(\gamma(\bar{B})\right)$.

Now, since a cut induced by a subset of edges B will have infinite capacity if $e \notin \gamma(B)$, a minimum cut in G' will indeed have the form $B \cup \{r\}$ with $e \in \gamma(B)$, hence, minimizing $|B| - 1 - \mathbf{x}(\gamma(B))$. As a consequence, any network flow algorithm can serve as an oracle for Cunningham's algorithm. Many polynomial implementations of network flow algorithms ([91], [92]) have been proposed since the proof of the Max-Flow Min-Cut theorem by Ford and Fulkerson [93] in 1962.



(a) Constructing the graph G' from G for the network flow algorithm.



(b) Illustrating the cut induced by $B \cup \{r\}$

Figure 4.10: Constructing the graph G' for the network flow algorithm. Figure 4.10(a) shows the construction of G' from G. The edge under consideration in this example is $e = 5$. Examples in Figures 4.10(b) show the cut induced by $B \cup \{r\}$ for $B \subseteq \mathcal{V}$. In the left figure, $B = \{a, b\}$ does not contain $j = 5$. The capacity of this cut is equal to infinity. In the right figure, $B = \{a, c\}$ which contains edge $e = 5$ (the only edge). As can be seen in the figure, the capacity of the cut induced by this choice of B is $2 + \mathbf{x}(1) + \mathbf{x}(2) + \mathbf{x}(3) + \mathbf{x}(4)$ which is finite.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis we make two main contributions to the literature. First, we illustrate the potential usefulness of Game Theory in security by modeling the interactions between attackers and defenders as games in three types of communication scenarios. By computing and analyzing the Nash equilibria of the games, we predict the adversaries' attacks, determine the set of assets that are most likely to be attacked, and suggest defense strategies for the defenders. Second, we have determined the structure of a particular Nash equilibrium for a class of bimatrix games and we completely characterize the structure of all Nash equilibria for a subset of those games. Also, we have proposed a polynomial-time algorithm to compute Nash equilibria for certain games.

The first game that we have studied models the communication scenario where not all network components taking part in a communication process are trustworthy. We study the strategic interaction of a receiver and an intruder as a game and analyze the set of Nash equilibria. We have found that when the chances that the intruder is present are not small, the receiver should never trust a received message. If those chances are relatively small, the receiver should always trust a received message, despite the fact that the best attacker always attacks. The equilibrium strategies also suggest that, in general, the most aggressive attackers are not always the most dangerous ones. Finally, we have seen that by introducing a *challenge-response* mechanism, the defender can completely deter the attacker from attacking.

The second game scenario is one where an intelligent virus is trying to infect a network protected by an Intrusion Detection System (IDS). We study the game where the virus is choosing its infection rate while the IDS is deciding how to set the best threshold for intrusion decision. By analyzing the Nash equilibrium strategies, we have found that aggressive viruses are not always the most dangerous ones. A virus that can intelligently tune its infection rate could stay longer in the system and cause more damage. We have also seen that an IDS that delays traffic while making security decision performs better than an IDS that does not. Of course, the challenge here is to

determine the optimal amount of delay. We did not address this challenge in this thesis.

The third example considers the scenario where a defender needs a subset of a set of resources to perform some critical task. To disrupt the task, an attacker is at the same time targeting one resource to attack from the set. We introduce the notion of critical subset of resources and use this notion to propose a vulnerability metric for the task. We find that, in Nash equilibrium, the attacker always targets a critical subset of resources and the defender should only use a *minimal* amount of resources in the critical subset.

We have illustrated this model with two examples of communication scenarios: a topology design game where a defender is trying to choose a spanning tree of a graph as communication structure; and a demand-supply game where a network manager is selecting a feasible flow to carry goods from a set of sources to a set of destinations.

In the spanning tree game, we have found that the usual *edge-connectivity* metric is not the right one for graphs when an adversary might be targeting the links. In the demand-supply game, we have found that the attacker always targets subsets of links that maximize the minimum fraction of goods that any flow carries per link. In both games, if the adversary launches an attack, she will do so by uniformly targeting the links in a critical subset of links.

Computing a Nash equilibrium of arbitrary 2-player game is known to be in general complex. In the third scenario considered above, we have determined the structure of a particular Nash equilibrium for all games. For the games with zero (or constant) costs of attack, we have characterized all Nash equilibria. In all equilibria, the attacker targets (a set of) critical subsets of resources. Also, we have proposed polynomial-time algorithms that can be used to compute critical subsets (hence Nash equilibria) for a certain class of 2-player games.

5.1.1 Discussion: Challenges for Game Theoretic Approaches to Security

As has been illustrated in this thesis and in the many cited prior works, the application of Game Theory to security is a young but promising research field. However, there are lots of challenges that need to be addressed in order to make Game Theory a viable approach for reasoning about security problems.

The first challenge is the need to quantify security parameters such as risk, privacy, trust, and reputation. This is in general needed for the purpose of security risk evaluation and management; but it is necessary for the game theoretic approach because of the need to define payoff (cost) functions for both attackers and defenders. One cannot define a cost function for the security players without a good quantification of those parameters. Computing an exact quantification is likely impossible. However, estimated values can be obtained by using data from past attacks. For that to happen, there is a need to develop a trust relationship with corporations and network services and operators, who currently share with researchers only a small fraction of the information about the attacks they are facing. Estimating attackers' payoffs is more challenging. Many efforts is being spent to understanding the *economics* of the cyber security black market. However, much

is unknown.

