Game Theory for Cyber Deception against Network SIR Epidemics

Serge Kamguia

Laboratoire d'Informatique d'Avignon
CERI
Avignon Université
Académie d'Aix-Marseille
République Française

Department of Mathematics and Computer Science
Faculty of Science
University of Dschang

Republic of Cameroon

DOCTORAL THESIS

Presented at Avignon University and the University of Dschang in fulfillment of the requirements for the DOCTORATE DEGREE

SPECIALTY: COMPUTER SCIENCE

École Doctorale 536 « Agrosciences & Sciences »
Laboratoire Informatique d'Avignon (EA 4128)

Dschang School of Science and Technology
Unité de Recherches en Informatique Fondamentale, Ingénieurie et Applications (URIFIA)

Game Theory for Cyber Deception against Network *SIR* Epidemics

By
Tsemogne Kamguia Serge Olivier,
INE : 0i0odu08288 AU
Reg. Num.: CM-UDS-19SCI3175 UDS

Publicly defended on December 16, 2022 before a jury composed of:

Abderrahim BENSLIMANE, Professor, LIA	President
Eitan ALTMAN, Director of Research, INRIA,	Reporter
Fei FANG, Professor,Carnegie Mellon University, (USA)	Reporter
Louis-aimé FONO, Professor, Lab. Maths, University of Douala,	Examiner
Alexandre REIFFERSMASSON, Ass. Professor, Université Bretagne Occ.,	Examiner
Yezekael HAYEL, Ass. Professor, LIA	Supervisor
Gabriel DEUGOUÉ, Ass.Professor, URMA, University of Dschang,	Supervisor
Charles KAMHOUA, Senior Electronics Engin., Army Research Lab., (USA)	Supervisor

Dedication

Te ***Tsemogne*** *(the first one), Meu* ***Kenmogne****,*

Peace *be with you.*

The bond that unites us is inalienable and sacred. May the eyes and thoughts of those who read this page reach you so that you may see more of who you are.

Peace *be with you.*

Acknowledgments

I would like to thank all those who made this thesis possible.

Special thanks go to my supervisors, Professor Hayel of Avignon University, Professor Deugoué of the University of Dschang and Dr. Kamhoua of US Army Research Laboratory.

This work was born by moral armament provided by Wabo Jules and his brother, encouraged by Mama Kenmogne Madeleine, Sado Alain and Maliedjie Florence, supported by Matchuim Kouam whose sacrifice I do not know the limits. The difficulty of the latter being mine, I thank those who supported her in her sacrifice, including Papa Kouam Alexis, Mama Youogo Rose and, of course, Fomkouong Mandela.

I think of my friends in the cyber security research group at the University of Dschang and my colleagues and friends at LIA. The interventions of Kouam Willie and the care of Sarkiss Moussa, Njifendjou Ahmed and Christian Cravotto are unforgettable. Thanks to Afaf Arfaoui, Naresh Modina, Samira Habli and Mandar Datar for their availability, sympathy and kindness.

The list is not exhaustive. The important collaboration of Oumaima Diami, for example, has not been mentioned. I would like to express my gratitude to all of them.

Résumé

L'un des aspects délétères de l'évolution des technologies de l'information et de la communication est la combinaison de l'efficacité et de l'efficience dans la propagation de codes malveillants, ce qui constitue clairement une menace pour la sécurité des utilisateurs de ces technologies. Le terme "utilisateur" recouvre ici les individus, les entreprises, les organisations gouvernementales ou non gouvernementales, les États, toute personne ou groupe de personnes qui communique en utilisant les nouvelles technologies. Parmi ces menaces, on peut citer les rumeurs dans un réseau social et le recrutement furtif d'utilisateurs naïfs dans une armée cyber-terroriste capable, par exemple, de causer de graves dommages à une entreprise dont les services sont utilisés par ces mêmes utilisateurs. Dans ces deux cas, comme dans beaucoup d'autres, les utilisateurs, trompés par des experts compétents, participent contre leur gré et contre leur propre intérêt à une cyberattaque dont ils ne sont pas conscients, le support de l'attaque étant la tromperie. De plus, les cybercriminels, contrairement aux cyber-défenseurs, violent les règles de la vie privée et sont donc les mieux, voire les seuls, informés de la vulnérabilité de la cible de l'influence.

Divers modèles issus de la théorie des jeux sont proposés dans la littérature afin d'aborder le contrôle épidémique sous l'angle d'un problème stratégique. Les jeux stochastiques (SGs) sont les types de jeu les plus adéquats pour étudier ce genre de problème pour deux principales raisons : (1) ils s'intéressent au résultat global, appelé utilité, plutôt qu'à la récompense de l'étape courante de jeu ; (2) ils intègrent l'incapacité des joueurs à contrôler l'évolution du système, ce qui traduit la naïveté des utilisateurs. Lorsqu'ils tiennent aussi compte de l'asymétrie liée au fait que les attaquants sont les seuls à connaître de la vulnérabilité des cibles potentielles, ces types de SG sont dits partiellement observables (POSGs).

L'existence d'un processus de propagation épidémique (virus informatique) s'explique par la naïveté des utilisateurs exploitée par des tricheurs. Un moyen d'arrêter les tricheurs est de leur tendre une embuscade. Puisque l'interaction entre les joueurs est répétée et les assaillants connaissent à chaque coup le résultat de leur évaluation, nous proposons donc d'utiliser une embuscade subtile dont la stratégie de positionnement ne sera pas inférée par les cyber attaquants. Cette hypothèse nous écarte des POSGs classiques à deux joueurs et à somme nulle, dans lesquels le joueur qui connaît l'état du système peut inférer l'action de son adversaire.

Dans cette thèse, nous proposons un nouveau modèle de jeu entre un défenseur trompant, dans et par les moyens du cyberespace, un attaquant qui, dans et par les moyens du même cyberespace, trompe les utilisateurs naïfs. Il s'agit d'un POSG à deux joueurs et à somme

nulle dans lequel un seul joueur a une information complète et aucun joueur n'a une information parfaite. Nous abordons aussi la notion d'utilité maximale en tenant compte que les joueurs sont intéressés non pas à la somme des résultats à chaque étape, mais plutôt au résultat le plus critique du processus. Nous proposons enfin un modèle de jeu bayésien (BG) qui prend en compte la topologie du réseau dans la résolution de la propagation active et furtive de l'épidémie.