Another challenge that game theorists for security need to address is the interpretation of (Nash) equilibrium. Equilibrium concepts such as *mixed strategy* Nash equilibrium need to be converted into real-world security strategies. How should an IDS understand a mixed strategy between action A (e.g. intensive monitoring) and action B (e.g. normal or light monitoring)? Even in the Game Theory community, the debate is not yet closed on the interpretation of mixed strategy NE. Osborne and Rubinstein have an elaborate discussion about this in their book [40]. In [34] "frequentist" and a "probabilistic" interpretations are suggested. In general, a meaningful interpretation is needed prior to the practical implementation of any game theoretic solution.

The complexity of computing an equilibrium of a game is yet another challenge to Game Theory for security. Computing an equilibrium of a game is in general a difficult task. For instance, in [47] the authors have shown that the NE computation problem for a general bimatrix game is PPAD-complete (polynomial parity argument for directed graphs). There is some belief that this class of problems are intractable. For the particular case of a zero-sum game, we can, via linear programming, derive polynomial algorithm to compute Nash equilibria. In this thesis, we have shown how for a class of *quasi-zero-sum* games, one can derive a polynomial algorithm by using Polymatroid Theory. Numerical solutions have also been widely used. Although numerical solutions fail to show a global picture of the solutions of a problem, they can help get *some* degree of understanding of the outcome of a game.

As a consequence of this complexity, game theoretic studies often adopt simplified models which leads to (less,un)-realistic scenarios. Like in this thesis, many authors assume *common knowledge* of payoffs and actions, zero-sum or quasi-zero-sum games, one-shot or small horizon dynamic games. The choice of such model is generally guided by intuition and usually lacks solid justifications. These shortcomings certainly need to be fixed. However, one should notice that those simplifications do not lead to trivial games; and in most of the cases, these models help derive conclusions that would not have been possible in a non-game theoretic setting.

One general criticism to Game Theory is its *rationality* assumption. Players in game theoretic models are often considered to be *unboundedly* rational. This contrasts with general observation. In fact, many experimental results of *behavioral economics* have shown that people do not always play the Nash equilibrium strategies predicted by the theory [94], [95]. In the context of communication security, even though one can assume some degree of rationality for sophisticated players such as cybercriminal, inside attackers, and nation-state intelligence agents, infinite rationality does not hold. Models using bounded rationality can be used to study such situations; however of research efforts are still needed to understand these models.

There is a diversity of games models (non-cooperative/cooperative, static/dynamic, Bayesian, Stochastic, complete/incomplete information) to study different communication security scenarios. However, for a given problem, the choice of the appropriate game model is still left at the intuition of the modeler. So far, there is no systematic way to choose (and justify) a model. Although a clear dichotomy and one-to-one correspondence between security problems and game models are not possible, a good selection methodology is needed. This will require a widespread understanding

of Game Theory by security specialists.

The game theoretic approach to security is still in its infancy. The potentials it presents for a better understanding of the security problem are enormous, as we have shown in the models considered in this thesis and in the many cited related works. However, there is still a lot to do in order to convert game theoretic results into practical security solutions. The bridge between the theory and the practice is still *under construction*.

5.1.2 Experimental Design

As we have mentioned in the previous section, one general criticism to Game Theory is the *rationality* assumption of the players. In fact, experimental studies have shown that agents are rarely fully rational. In chapters 2-4 of this thesis we have predicted the actions of both rational attackers and a rational defenders by analyzing the Nash equilibria for three communication scenarios. One interesting follow up of this study is to test the theoretical conclusions against real-life (laboratory) data. A design of such experimental study is discussed in the appendix section B.

Bibliography

[1] S. TZU. (-6 BC) The Art of War. Translated from the Chinese By Lionel Giles, M.A. (1910). [Online]. Available: http://www.chinapage.com/sunzi-e.html

[2] SocyBerty. (2008, April) How the Internet Has Changed the World. [Online]. Available: http://socyberty.com/society/how-the-internet-has-changed-the-world/

[3] B. Meadowcroft. (2001, june) The Impact of Information Technology on Work and Society. [Online]. Available: http://www.benmeadowcroft.com/reports/impact/

[4] M. Jones and I. Alony, "The Cultural Impact of Information Systems - Through the Eyes of Hofstede A Critical Journey," *Issues in Informing Science and Information Technology*, vol. 4, pp. 265 – 283, 2007.

[5] CERT/CC. (2001, Sept) CERT Advisory CA-2001-26 Nimda Worm. [Online]. Available: http://www.cert.org/advisories/CA-2001-26.html

[6] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "Inside the Slammer Worm," *Security Privacy, IEEE*, vol. 1, no. 4, pp. 33 – 39, july 2003.

[7] Conficker-Working-Group. (2009, March) Press Releases. [Online]. Available: http://www.confickerworkinggroup.org/wiki/pmwiki.php/PR/PressReleases

[8] ——. (2009, Oct) Infection Tracking. [Online]. Available: http://www.confickerworkinggroup.org/wiki/pmwiki.php/ANY/InfectionTracking

[9] ——, "Conficker Working Group: Lessons Learned," White paper, Conficker Working Group, Jan 2011, available online (59 pages). [Online]. Available: http://www.confickerworkinggroup.org/wiki/#toc3

[10] Cyber-Secure-Institute. (2009, Apr) Cyber Secure Institute on the Conficker Controversy. [Online]. Available: http://cybersecureinstitute.org/blog/?p=12

[11] N. Falliere, L. O. Murchu, and E. Chien, "W32.Stuxnet Dossier," White paper, Symantec Security Response, Feb 2011, available online (69 pages). [Online]. Available: http://www.symantec.com/.../w32_stuxnet_dossier.pdf

[12] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, "Your Botnet is my Botnet: Analysis of a Botnet Takeover," in *Proceedings of the 16th ACM conference on Computer and communications security*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 635–647. [Online]. Available: http://doi.acm.org/10.1145/1653662.1653738

[13] CNET-News. (2007, May) Cyberattack in Estonia–What it Really Means. [Online]. Available: http://news.cnet.com/Cyberattack-in-Estonia-what-it-really-means/2008-7349_3-6186751.html

[14] BBC-News. Security Experts Say GGoogle Cyber-Attack Was Routine, month = Jan, year = 2010, url = http://news.bbc.co.uk/2/hi/8458150.stm.

[15] TechWorld. (2010, Dec) Anonymous Knocks Swiss Bank Offline Over Wikileaks Attack. [Online]. Available: http://news.techworld.com/security/3252222/anonymous-knocks-swiss-bank-offline-over-wikileaks-attack/

[16] The-Washington-Post. (2001, May) Washington Post, White House, FAA, DoD, Others, Targeted in Online Attack. [Online]. Available: http://voices.washingtonpost.com/securityfix/2009/07/washington_post_white_house_fa.html

[17] US-CERT. (2010, Dec) Technical Cyber Security Alerts. [Online]. Available: http://www.us-cert.gov/cas/techalerts/index.html

[18] P. T. Leeson and C. J. Coyne, "The Economics of Computer Hacking," *Journal of Law, Economics and Policy*, 2006.

[19] BBC-News. (2010, Dec) Pro-Wikileaks Activists Abandon Amazon Cyber Attack. [Online]. Available: http://www.bbc.co.uk/news/technology-11957367

[20] M. Cremonini and D. Nizovtsev, "Understanding and Influencing Attackers Decisions: Implications for Security Investment Strategies," June 2006.

[21] T. Alpcan and T. Baser, *Network Security: A Decision and Game-Theoretic Approach*, 1st ed. Cambridge University Pres, November 2010.

[22] G. Gilder. (1995, Nov) Metcale's Law And Legacy. [Online]. Available: http://www.seas.upenn.edu/ gaj1/metgg.html

[23] V. Sekar, Y. Xie, D. A. Maltz, M. K. Reiter, and H. Zhang, "Toward a Framework for Internet Forensic Analysis," in *In Third Workshop on Hot Topics in Networking (HotNets-III)*, 2004.

[24] H. Burch and B. Cheswick, "Tracing Anonymous Packets to Their Approximate Source," in *in Usenix LISA*, 2000.

[25] L. Kleinrock. (1996, Aug) The Birth of the Internet. [Online]. Available: http://www.cs.ucla.edu/ lk/LK/Inet/birth.html

[26] X. Ao, "Report on DIMACS Workshop on Large-scale Internet Attacks," Sept 2003.

[27] WASHINGTON-(Reuters). (2010, December) U.S. Code-Cracking Agency Works as if Compromised. [Online]. Available: http://ca.reuters.com/article/technologyNews/idCATRE6BF6BZ20101216

[28] R. Anderson, "Why Information Security is Hard-An Economic Perspective," in *Proceedings of the 17th Annual Computer Security Applications Conference*. Washington, DC, USA: IEEE Computer Society, 2001, pp. 358–.

[29] J. Václav Matyás and Z. Ríha, "Biometric Authentication - Security and Usability," in *Proceedings of the IFIP TC6/TC11 Sixth Joint Working Conference on Communications and Multimedia Security: Advanced Communications and Multimedia Security*. Deventer, The Netherlands, The Netherlands: Kluwer, B.V., 2002, pp. 227–239.

[30] N. F. Johnson, Z. Duric, and S. Jajodia, "Information Hiding: Steganography and Watermarking - Attacks and Countermeasures," in *Advances in Information Security*, vol. 1. Springer, 2000.

[31] M. Sabhnani, "Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset Within Misuse Detection Context," in *International Conference on Machine Learning: Models, Technologies, and Applications*, 2003, pp. 209–215.

[32] R. Khanna and H. Liu, "Distributed and Control Theoretic Approach to Intrusion Detection," in *Proceedings of the 2007 international conference on Wireless communications and mobile computing*, ser. IWCMC '07. New York, NY, USA: ACM, 2007, pp. 115–120.

[33] G. Giacinto, R. Perdisci, M. D. Rio, and F. Roli, "Intrusion Detection in Computer Networks by a Modular Ensemble of One-Class Classifiers," *Information Fusion*, vol. 9, 2008.

[34] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A Survey of Game Theory as Applied to Network Security," *Hawaii International Conference on System Sciences*, vol. 0, pp. 1–10, 2010.

[35] M. Manshaei, Q. Zhu, T. Alpcan, T. Basar, and J.-P. Hubaux, "Game Theory Meets Network Security and Privacy," EPFL, Lausanne, Tech. Rep., 2010. [Online]. Available: http://infoscience.epfl.ch/record/151965/files/GamesecSurvey-SubmittedVersion.pdf

[36] D. A. Burke, "Towards a Game Theoretic Model of Information Warfare," Air force Institute of Technology, Tech. Rep., 1999.

[37] H. Cavusoglu, S. Raghunathan, and W. Yue, "Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment," *J. Manage. Inf. Syst.*, vol. 25, pp. 281–304, September 2008.