Nous démontrons que l'algorithme de résolution des POSGs classiques peut être utilisée pour notre nouveau modèle de POSG, même lorsque l'utilité est vue comme la valeur la plus critique du processus. De plus, nous résolvons le jeu bayésien, qui répond au problème de passage à l'échelle mieux que le jeu stochastique.

En plus d'améliorer la cybersécurité en intégrant la cybertromperie dans le contrôle épidémique, le travail de cette thèse propose, d'une part, une idée originale pour la résolution des jeux stochastiques dont l'utilité est l'extremum, d'autre part, d'améliorer la scalabilité de l'algorithme d'itération de valeur en transformant un SG sur un réseau en un jeu de centralité.

Ces contributions sont résumées dans les articles que nous avons publiés au cours de cette thèse, qui sont tous évalués par des pairs. Notre premier travail porte sur la définition et la résolution d'un modèle de jeu qui rend compte des interactions entre un attaquant qui tente d'infecter un réseau pour en prendre le contrôle et un défenseur qui tente de l'en empêcher. L'article "Partially Observable Stochastic Games for Cyber Deception against Network Epidemic" construit un tel modèle de jeu. Il fournit la première démonstration de la convergence de l'algorithme d'itération de valeur dans un POSG à somme nulle à deux joueurs lorsque aucun joueur ne dispose d'une information parfaite. Dans le cas particulier du contrôle des épidémies, la récompense du défenseur est une somme de récompenses résultant des transitions individuelles des nœuds entre les états infecté, sensible et résistant. Nous généralisons cette récompense du défenseur à toute agrégation de récompenses partielles fournies par des transitions individuelles dans "A Partially Observable Stochastic Zero-sum Game for a Network Epidemic Control Problem". Cette généralisation est, avec la preuve de convergence de l'algorithme d'itération de valeur, le sujet du chapitre 5. La fonction d'utilité est ici la somme actualisée. En considérant le facteur d'actualisation assez proche de l'unité, et pour certains paramètres des récompenses partielles, l'utilité de l'attaquant est presque égale au pic épidémique.

Intéressés par cette utilité plus réaliste qu'est le pic épidémique, nous résolvons dans l'article "Optimizing Intrusion Detection Systems Placement against Network Virus Spreading using a Partially Observable Stochastic Minimum-Threat Path Game", et c'est nouveau, les POSGs pour lesquels l'utilité d'un des joueurs est égale à sa plus grande récompense. Nous proposons ensuite un algorithme qui converge vers le pic épidémique optimal. Limités par la non-scalabilité de l'algorithme d'itération de valeur, nous présentons des stratégies de défense intelligentes dans l'article "Game-Theoretic Modeling of Cyber Deception Against Epidemic Botnets in Internet of Things". Ces deux articles sont discutés au chapitre 6.

L'article "A Network Centrality Game for Epidemic Control" repousse cette limite de manière significative avec une autre nouveauté, qui consiste à prendre en compte l'influence

relative des nœuds dans la lutte pour le contrôle du réseau. Le jeu qui en résulte est un jeu bayésien dans lequel les joueurs gagnent l'influence des nœuds qu'ils conquièrent. Par conséquent, la stratégie intelligente consiste à conquérir uniquement les nœuds les plus influents. Le chapitre 7 fournit les détails de ce résultat, qui nous a valu la mention honorable du meilleur article à la conférence internationale GameSec 2022.

Bilan des réalisations

Les résultats de nos travaux ont été examinés par des pairs et publiés dans les revues suivantes ou présentés aux conférences suivantes :

1. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *Partially observable stochastic games for cyber deception against network epidemic*. In International Conference on Decision and Game Theory for Security (GameSec) 2020.

2. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *Game-Theoretic Modeling of Cyber Deception Against Epidemic Botnets in Internet of Things*. IEEE Internet of Things Journal 2021.

3. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *A Partially Observable Stochastic Zero-sum Game for a Network Epidemic Control Problem*. Dynamic Games and Applications 2022.

4. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *Optimizing Intrusion Detection Systems Placement against Network Virus Spreading using a Partially Observable Stochastic Minimum-Threat Path Game*. In International Conference on Decision and Game Theory for Security (GameSec) 2022.

5. Tsemogne O, Kouam W, Anwar A, Hayel Y, Kamhoua C, Deugoué G. *A Network Centrality Game for Epidemic Control*. In International Conference on Decision and Game Theory for Security (GameSec) 2022. **Best Paper Honorable Mention**

Mots clés: Cyber-attaque · Cyber-tromperie · Botnet · Contrôle épidémique · Jeu bayésien · Jeu du chemin minimal · Jeu du chemin à menace minimale · Itération de la valeur · Centralité dans les réseaux

Abstract

One of the deleterious aspects of the evolution of information and communication technologies is the combination of efficiency and effectiveness in the malware spread, which clearly constitutes a threat to the security of the users of these technologies. The term "user" here covers individuals, companies, governmental or non-governmental organizations, states, in short, any person or group of persons who communicate using the new technologies. Among these threats, we can cite rumors in a social network and the stealthy recruitment of naive users into a cyber terrorist army capable, for example, of causing serious damage to a company whose services are used by these same users. In these two cases, as in many others, users, tricked by skilled experts, participate against their will and against their own interest in a cyber attack of which they are not aware, the bearer of the attack being deception. Moreover, cybercriminals, unlike cyber defenders, violate the rules of privacy and are therefore the best, if not the only, informed of the vulnerability of the target of influence.

Various game models have been proposed in the literature that approach epidemic control from a game theory perspective. Stochastic games (SGs) are the most appropriate for two main reasons: (1) they focus on the global outcome, called utility, rather than the reward of the current game stage; (2) they assume the inability of the players to control the evolution of the system, which reflects the naivety of the users. When they also take into account the asymmetry related to the fact that the attackers are the only ones to know about the vulnerability of the potential targets, they are said to be partially observable (POSGs).

The existence of the epidemic can be explained by the naivety of the users, which is exploited by cheaters. The only way to stop the cheaters is to ambush them. Since the process is open-ended and the attackers know the result of their evaluation at each move, we propose to use a subtle ambush whose positioning strategy will not be inferred by the attackers. This assumption sets us apart from classical two-player zero-sum POSGs, in which the player who knows the state of the system can infer the action of his opponent.

We propose a game model between a defender cyber deceiving an attacker who cyber deceives naive users.This is a two-player zero-sum POSG in which only one player has complete information and no player has perfect information. We also address the notion of utility by taking into account that players are not interested in the sum of step outcomes, but rather in the most critical outcome of the process. Finally, we propose a Bayesian game model (BG) that is based on the topology of the network to solve the active and stealthy propagation of the epidemic.

We show that the algorithm for solving classical POSGs holds for our new POSG model,

even when utility is seen as the most critical value of the process, and then we significantly increase the scalability of the solution by solving the Bayesian game.

In addition to improving cyber security by integrating cyber deception into epidemic control, this work proposes, on the one hand, a novel idea for solving stochastic games whose utility is the extremum, on the other hand, one to improve the scalability of the value iteration algorithm by transforming an SG on a network into a centrality game.

These contributions are summarized in the papers we published during this thesis, all of which are peer-reviewed. Our first work deals with the definition and resolution of a game model that captures the interactions between an attacker who tries to infect a network to take control of it and a defender who tries to prevent it. The paper "Partially Observable Stochastic Games for Cyber Deception against Network Epidemic" builds such a game model. It provides the first demonstration of the convergence of the VI algorithm in a two-player zero-sum POSG when no player has perfect information. In the particular case of epidemic control, the defender's reward is a sum of rewards resulting from individual node transitions between infected, susceptible and resistant states. We generalize this defender reward to any aggregation of partial rewards provided by individual transitions in "A Partially Observable Stochastic Zero-sum Game for a Network Epidemic Control Problem". This generalization is, together with the proof of convergence of Algorithm VI, the subject of chapter 5. The utility function is here the discounted sum. Considering the discount factor close enough to unity, and for some parameters of the partial rewards, the utility of the attacker is almost equal to the epidemic peak.

Interested in this more realistic utility that is the epidemic peak, we solve in the paper "Optimizing Intrusion Detection Systems Placement against Network Virus Spreading using a Partially Observable Stochastic Minimum-Threat Path Game", and this is new, the POSGs for which the utility of one of the players is equal to its largest reward. We then propose an algorithm that converges to the optimal epidemic peak. Limited by the non-scalability of Algorithm VI, we present intelligent defense strategies in the paper "Game-Theoretic Modeling of Cyber Deception Against Epidemic Botnets in Internet of Things". These two papers are discussed in chapter 6.

The paper "A Network Centrality Game for Epidemic Control" pushes this limitation significantly with another novelty, which is to take into account the relative influence of nodes in the struggle for network control. The resulting game is a Bayesian game in which players win the influence of the nodes they conquer. As a result, the smart strategy is to conquer only the most influential nodes. Chapter 7 provides the details of this result, which earned us the Best Paper Honorable Mention at the GameSec 2022 international conference.

Achievements

The results of our work have been peer-reviewed and published in the following journals or presented at the following conferences:

1. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *Partially observable stochastic games*

for cyber deception against network epidemic. In International Conference on Decision and Game Theory for Security (GameSec) 2020.

2. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *Game-Theoretic Modeling of Cyber Deception Against Epidemic Botnets in Internet of Things*. IEEE Internet of Things Journal 2021.

3. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *A Partially Observable Stochastic Zero-sum Game for a Network Epidemic Control Problem*. Dynamic Games and Applications 2022.

4. Tsemogne O, Hayel Y, Kamhoua C, Deugoué G. *Optimizing Intrusion Detection Systems Placement against Network Virus Spreading using a Partially Observable Stochastic Minimum-Threat Path Game*. In International Conference on Decision and Game Theory for Security (GameSec) 2022.

5. Tsemogne O, Kouam W, Anwar A, Hayel Y, Kamhoua C, Deugoué G. *A Network Centrality Game for Epidemic Control*. In International Conference on Decision and Game Theory for Security (GameSec) 2022. **Best Paper Honorable Mention**

Keywords: Cyber attack · Cyber deception · Botnet · Epidemic control · Bayesian game · Shortest path game · Minimum threat path game · Value iteration · Network centrality

List of Figures

List of Tables

Nomenclature

$\Delta(X)$ Set of probability distributions over the set X, page 24

$\langle \cdot, \cdot \rangle$ Scalar product, page 33

$\mathbb{1}_P$ Kronecker delta: $\mathbb{1}_P = 1$ if P is true, $\mathbb{1}_P = 0$ if P is false, page 24

$\mathbb{S}_z$, $\mathbb{S}_b$ Stake, feasible stake, page 61

$\mathbb{V}$ Set of value functions, page 69

$\mathbf{R}$ Reward matrix, page 66

$\mathbf{T}$ Transition matrix, page 65

$\mathcal{P}_{\leqslant x}(X)$ Set of all subsets of at most x elements of a set X, page 63

$\mathcal{R}$ Reward function, page 61

$\mathcal{R}^{\text{imm}}_{\pi_1,\pi_2}$ Immediate reward in the stage game, page 70

$\mathcal{R}^{\text{subs}}_{\pi_1,\pi_2}$ Subsequent reward in the stage game, page 70

$\mathfrak{X}$ No observation on node i transition, page 65

$\text{BR}_i(a_{-i})$ Set of best responses of player i to opponents' actions a_{-i}, page 23

lin_X Set of linear functions over a linear space X, page 33

supp Support. $\text{supp}(Pr)$ is the support of any probability distribution Pr, page 118

Tar_i Set of edges through which from node i the attacker propagates the malware, page 63

μ Reward generated to the defender because of its individual transition, page 66

∇ Border of an action or action profile, page 63

ω Observation generated to the defender. Can be infered from the action and the effective network state transition, page 66

∂ Footprint of an action or action profile, page 63

θ, θ_1, θ_2 Path of the game, player 1 and 2's paths, page 31

$a(z)$ Network state at the end of the first stage, resulting from action a taken in network state z, page 63