[38] S. Flowerday and R. von Solms, "Trust: An Element of Information Security," in *Security and Privacy in Dynamic Environments*, ser. IFIP International Federation for Information Processing. Springer Boston, 2006, vol. 201, pp. 87–98.

[39] L. Buttyan and J.-P. Hubaux, *Security and Cooperation in Wireless Networks: Thwarting Malicious and Selfish Behavior in the Age of Ubiquitous Computing*. New York, NY, USA: Cambridge University Press, 2007.

[40] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, ser. MIT Press Books. The MIT Press, June 1994, vol. 1.

[41] D. Fudenberg and J. Tirole, *Game Theory*, ser. MIT Press Books. The MIT Press, 1991.

[42] J. V. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[43] M. Shubik, "Game Theory: Models of Strategic Behavior and Nuclear Deterrence," Cowles Foundation for Research in Economics, Yale University, Cowles Foundation Discussion Papers 829, Mar 1987. [Online]. Available: http://ideas.repec.org/p/cwl/cwldpp/829.html

[44] J. C. Harsanyi, "Games with Incomplete Information Played by "Bayesian" Players, I-III. Part III. The Basic Probability Distribution of the Game," *Management Science*, vol. 14, no. 7, pp. 486–502, Mar 1968.

[45] J. Nash, "Non-Cooperative Games," *The Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, Sept 1951.

[46] V. Robert, *Linear Programming: Foundations and Extensions*. Springer, May 2001.

[47] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The Complexity of Computing a Nash Equilibrium." ACM Press, 2006, pp. 71–78.

[48] P. Liu and W. Zang, "Incentive-Based Modeling and Inference of Attacker Intent, Objectives, and Strategies," in *Proceedings of the 10th ACM conference on Computer and communications security*, ser. CCS '03. New York, NY, USA: ACM, 2003, pp. 179–189.

[49] J. Grossklags, N. Christin, and J. Chuang, "Predicted and Observed User Behavior in the Weakest-Link Security Game," in *Proceedings of the 1st Conference on Usability, Psychology, and Security*. Berkeley, CA, USA: USENIX Association, 2008, pp. 8:1–8:6.

[50] R. Wash and J. K. MacKie-Mason, "Security When People Matter: Structuring Incentives for User Behavior," in *Proceedings of the ninth international conference on Electronic commerce*, ser. ICEC '07. New York, NY, USA: ACM, 2007, pp. 7–14.

[51] L. Jiang, V. Anantharam, and J. Walrand, "Efficiency of Selfish Investments in Network Security," in *Proceedings of the 3rd international workshop on Economics of networked systems*, ser. NetEcon '08. New York, NY, USA: ACM, 2008, pp. 31–36.

[52] N. Shetty, G. Schwartz, and J. Walrand, "Can Competitive Insurers Improve Network Security?" in *Proceedings of the 3rd international conference on Trust and trustworthy computing*, ser. TRUST'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 308–322.

[53] H. Otrok, N. Mohammed, L. Wang, M. Debbabi, and P. Bhattacharya, "A Game-Theoretic Intrusion Detection Model for Mobile Ad Hoc Networks," *Comput. Commun.*, vol. 31, pp. 708–721, March 2008.

[54] T. Alpcan and T. Baser, "A Game Theoretic Analysis of Intrusion Detection in Access Control Systems," in *in Proc. of the 43rd IEEE Conference on Decision and Control*, 2004, pp. 1568–1573.

[55] K.-w. Lye and J. M. Wing, "Game Strategies in Network Security," *International Journal of Information Security*, vol. 4, pp. 71–86, 2005.

[56] D. Liu, X. Wang, and L. J. Camp, "Game theoretic modeling and analysis of insider threats," *International Journal of Critical Infrastructure Protection*, vol. 1, pp. 75–80, 12 2008.

[57] H. Cavusoglu and S. Raghunathan, "Configuration of Detection Software: A Comparison of Decision and Game Theory Approaches," in *Informs*, vol. 1. Addison Wesley, Sept 2004, pp. 131–148.

[58] T. Alpcan, L. Buttyán, and J. S. Baras, Eds., *Decision and Game Theory for Security - First International Conference, GameSec 2010, Berlin, Germany, November 22-23,2010. Proceedings*, ser. Lecture Notes in Computer Science, vol. 6442. Springer, 2010.

[59] T. Alpcan and T. Basar, "A Game Theoretic Approach to Decision and Analysis in Network Intrusion Detection," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 3, dec 2003, pp. 2595 – 2600.

[60] ——, "A game Theoretic Analysis of Intrusion Detection in Access Control Systems," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 2, dec 2004, pp. 1568 – 1573.

[61] A. Patcha and J.-M. Park, "A Game Theoretic Approach to Modeling Intrusion Detection in Mobile Ad Hoc Networks," in *Information Assurance Workshop, 2004. Proceedings from the Fifth Annual IEEE SMC*, 2004, pp. 280 – 284.

[62] Q. Zhu and T. Basar, "Dynamic policy-based ids configuration," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, 2009, pp. 8600 –8605.

[63] S. Zonouz, H. Khurana, W. Sanders, and T. Yardley, "RRE: A Game-Theoretic Intrusion Response and Recovery Engine," in *Dependable Systems Networks, 2009. DSN '09. IEEE/IFIP International Conference on*, july 2009, pp. 439 –448.

[64] Y. Liu, C. Comaniciu, and H. Man, "A Bayesian Game Approach for Intrusion Detection in Wireless Ad Hoc Networks," in *Proceeding from the 2006 workshop on Game theory for communications and networks*, ser. GameNets '06, vol. 4, New York, NY, USA, 2006.