A_{-i} Cartesian product of the list $(A_i)_{i \in N}$ deprived of its i-th set, page 22

a_{-i} Tuple $(a_i)_{i \in N}$ deprived of its i-th component, page 22

a_i Player i's action, $i \in \{1, 2\}$, page 63

A_i, A Set fo player i actions, set of action profiles, page 61

b^0 Initial belief of the defender upon the network state, page 61

E Set of edges of the network, page 61

H Backup operator, page 70

h Maximum number of IPSs, page 63

I Infected compartment, infected state, page 55

O Observation space, page 61

o_i Observation related to node i transition, page 65

R Removed compartment, removed state, page 56

R Resistant compartment, resistant state, page 58

S Susceptible compartment, susceptible state, page 55

T Transition function, page 61

U_π^V Reward in the stage game, page 70

V Set of nodes (vertices) of the network, page 61

w Aggregation of marginal rewards generated to the defender, page 66

$X \to Y$ Transition of an individual state from values X to Y, page 56

$x^{[t,\hbar]}$ Value of token x at period t if the history is $\hbar$, page 31

$x^{[t]}$ Random variable corresponding to the value at period t of a random sequence $\left(x^{[t]}\right)_{t=0}^{\infty}$. This differs from notation x^t in which t may even not stand for the period, page 30

Z State pace of the system, page 27

z Network (global) state, page 61

z_i Node i's (individual) state, $i \in \{1, 2, \ldots\}$, page 61

CPU Central processing unit, page 45

DDoS Distributed denail of service, page 45

DoS Denail of service, page 44

IoT Internet of things, page 38

NE Nash equilibrium, page 23

players 1, 2 The attacker, the defender, page 61

PWLC pointwise linear and convex functions, page 77

SG Stage game, page 70

Contents

Chapter 1

Introduction

Contents

1.1 Active and Furtive Spread of an Epidemic

According to a report presented by [29], "38% of companies that provide financial services or operate online services for the public experienced a DDoS (distributed denial of service) attack between April 2013 and May 2014." Each incident cost an average of $\$52,000$ for small and medium-sized businesses and $\$444,000$ for large businesses. According to another report, the number of attacks increased by 14% in 2021 compared to 2019, reaching a total of 9.84 million DDoS. [1] The fury of these attacks is largely due to the fact that many people contribute to them without being aware of their involvement. Indeed, DDoS follows a furtive preliminary recruitment of IoT devices into a zombie army called a *botnet* [47]. Armed with their IoT device control system, highly organized hackers with little regard for privacy scan normal users' devices for vulnerabilities. Judicious exploitation of a vulnerability allows code to be injected into the device to gain control of it. However, the device continues to behave as usual, at least in appearance. Under the command of the above control system, it will participate in scanning the devices connected to it and injecting code into one or more others, according to the controller's plans. This process has all the characteristics of an epidemic, in which the subjects are the IoT devices and the pathogen is the code transmitted from subject to subject. In addition, not all subjects have the same response to an exposure: some are vulnerable and others are not. This spread of what is now called an epidemic is the preliminary to the DDoS attack. Indeed, the army of zombies thus formed, once it

[1] `https://www.hipaajournal.com/2021-saw-record-numbers-of-ddos-attacks-on-the-healthcare-industry/`

has reached a sufficiently large number, will be a foothold for a sufficiently large number of requests to a device targeted in advance. As a result, the target device will have exhausted all its resources or bandwidth and will no longer be able to provide all its services, as one of which is targeted by the hackers. On April 2009, on Twitter, rumors falsely present pork consumption as a vector for the spread of the H1N1 swine epidemic [11]. This time, the pathogen is the rumor about the never proven, but easily acquired, link between flu and pork. We assume that this kind of propagation is common in activities such as advertising campaigns, where the subject, as in the above rumor, is the human, but the agent is the desire for the promoted product.

Despite its resemblance to ordinary epidemics, the botnet epidemic differs from others in three ways concerning the vector agent. Indeed, in the case of the botnet, the vector, the team of hackers who control the propagation, is intelligent, rational and cunning. By intelligent and rational, we mean that the actions will always be in line with their selfish interests. Cunning is the expression of their stealth, their stratagem to silently spread the threat. So this is a situation that can be approached from different angles: epidemic patterns, adversity, relationships between subjects and deception. In other words, in addition to being able to be addressed from the point of view of the mathematical study of epidemic models, the mathematical theory of conflicts (game theory), graph theory and cyber deception.

The best response to a distributed attack is obviously a decentralized defense in which nodes make decisions autonomously, as the authors suggest in [86]. However, if we consider that each agent tries to respond to the threat as well as possible and thus design a scenario with a very large number of rational decision makers, we are forced to allow the unrealistic solution that consists in saying how many players of these decision makers should have a given behavior, instead of telling each of them how he should behave [86]. Moreover, in many networked epidemic attacks, most actors do not meet the requirement of rationality and intelligence. Many IoT users do not rigorously evaluate their possible actions and may not even take the best ones, even if they know it. On figure 1.1 for example, the attacker could control the spreading of a virus in a network starting from an input device, and taking control of the infected device each time. Thus, to better mitigate the actions of the attacker, he must be opposed by an opponent who is also rational and intelligent. In other words, effective and realistic control of such a spread requires a defender. In the case where in addition the offensive action is carried out secretly on agents whose decision cannot be controlled, the game theoretical model used should be that of the stochastic game (SG). SGs are realistic because they compare players' strategies, not step-by-step based on the score of each step, but rather by taking into account an aggregation of all step scores of each player. The aggregated score over long periods of time is called utility. The importance of the utility is justified by the fact that the order produced is not a total order. In other words, a strategy may be better at one stage and worse at another.

Generally speaking, SGs are not easy to solve. The author of [32] proposes a convergent algorithm for solving two-player zero-sum partially observable stochastic games (POSGs) in a context where one of the players does not know the state of the system in which he acts. The value iteration algorithm proposed is a version of the same well known algorithm in

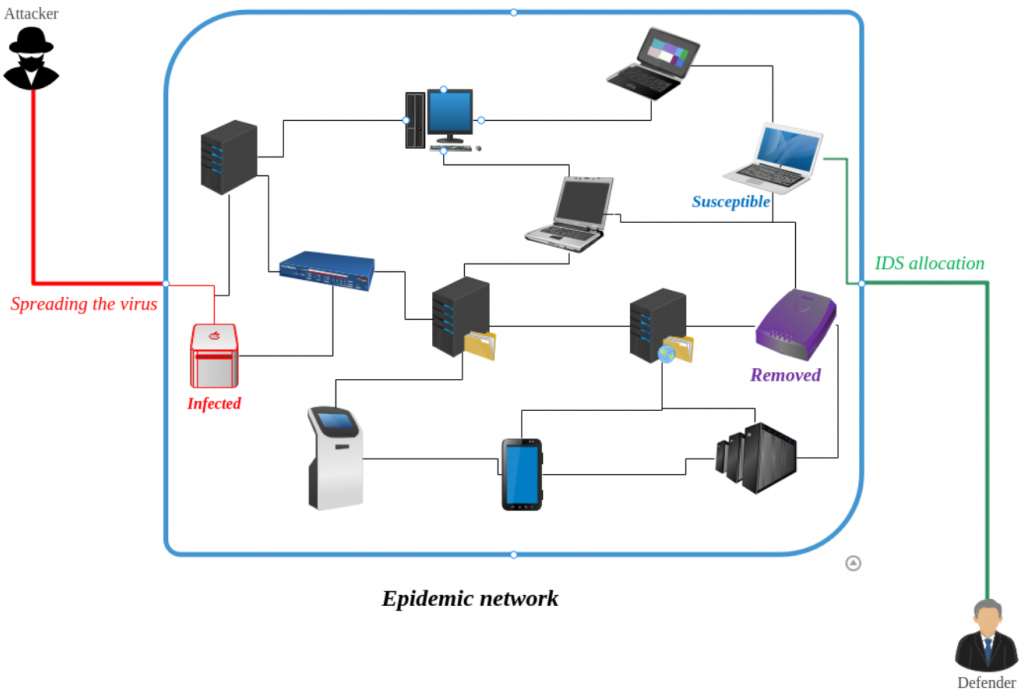


Figure 1.1: Epidemic in network

the context of Markov decision processes (MDP) in [67]. It is used in a lateral movement context, in which the player who knows the state of the network can infer the actions of his opponent. This does not reduce the asymmetry resulting from the one-sidedness of the observability of the system on the one hand, and from the stealthy nature of the attackers' approach on the other.

1.2 Our Contributions

The most general model of epidemic processes is the SIR model. Indeed, on the one hand, the SIS model and its corollary SI are deducible from SIR by abstraction of the compartment R of the withdrawn subjects. On the other hand, the other models are obtained by partitioning one of the compartments S, I and R. The dynamics related to the evolution of an epidemic in a network do not escape this compartmentation. Hackers must assess the vulnerabilities of a device before attacking it. This is a reminder that some devices are invulnerable, i.e., resistant to their eventual assault. This corresponds to the R compartment. Hackers only attack vulnerable devices, i.e., devices that are susceptible to infection. This corresponds to the S compartment. Compartment I is for infected devices with the malicious code and therefore become part of the botnet. Removing a device from this compartment is

equivalent to an etiological treatment of the device, after which the network user can, if he or she chooses, take the action to make the device no longer vulnerable (like installing a patch or an update of a software). Both of these decisions, i.e., moving from infected to susceptible and moving from susceptible to resistant, are beyond the control of the decision makers: those attacking the network and those protecting it. Under the assumption that the attackers constitute one team and the protectors another team, we suppose that each team equips itself with all the hardware and software resources to confront the other. We refer to these teams and their arsenal as the attacker and defender respectively, and assume them to be intelligent and rational. The attackers team (further called attacker) and the defenders team (further called defender) thus faces each other in a battle whose object is epidemic control, the first aiming at increasing the number of victims, the second aiming at reducing it. Any situation like this, involving intelligent and rational individuals with antagonistic interests, can be mathematically modeled and studied considering a game [61]. It is therefore appropriate to approach epidemic control from the perspective of mathematical game theory. In doing so, four essential aspects of the confrontation emerge from the presentation made so far.

- The first aspect is the temporal structure of the interventions of our two intelligent and rational actors, which game theory calls players. The players make their actions several times and in a discontinuous way. We then model their interactions as a repeated game.

- Second, the two players are not the only actors in the confrontation, as the decisions of the network users influence their outcomes and the conditions under which they make their decisions. Because their decisions are not monitorable, we assume that they act probabilistically and thus constitute a random decision entity that can influence the course of the game. It follows that the game is stochastic, the system being the free entity that the devices constitute.

- The third aspect concerns the satisfaction of the two players. Since their objectives are exactly opposite, i.e., one wants to minimize what the other wants to maximize, we are dealing with a zero-sum game.

- Finally, we have to take into account the relationship of each player to the observation. The defender has incomplete and imperfect information, since he does not know the state of the system and cannot infer the decision of the attacker. The attacker has complete information, since he observes the state of the network before taking an action.

We thus obtain a two-player, zero-sum, partially observable SG (POSG). Solutions of standard POSGs can be obtained by the value iteration (VI) algorithm [35]. Authors present an asymmetric battle between a player with incomplete information, the defender, and a player with complete information, the attacker. In the context of botnet propagation for example, the incompleteness of the defender's information comes from the privacy rules, which the attacker overrides. In an effort to preserve an asymmetry to its advantage, the attacker

further complicates defensive action by hiding its presence from the IoT user. Because of this obfuscation, the defender, even if he has a patch, cannot effectively convince users of the need for that patch and, thus, may not always get their approval to fight a common enemy. The target of this deception is the defender, and this strategy of the attacker is cyber deception, i.e., deception in cyber space with an object in cyber space [73].

"If the attacker deceives the defender, let the defender also deceive the attacker!"

In an attempt to reduce the asymmetry, we propose that the defender responds to cyber deception with cyber deception. The strategy will be to detect the transmissions and only react when the attacker is confident that he has succeeded. Thanks to this mixing, the attacker will not be able to infer the action of the defender and, therefore, there will be no asymmetry in the quality of information, i.e. both players will have imperfect information. This POSG model differs from the model solved in [35]. Nevertheless, we establish the convergence of the VI algorithm for this singular model, which by the way, we did not find in the literature. We then review the notion of utility in modeling threat propagation. The utility studied in [35] is the discounted sum, which assumes that the step scores are summed after being discounted. This utility has the advantage of being interchangeable with the mean, but does not take into account the non-additivity of the threat. We are interested in the utility seen as an extremum (minimum or maximum), which is realistic, but left aside in the literature for its complexity. We prove that the VI algorithm is still converging but not scalable. Looking for an answer to this non-scalability issue, we are interested in an intelligent behavior of the players based on the evaluation of the influence of the nodes in the network. The study of the influence of nodes in graph theory brings out the notion of centrality. We thus transform epidemic control into a battle for centrality.

To the best of our knowledge, this work is the first to approach epidemic diffusion from the perspective of POSGs, to study SGs by taking the extremum as a utility, and to propose a smart control strategy based on network topology measures. Our contributions are as follows:

(i) An SG model is defined to address the control of epidemics. This model offers the advantage of taking into account the impossibility for the cyber defender to rely on the collaboration of the network users.