[65] T. Alpcan and S. Buchegger, "Security Games for Vehicular Networks," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 280–290, 2011.

[66] C. Castelfranchi and Y.-H. Tan, Eds., *Trust and Deception in Virtual Societies*. Norwell, MA, USA: Kluwer Academic Publishers, 2001.

[67] J. Zander, "Jamming Games in Slotted Aloha Packet Radio Networks," in *Military Communications Conference, 1990. MILCOM '90, Conference Record, A New Era. 1990 IEEE*, vol. 2, sep-oct 1990, pp. 830 –834.

[68] E. Altman, K. Avrachenkov, and A. Garnaev, "A Jamming Game in Wireless Networks with Transmission Cost," in *Network Control and Optimization*, ser. Lecture Notes in Computer Science, T. Chahed and B. Tuffin, Eds. Springer Berlin / Heidelberg, 2007, vol. 4465, pp. 1–12.

[69] A. Kashyap, T. Basar, and R. Srikant, "Correlated Jamming on MIMO Gaussian Fading Channels," in *Communications, 2004 IEEE International Conference on*, vol. 1, june 2004, pp. 458 – 462.

[70] M. Mavronicolas, V. Papadopoulou, A. Philippou, and P. Spirakis, "A Graph-Theoretic Network Security Game," in *Internet and Network Economics*, ser. Lecture Notes in Computer Science, X. Deng and Y. Ye, Eds. Springer Berlin / Heidelberg, 2005, vol. 3828, pp. 969–978.

[71] D. Avis, G. Rosenberg, R. Savani, and B. von Stengel, "Enumeration of Nash Equilibria for Two-Player Games," *Economic Theory*, vol. 42, pp. 9–37, 2010.

[72] L. A. Wolsey and G. L. Nemhauser, *Integer and Combinatorial Optimization*. Wiley-Interscience, November 1999.

[73] A. Toshev, "Submodular Function Minimization," University of Pennsylvania, Philadelphia, Tech. Rep., 2010. [Online]. Available: www.seas.upenn.edu/ toshev/Site/About_Me_files/wpii-2.pdf

[74] R. J. Vanderbei, *Linear Programming: Foundations and Extensions*, 2nd ed. Springer, 2001. [Online]. Available: http://www.princeton.edu/~rvdb/LPbook/

[75] P. B. Miltersen and T. B. Sørensen, "Computing Sequential Equilibria for Two-Player Games," in *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*. New York, NY, USA: ACM, 2006, pp. 107–116.

[76] R. A. M. Jaeyeon Jung and V. Paxson, "On the Adaptive Real-Time Detection of Fast-Propagating Network Worms," in *Fourth GI International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, July 2007.

[77] L. Billings, W. M. Spears, and I. B. Schwartz, "A unified prediction of computer virus spread in connected networks," *Physics Letters A*, vol. 297, no. 3-4, pp. 261–266, 2002.

[78] G. Serazzi and S. Zanero, "Computer virus propagation models," 2003. [Online]. Available: citeseer.ist.psu.edu/serazzi03computer.html

[79] S. Goldman and J. Lightwood, "Cost optimization in the sis model of infectious disease with treatment," *Topics in Economic Analysis & Policy*, vol. 2, no. 1, pp. 1007–1007, 2002.

[80] L. A. Wolsey and G. L. Nemhauser, *Integer and Combinatorial Optimization*, 1st ed. Wiley-Interscience, November 1999.

[81] D. R. Fulkerson, "Blocking and Anti-Blocking Pairs of Polyhedra," *Math. Programming*, no. 1, pp. 168–194, 1971.

[82] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004.

[83] S. Chopra, "On the Spanning Tree Polyhedron," *Operations Research Letters*, vol. 8, no. 1, pp. 25 – 29, 1989.

[84] D. Gusfield, "Connectivity and Edge-Disjoint Spanning Trees," *Information Processing Letters*, no. 16, pp. 87–89, 1983.

[85] W. T. Tutte, "On the Problem of Decomposing a Graph into N Connected Factors," *Journal of the London Mathematical Society*, no. 36, pp. 221–230, 1961.

[86] J. A. Nash-Williams, "Edge-Disjoint Spanning Trees of Finite Graphs," *Journal London Math. Soc*, no. 36, pp. 445–450, 1961.

[87] W. H. Cunningham, "Optimal Attack and Reinforcement of a Network," *J. ACM*, vol. 32, no. 3, pp. 549–561, 1985.

[88] P. A. Catlin, H.-J. Lai, and Y. Shao, "Edge-Connectivity and Edge-Disjoint Spanning Trees," *Discrete Mathematics*, vol. 309, no. 5, pp. 1033 –1040, 2009.

[89] D. R. Fulkerson and D. B. Weinberger, "Blocking Pairs of Polyhedra Arising from Network Flows," *Journal of Combinatorial Theory, Series B*, vol. 18, no. 3, pp. 265 – 283, 1975.

[90] D. Welsh, *Matroid Theory*. London, New York, San Francisco: Academic Press, 1976.

[91] M. Stoer and F. Wagner, "A Simple Min-Cut Algorithm," *J. ACM*, vol. 44, no. 4, pp. 585–591, 1997.

[92] D. R. Karger and C. Stein, "An o(n2) Algorithm for Minimum Cuts," New York, NY, USA, 1993, pp. 757–765.

[93] D. F. L Ford, "Flows in Networks." Princeton Univ. Press, 1962, pp. 453–460.