(ii) A mathematical study of a two-player, zero-sum POSGs with extremum value as utility function is given. We propose an algorithm converging to its solution. It is important to recall that this issue concerned also Markov decision processes (MDPs). Because of the non-interchangeability of the sum and the maximum, MDPs whose utility is given by the sum were obsolete. Our resolution applies *mutatis mutandis* to them.

(iii) An answer to the cyber deception of hackers is proposed.

(iv) The use of cyber deception to thwart the active and stealthy propagation of an epidemic is considered. It is therefore a response to cyber deception with cyber deception.

(v) An optimal strategy to obtain advantageous positions in a network in the sense of better ensuring the propagation or control of a threat.

1.3 Strurcture of the Thesis

The remainder of this thesis is organized as follows: after the related work, the next two chapters present the elements necessary to the understanding of this work. Chapter 3, is a brief introduction to game theory. The game models we present here include, in order to simplify the understanding, those we discuss above, while respecting the mathematical thoroughness that must be attached to them. In Chapter 4 we review numerous cyber attacks and present the basic principles of cyber security and cyber deception. The problem we are solving is about epidemic control and therefore its understanding goes through the understanding of epidemic processes. In Chapter 5, after a comprehensive and elicited classification of known epidemics, we identify the one that is the focus of our investigation and then we integrate it into the framework of a new POSG framework that we solve. Thanks to its resolution, in Chapter 6, we tackle the epidemic control problem in a more frontal way, taking into account the non-additivity of the threat. We solve the POSGs whose utility is the extremum value and, at the same time, we bring a realistic solution to the epidemic control problem. We also suggest some intelligent behaviors for the defender. In Chapter 7, to circumvent the non-scalability of the VI algorithm, we replace the epidemic control by a Bayesian game of improved scalability. In the conclusion, we summarize the rules and evidence that guided the choice of a stochastic game and a Bayesian game, and then open perspectives on the exploitation and improvements of our solutions developed in this thesis.

Chapter 2

Related Work

Contents

2.1 Non-Adversarial Study of Epidemics

In mathematics and computer science, a broad scale of epidemic study addresses the issue of virus propagation as in biology, i.e., without consideration of the virus goal. In this scope, [44] uses terms from mathematical modeling of infectious diseases, classifies epidemics and study their propagation in a network through deferential analysis. The epidemic model name relies on the possible compartments of the individuals, susceptible (S), infectious (I), recovered (R), and the possible transitions (S to I, I to R or others) of any individual from one class to another. Intractable differential equations that apply not only in the domain of computer network study put in relations the proportion of each compartment and their evolution. Focusing on network diseases and attempting to immunize network systems, [17], as well as [10,91], presents the Analytical Active Worm Propagation (AAWP) model, which characterizes the propagation of worms that employ random scanning. This model differs from the epidemiological model, which uses continuous time and never considers the patching rate nor the time it takes to a worm to infect a machine. Long before the attacker releases the worm, she scans the system to find out vulnerable machines and establish a "hitlist" of machines that she will further infect and use as "stepping stones" to infect other

vulnerable machines. With random scanning, if there are m_i vulnerable machines and n_i infected ones with a scanning rate of s, then the number of newly infected machines at the following stage will be $(m_i - n_i)\left[1 - \left(1 - \frac{1}{2^{32}}\right)^{sn_i}\right]$ in average. For many authors, there exists an epidemic threshold under which the epidemic dies out. However, the description via Markov chain shows the existence of absorbing states. This paradox suggested in [91] an application of mean field approximation, which is called the N-intertwined model. For the sake of effectiveness, an important aspect of the virus study is the immunization strategies among which the targeted immunization strategy, which is based on immunizing the highest betweenness centrality nodes or links. However, taking possible size of the network that could be infected serves as the performance measure of the immunization procedure. In [76], the authors introduce a method which is significantly more efficient in preventing disease spreading. Another type of immunization, namely random immunization, also presents some weaknesses that [19] overcomes. On the one hand, almost all of the nodes need to be immunized before an epidemic is stopped; on the other hand, when the most highly connected nodes are targeted first, removal of just a small fraction of the nodes results in the network's disintegration. [19] presents an effective strategy based on the immunization of a small fraction of random acquaintances of randomly selected nodes.

Unlike in Biology, network viruses, among which botnets (which are used in DDoS) have targets and are discussed in the literature [2, 47]. The Mirai botnet was first found in 2016 and has been used in some of the largest and most DDoS, including an attack in September 2016 on computer security journalist Brian Krebs' website, an attack on French web host OVH, and the October 2016 Dyn cyber attack. Studying DDoS activities from August 2016 to February 2017, Antonakakis et al. noticed a capability to infect more than 60,000 IoT devices in its first 20 hours and the existence of a steady state. Left apart a frequent check for network status and new prospective target victims, the Mirai attack consists in six steps including a brute-force attack to discover the default credentials of weakly configured IoT devices, the report to the "report server", the issuing of an infected command, the loading in the target devices of an instruction to download the malware (which protects the new recruited bot instances from other malware), (after a sufficient number of hosts is achieved) an instruction to commence an attack against a target server, the attack against the server. The victims range from game servers, telecoms, and anti-DDoS providers, to political websites and relatively obscure Russian sites [2, p. 104].

2.2 Adversarial Study of Epidemics

By considering the intelligence and the rationality of the IoT and the social network actors, it is unavoidable to go against virus propagation with the sight of game theory. Some results permit such studies. For example, [58] generalizes the study in network games in which not all players are connected, proves existence and uniqueness of Nash Equilibrium under few assumptions and outlines the influence of any exogenous shock or addition of links on the Nash equilibrium, under assumption of incomplete information. This Nash Equilibrium

concept is used in [86, 87] to stop the spread of SIS (susceptible-infectious-susceptible) epidemics and to optimize influence in competitive contexts. To compare the advantages of centralized and decentralized protection of a network against threats, Trajanovski et al. [86] discuss the price of anarchy (PoA) in single community, bipartite and multi-community networks. They prove the existence of the Nash equilibrium and outline an algorithm to find the NE in pure strategies. They bounded the PoA (particularly in single community and bipartite networks) analytically and in concrete examples. The upper bound relies on the costs of possible individual decisions (to invest for protection or not), on the number of individuals in the communities and on the spreading rate. To address the issue of designing an optimal network topology while balancing multiple, possibly conflicting objectives such as cost, performance and resiliency to viruses, the authors in [87] model the SIS epidemic with the N-intertwined mean field approximation and consider a network formation game model. As result, they give a new upper bound of PoA, show that the prize of stability is equal to 1 and the NE is achieved only if the graph is a star or a path graph.

Authors in [14, 49] use game theory analysis to tackle virus spread in networks. [49] optimizes the action of a defender monitoring normal nodes against the action of an attacker monitoring attack nodes towards a target server. [14] on its side notices that the impact of the attacker on monitoring scheduling strategy has been neglected in some woks. Thus, regarding the balance between limited resources and scheduling monitoring needs to be considered, it addresses the following problem: the outbreak of propagation process is dynamic and the outbreak detection time is uncertain. It designs a Stackelberg game model (the defender is the leader, the attacker is the follower) for adversarial outbreak detection, through the probability of infection of each adjacent node and proves that the defender's optimal response is NP-hard. Nonetheless, it outlines an algorithm that approximates the optimal response under the assumption that the detection of one infected node inhibits further attacks.

Another game ancient and broadly studied concept, termed stochastic game (SG), is used to improve networks security. [82] generalizes to infinite games (and finite number of states), the existence of equilibrium point proved by Nash in 1950. However some assumptions (observability, reachability of information) generally do not concern network defenders. In addition, there is no indication to draw an equilibrium point.

2.3 Games of Incomplete Information

There is also important studies [5, 33, 79] on games in which some players may not have complete information on the game. Addressing the question of how players use the knowledge of opponent's actions to update his belief about the state, [5] proves the existence of a saddle-point and of the non-equality of minmax and maxmin values. In [33], the authors discuss partially observable stochastic games with public observation (PO-POSG) with two players, i.e., each player has at each stage his private state, his own publicly known observations, his own transition function and his own belief over his opponent states. Motivated by intractability to compute optimal strategies, the authors propose a novel algorithm for

one-sided partially observable stochastic games that converge to the optimal values: the algorithm draws a convenient approximation of the value. These works do not apply directly to control of epidemic process over networks, in which the state is not private and some players who study vulnerabilities of the system before their intrusion (the attackers) know more information than the others (defenders) who observe what the attackers allow them to observe. To overcome this asymmetry, it seems obvious that defenders have to implement deceptive strategies.

2.4 Cyber Deception

Several methods of deception exist in the literature. The article [72] categorizes deception methods based on techniques such as impersonation, delays, forgeries, camouflage, false excuses, and social engineering. Further on, [74], motivated by the observation that traditional cyber security methods involve a never-ending cycle of detection and response to new vulnerabilities and threats, defines concepts related to game theory and clearly outlines eight types of games. Better than humans, computers can examine the huge number of possible threat scenarios in the cyber system. However, here no one is guaranteed to dominate the information in terms of intelligence and accessibility. Hence the importance of game theory for cyber security. The game models used can be static - then imperfect information - or dynamic and help in many tasks such as risk assessment, worm response design, intrusion detection modeling or interactions between attackers and defenders, among others. Their limitations lie in the stringent assumptions (perfect information, fixed transition probabilities and synchronization of player actions) and scalability with size and complexity. This study of game theory and cyber deception is devoted to existing cyber security problems studied from a game theory perspective. After a survey of the six types of deception (perturbation, obfuscation, mixing, moving target, honey-X, and attacker engagement), [66] distinguishes between the different types to allow for accurate game-theoretic modeling of each. It also provides a clear understanding of the challenge of applying game theory to cyber security and privacy and presents the nature of the proposed game concepts to achieve each type of deception.

In the domain of lateral movement, the value iteration (VI) method is used to find the solution of a game in terms of value. Replacing the worst case distinctiveness (wcd) by the more intuitive notion wcd_{ag} and the new expected-case distinctiveness (ecd), [96] proves the existence of solutions. On the one hand, the wcd, which is the longest sequence of actions such that even if it is known that agent is taking these actions to a certain pair of goals, the precise goal remains not revealed. It is bases on observation of pairs of goals. On the other hand, wcd_{ag} takes into consideration the set of all goals and ecd takes into account prior information on the agent goals. [96] also proves that: wcd_{ag} is slightly faster than computing wcd and, in several occasions, ecd is reduced while the wcd_{ag} remains unchanged. Using an heuristic search with the VI (HSVI), [34] replaces the representation of the beliefs of the possible states with a summarised abstraction and increments strategy generation technique

to iteratively expand the strategy space of players. However, HSVI does not serve only for lateral movement.

2.5 Value Iteration in POSGs

The standpoint for mitigating an attack is obviously to get able to observe it. The need of this ability explains the works on intrusion protection and detection systems proposed in [27,38]. Once an intrusion detection system is performed, the network becomes a battlefield with repeated scenario between at least one attacker and one defender. The non-cooperative game theoretic solution concept is generally the Nash equilibrium [48, 61]. This solution is approximated through the value iteration (VI) method in the domain of lateral movement [28]. Replacing the worst case distinctiveness (wcd) by the more intuitive notion wcd_{ag} and the new expected-case distinctiveness (ecd), [96] proves the existence of solutions. On the one hand, the wcd, which is the longest sequence of actions such that even if it is known that the agent is taking these actions to a certain pair of goals, the precise goal remains not revealed. It is based on observation of pairs of goals. On the other hand, wcd_{ag} takes into consideration the set of all goals and ecd takes into account prior information on the agent goals. [96] also proves that: wcd_{ag} is slightly faster than computing wcd and, in several occasions, ecd is reduced while the wcd_{ag} remains unchanged. Using an Heuristic Search with the Value Iteration principle (called HSVI), [34] replaces the representation of the beliefs of the possible states with a summarized abstraction and increments strategy generation technique to iteratively expand the strategy space of players. However, HSVI does not serve only for lateral movement application. HSVI algorithm is discussed in the study of POMDPs (see [78] for example) and, instead of looking for solutions in the value space as usual, [30] searches solutions in the policy space and guarantees the convergence after a finite number of states. This HSVI algorithm is generalized in [35] to the study of OS-POSGs which respond to the undecidability of real-world security scenarios, that require POSGs and no strictly defined horizon. Horák et al. use a contracting mapping over the set of (possible) value functions to guarantee the convergence to the value of the game and outline an algorithm that approximates optimal strategies in OS-POSGs. They assume the worst case scenario when the attacker has complete observation. Later, to overcome the complexity of representing, updating and reasoning about the uncertainty of the defender on a very large state space, [31] improves the scalability of POSG algorithms by replacing the representation of the beliefs by a characteristic vector that captures key information but reduces the dimension of the beliefs. Algorithms are applicable in a network of 11 nodes with not more than 1 GB of memory consumption are strategy found are very near the optimal solutions for small cases.