[94] J. K. Goeree and C. A. Holt, "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions," *The American Economic Review*, vol. 91, no. 5, pp. pp. 1402–1422, 2001. [Online]. Available: http://www.jstor.org/stable/2677931

[95] U. Gneezy, "Deception: The Role of Consequences," *American Economic Review*, vol. 95, no. 1, pp. 384–394, March 2005. [Online]. Available: http://ideas.repec.org/a/aea/aecrev/v95y2005i1p384-394.html

Appendix A

Computing Critical Subsets

A.1 A binary search algorithm for computing a critical subset and θ

In this appendix section, we present an algorithm to compute a critical subset of a graph. A first algorithm was discussed in section 4.6.2.
Recall that for the spanning tree game, the vulnerability $\theta = \max\left(\frac{\mathcal{M}(E)}{|E|}\right)$ of the graph takes the form of a ratio $\frac{p}{q}$, where $0 \leq p \leq |\mathcal{V} - 1|$ and $1 \leq q \leq |\mathcal{E}|$. Letting $\sigma := \frac{p}{q}$, we have also shown in section 4.6.2 that,

$$\theta \leq \sigma \quad \Leftrightarrow \quad 0 \leq \min_{E \subseteq \mathcal{E}} (\sigma|E| - \mathcal{M}(E)) \ . \tag{A.1}$$

Now, assuming that the minimization above can be efficiently computed, we propose 2-dimensional *binary search* algorithm. Figure A.1 illustrate the algorithm that we present in Table A.1.

The algorithm works as follow.
We keep a set of candidate values Pr for p, and for each $p \in Pr$, a range $\{q_{min}(p), \ldots, |\mathcal{E}|\}$ of values of q for which the test in (4.193) will be carried out.

At each iteration, for some $p \in Pr$ and $q \in \{q_{min}(p), \ldots, |\mathcal{E}|\}$, a call is made to the minimization oracle; then Pr and q_{min} are updated. Pr is defined as $P_r = \{1, \ldots, |\mathcal{V}| - 1\}$ at initial time, and maintained as follows.
Since θ is always less than or equal to 1, the values of p and q for which $p/q > 1$ can be ignored from the test. These values correspond to the "dark" (blue) region above the first diagonal of Figure A.1 (if the graph does not contain a bridge, one can eliminate the values in the first diagonal as well). This implies that for each p, there is a minimum value for q, call it $q_{min}(p)$; i.e. when p is considered in a given iteration, only values of q in the range $\{q_{min}(p), \ldots, |\mathcal{E}|\}$ need to be used for testing.

Also, if $\theta \leq \frac{p_0}{q_0}$ for some fixed (p_0, q_0), then $\theta \leq \frac{p}{q}$ for all $\frac{p}{q} > \frac{p_0}{q_0}$. As of such, those values can

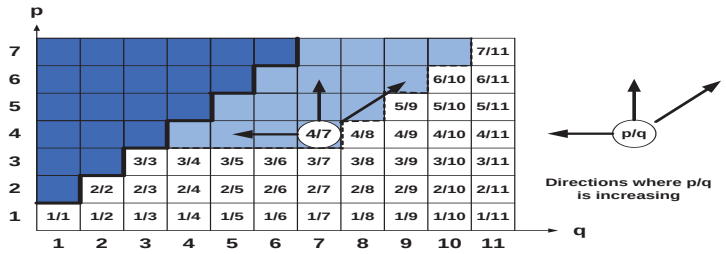


Figure A.1: An illustration of the 2-dimensional search algorithm to find the vulnerability of a graph. The dark (blue) region consists of p and q verifying $p/q > 1$. Since $\theta \leq 1$, those values do not need to be tested. The light (blue) consist of values of p and q such that $\frac{p}{q} > \frac{p_0}{q_0}$ (here $\frac{p_0}{q_0} = \frac{4}{7}$). If $\theta < \frac{p_0}{q_0}$, then, those values can be discarded from the test. The remaining (uncolored) values are the only ones that need to be tested.

be safely discarded from the set of values to be tested. In Figure A.1, that set is represented by the "light" (blue) region for $p_0 = 4$ and $q_0 = 7$. It is the set of numbers that are located in the 135 degrees range, from the first diagonal to the horizontal axis (traveling counterclockwise).
After removing this set, the values of $q_{min}(p)$ need to be updated for all $p \geq p_0$. If q_0 is the first value of q (starting from $|\mathcal{E}|$ going down) for which the test succeeds (i.e. $\theta \leq \frac{p_0}{q_0}$), then $q_{min}(p_0) = q_0 + 1$, and for $p \in \{p_0 + 1, \ldots, \mathcal{V} - 1\}$, $q_{min}(p)$ is obtained by adding 1 to $q_{min}(p-1)$. If $q_{min}(p) > |\mathcal{E}|$, then p can be removed from the set Pr of candidate values for p. If for some p, the test fails for all $q \in \{q_{min}, \ldots, |\mathcal{E}|\}$, then p can also be discarded from Pr.
The algorithm stops when the test succeeds and $|Pr| = 1$.

For each value of p, the algorithm makes less than $|\mathcal{E}|$ calls to the oracle, and there are at most $|\mathcal{V}|$ possible values for p (this is the worst case). Thus, computing a critical subset will take a polynomial time provided that the minimization in (4.191) can be done in polynomial time.

A.2 Argument for Remark in Section 4.4.1

In section 4.4.1, we have claimed that in the definition of critical subset we do not need the *feasibility* requirement of E. The following lemma justifies our claim. Recall that a subset $E \subseteq \mathcal{E}$ of edges is called *feasible* if *E is such that, for every edge $e \in E$, adding e to $G_{\bar{E}}$ (the graph obtained by removing from G, the edges in E) decreases its number of connected components by 1.*. A critical subset is one that achieves the maximum below

$$\theta(E) = \max_{\tilde{E}} \left(\frac{\mathcal{M}(\tilde{E}) - \boldsymbol{\mu}(\tilde{E})}{|\tilde{E}|} \right). \tag{A.2}$$

Lemma 10 *Let $E \subseteq \mathcal{E}$. If E is critical, then it is feasible.*

Table A.1: *BinarySearch2D* algorithm to compute θ and a critical subset. Algorithm *Cunningham-Min* is discussed in section 4.6.2. Method *update* method is presented in Table A.2.

BinarySearch2D
Input: connected graph $G = (\mathcal{V}, \mathcal{E})$, $\mathcal{V} = n$, $\mathcal{E} = m$
Output:θ, $E \subseteq \mathcal{E}$ critical

```
1 begin
2   Pr = {1,2,...,n-1}
3   qmin = {1,2,...,n-1}
4   while |Pr|>0
5      p <-- random(Pr)
6      for q=m downto qmin(p)
7         (E,minpq) = CunninghamMin((p/q)*1,G)
8          if n-1 <= minpq then
9             (Pr,qmin) = update(Pr,p,q)
10            goto 4
11         end  //if
12     end       //for
13     Pr = Pr-p
14  end           //while
15  return E, minpq
16 end             // begin
```

Table A.2: Pseudocode of the *Update* method used in the *BinarySearch2D* algorithm.

Update
Input: Pr, $p \in Pr$, $q \in \{q_{min}, |\mathcal{E}|\}$
Output: new Pr, q_{min}

```
1   begin
2     qmin(p) = q+1
3     for j=p+1 to |n|-1
4         qmin(j) = qmin(j-1)+1
5         if qmin(j)>m
6              Pr = Pr - j
7         end    //if
8     end        //for
9     return Pr, qmin
10   end           //begin
```

Proof: Let E be critical and not feasible. From Lemma 3, we have that $\mathcal{M}(E) = \min_T (|E \cap T|) = Q(G_{\bar{E}}) - 1$, where $Q(G_{\bar{E}})$ is the number of connected components of the graph $G_{\bar{E}}$ obtained by removing from G the edges in E. Now since E is not feasible, there exists an edge $e \in E$ such that adding e to $G_{\bar{E}}$ does not decrease its number of connected components. In other words, $Q(G_{\bar{E}}) = Q(G_{\overline{E\backslash\{e\}}})$, and as a consequence $\mathcal{M}(E) = \mathcal{M}(E\backslash\{e\})$. This implies that

$$\frac{\mathcal{M}(E\backslash\{e\}) - \boldsymbol{\mu}(E\backslash\{e\})}{|E\backslash\{e\}| - 1} = \frac{\mathcal{M}(E) - (\boldsymbol{\mu}(E) - \boldsymbol{\mu}(e))}{|E| - 1} = \frac{\mathcal{M}(E) - \boldsymbol{\mu}(E) + \boldsymbol{\mu}(e)}{|E| - 1}. \tag{A.3}$$

This is strictly larger than $\frac{\mathcal{M}(E)}{|E|}$ (the denominator has strictly decreased and the numerator has increased), which means that $E\backslash\{e\}$ is strictly more critical than E. This contradicts the hypothesis that E is critical. ■

Appendix B

Experimental Study

In chapters 2-4 of this thesis we have predicted the actions of both a rational attacker and a rational defender by analyzing the Nash equilibria for three communication scenarios. In this section, our goal is to design an experimental study in order to generate real-life (laboratory) data against which our theoretical predictions will be tested. In our experiments, we will only consider the spanning tree game described in section 4.4.1. In that game, a defender is choosing a spanning tree of a graph to connect all nodes, while an attacker is targeting one link of the graph in order to break the tree.

B.1 Motivations

Edges-connectivity has been widely considered as the vulnerability metric for graphs. The edge-connectivity of a connected graph is the minimum number of edges that need to be removed in order to disconnect the graph; the corresponding set of edges is called a minimum cutset. In figure B.1, the minimum cutset of the graph corresponds to edges 1 and 4, giving a vulnerability (or edge-connectivity) of 2. The *bold* lines show an example of spanning tree of the network. Given this vulnerability metric, common sense might suggest that in the spanning tree game the attacker

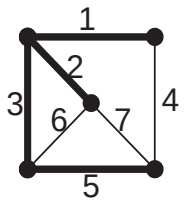


Figure B.1: Example of network graph. The bold lines show one way to connect all nodes in the graph without loop.

will target edges of the minimum cutset to increase her chances of cutting the chosen spanning tree. This is so because the minimum cutset contains at least one link of each tree and it is the minimum set of links with this property.

In section 4.4.1, we have defined a critical subset of edges of a graph to be a subset that maximizes the minimum fraction of edges it has in common with any tree ($\frac{\min_T(|T\cap E|)}{|E|}$). We have proposed this maximum value as the vulnerability metric of the graph. We have also shown that a minimum cutset of a graph is not necessarily critical and that a rational attacker will target edges on a critical subset (not a minimal cutset) and attack them with the same probability. Furthermore, we have shown that if the attacker incurs a cost of attacking the links, she will target a subset that maximizes ($\frac{\min_T(|T\cap E|)}{|E|} - \frac{\boldsymbol{\mu}(E)}{|E|}$) which is the minimum fraction of edges it has n common with any tree minus the average cost of attack of the subset. The attacker does not launch an attack if this difference is negative. Otherwise, she attack links in a maximizing subset with the same probability.

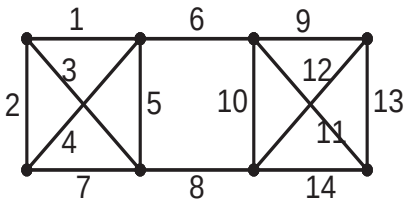
The goal of this experimental study is to test those predictions with subjects in a laboratory setting. More precisely, we would like to assess the minimum cutset assumption (game 1), the critical subset attack (game 2), the uniformity of attack (all games), and the dependency on the attack cost (games 3,4,5).

The games considered in our experiments, are played by a user (the subject) and a machine (the defender). The role of the user is to attack one link of the network in order to break the spanning tree selected by the machine. The machine chooses a spanning according to the distribution predicted by the Nash equilibria analysis.

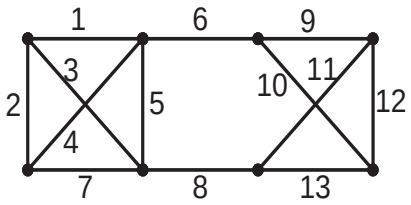
In each game, the subject will be asked to choose one link to attack and win up to 70 tokens. The number 70 is just chosen for convenience reasons. To simulate mixed strategies, the subject is given the choice to bet all tokens on one link, or to distribute the tokens on a selected subset of links and indicate how many tokens to bet on each link. When there is no cost of attack, the total number of tokens bet on links belonging to the selected spanning tree is won by the attacker. If the cost of attacking any link is positive, the reward of the users is given by the total number of tokens bet on links belonging to the selected spanning tree minus the cost of attacking such links.

The questions that we would like to answer with these experiments are the following:

- Which set of links will be attacked and how many tokens are bet by the subjects? By dividing the number of tokens bet on each link by the total number of tokens, we get a probability distribution that can be interpreted as the mixed strategy of the player.
- Do subject follow the “common sense” attack that targets a minimum cutset of the graph?
- Do subjects target links uniformly?
- How do subjects behave when the cost of attack is larger/equal/smaller than the maximum expected gain?



(a) Topology of the network used in experiment 1. The network contains one minimum cutset ($\{6, 8\}$) which also is a critical subset.



(b) Topology of the network used in experiment 1. The network contains one minimum cutset ($\{6, 8\}$). This minimum cutset is however not a critical subset.

Figure B.2: Networks considered in experiments 1 (B.2(a)) and 2 (B.2(b)).

To answer these questions, we collect data by letting the subjects play the following 5 games. In the first two game, there is no cost of attack. In the last three, the attacker (subject) incurs a positive cost when attacking a link.

B.2 Game 1

The graph considered in this game is depicted in figure B.2(a). It contains one minimum cutset ($\{6, 8\}$, the only one) which also is a critical subset. There are three other critical subsets: the set of all links, the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$, and the set $\{6, 8, 9, 10, 11, 12, 13, 14\}$. In this experiment, we would like to test the *minimum cut attack prediction*, which is the optimal strategy in this case because the minimum cutset is critical. Also, because of the symmetry of the graph, we expect subjects to uniformly target edges in the minimum cutset. The predicted (average) attack reward is 35 tokens which can be achieved by betting 35 tokens for each links in the minimum cutset.

B.3 Game 2

Game 2 is played on the graph shown in figure B.2(b). As before, it contains one minimum cutset ($\{6, 8\}$). However, unlike this previous case, the minimum cutset in this figure is not critical. The graph contains one critical subset that corresponds to edges $\{6, 8, 9, 10, 11, 12, 13\}$. The NE analysis predicts that a rational attacker will only target edges in the critical subset and will attack them with the same probability. The predicted (average) attack reward is 40 tokens which can be achieved by betting 10 tokens on each link of the critical subset. In this experiment, we test whether users plays according to the NE prediction.

B.4 Games 3,4,5

The next games are played on the graph in B.2(b). In game 3, we assume that the attacker incurs a cost of $\boldsymbol{\mu}$ = [22 39 27 5 17 3 22 19 42 35 21 38 6 17] tokens on the links. With this cost, the predicted average attack reward is equal to 24 token and the maximizing subset corresponds to the minimum cutset $\{6, 8\}$. With this reward, a rational attacker should always attack uniformly the edges 6 and 8. Game 4 is played with an attack cost equal to $\boldsymbol{\mu}$ = [5 63 5 31 58 28 43 58 63 66 14 19 63 42]. The corresponding maximum average attack reward is equal to -1 and is achieved at subset $\{1, 2, 3\}$. In this case, the NE analysis suggests that a rational player will opt to not attack the network. In the last game (game 5), the cost incurred by the attacker is equal to $\boldsymbol{\mu}$ = [10 61 30 29 68 53 69 17 7 27 36 40 69 35]. It gives a maximum average reward of zero. Two subsets achieve this reward: the minimum cutset $\{6, 8\}$ and the subset $\{6, 8, 9, 10, 11\}$. This is a case where attacking and not attacking give the same reward to the attacker.

In these three experiments, we are interesting in determining how the introduction of the cost of attack will influence the subjects' attack strategies. In the first experiment, the attack costs are not too high and there is an strategy that gives positive reward. In the second one, no strategy can give positive reward and subjects are expected to not play. In the last game, playing and not playing give the same payoff to the subjects.