There also exists games where players should not be viewed as optimizers, but where one player (preciser) may be trying to have a precise value. [15] studies such games played on a *stochastic game arena*, i.e. a graph whose vertices are partitioned into those controlled by either of the players. It proves that: the condition of existence of winning strategies for Preciser solves counter-strategy problem and discounted reward controller synthesis problem.

The game is not determined and to apply it, one should know the value targeted by the preciser, which is likely a non-realistic challenge.

2.6 Centrality measures

A wide range of real-world phenomena, from social and information to technological and biological networks, can be described using complex networks which can be modeled using graphs for their performance analysis [26]. Diffusion, as a means of studying a complex network's dynamic behaviors, has been one of the most important topics in this area. Diffusion on the network is transferred from one node to another and it starts on a small scale and then affects more neighbors. Moreover, in a diffusion scenario such as epidemic propagation, the goal is to find influential nodes which have a higher diffusion power in comparison with other nodes. Diffusion in complex networks has a lot of applications and based on the nature of a problem, specific influential nodes can be used to accelerate, control or prevent diffusion. For instance, in marketing on social networks, the highest amount of ads can be diffused with the least amount of time and resources using influential nodes. In computer networks, the spreading of viruses can be prevented by securing the most suitable nodes.

Since finding influential nodes is an NP-hard problem, some approximate methods called centrality measures are used and can be divided into three types: Local measures, global measures and semi-local measures [55].

- The first and simplest centrality measure in the category of local measures is the degree centrality. In this measure, only the first-degree neighbors of a node are considered important. In fact, a node is regarded as important if it has a higher degree. This measure can determine a node's importance to some extent but nodes with the same degree do not necessarily have the same essential role in the graph. This measure because of being local and ignoring the graph's global information, and with a linear time complexity of O(n), does not have high accuracy [26]. One of the popular measures in the category of local measure is the degree centrality that Measures how many neighbors a node contains. There are two types of degree centrality (in-degree and out-degree) according to direction. It measures the immediate influence and it is used to calculate the central nodes in the simulation network, assesses how difficult it would be to isolate a given node (using edge disjoint k-path centrality), etc.

- In contrast, global measures need the entire graph's information to do so, and therefore have a higher accuracy and time complexity. As a result, they cannot be used for large-scale networks. One of the important measures in the category of global measures is the betweenness centrality. The goal of betweenness centrality is to determine the importance of a node based on the information flow existing within the graph. It is based on the number of times a node is located in the shortest paths among all the pairs of nodes in the graph. The high betweenness centrality of a node indicates that it is located between most of the shortest paths available in the graph [25].

- In recent years, due to increases in the size of real-life networks, a need for measures with high accuracy and low time complexity compared to global measures was felt. Measures in the semi-local category have been introduced for this reason. One of the most important and early semi-local measures is the LC measure. The LC centrality is an advanced version of the degree measure in which, in addition to the first-degree neighbors, second-degree neighbors are taken into consideration as well. Another measure recently developed in this category is measure based on clustering coefficient [104].

2.7 Bayesian Game

Chen [16] in his doctoral dissertation used game theoretic model to design the response for the importance-scanning Internet worm attack. The main idea is that defenders can choose how to deploy an application, that is the group distribution, when it is introduced to Internet to minimize the worm propagation speed. The attacker can choose the optimal group scanning distribution to maximize the infection speed. Thus a game would be played between the attacker and the defender. The attacker should choose so as to maximize the minimum speed of worm propagation, while defender wants to minimize the maximum speed of worm propagation. By framing the problem this way it turns out to be a zero sum game and a min-max problem. The optimal solution for this problem is that defender should deploy the application uniformly in the entire IP-address space or in each enterprise network, so that the best strategy that the attacker exploits is equivalent to random scanning strategy. This work gave a game theoretical framework to design the locations of vulnerable and high value hosts over a network.

Patcha et al. [65] proposed a game theoretic approach to model intrusion detection in mobile ad-hoc networks. The authors viewed intrusion detection as a game played between the attacker node and the IDS hosted on the target node. The objective of the attacker is to send a malicious message with the intention of attacking the target node. The modeled game is a basic signaling game which falls under the domain of multi-stage dynamic non-cooperative game.

Bloem et al. [7] modeled intrusion response as a resource allocation problem based on game theory. A cost is associated with attacks and responses. This problem, including imperfections in the sensor outputs, was first modeled as a continuous game. The strategies are discretized both in time and intensity of actions, which eventually leads to a discretized model. The reaction functions uniquely minimize the strictly convex cost functions. After discretization, this becomes a constrained integer optimization problem. To solve this they introduced their dynamic algorithm, Automatic or Administrator Response algorithm (AOAR). They classified attacks into those resembling previous attacks and those that do not, and many such intuitive classes with Kohonen self-organizing maps, a neural net, and the response cost is minimized. The simulations captured variation in vulnerability, value and cost of actions. Their results showed system performs improves after using AOAR.

Chapter 3

Game Theory

Contents

3.1 Introduction

In a general case of decision problems, one or more actors (the decision makers) have to choose between more than one alternative (called ***action***), with aim to optimize the outcome, that relies on the choice of the decision makers. In this thesis, we pay interest only to rational, intelligent decision makers, called ***agents*** – or players, if they are many. The set of agents who do not meet this requirement is referred to as the ***system***.[1] The adjective ***intellegent*** here means that the decision maker "knows everything that we know" about the problem, and "he can make any inferences about the situation that we can make"; the ***rationality*** means that the decision maker "makes decisions consistently in pursuit of his own objectives" [61]. The different ways the system can influence the outcome of the decision makers is refered to as ***states***. Consider, for example, a firm that suspect a malicious activity that may exploit IoT devices to launch a distributed denial of its service. The firm can be considered as an agent, while the IoT users should be regarded as the system. To meet its objectives, the firm may decide to choose two devices and monitor the data exchange between them (action). The relevance of the decided action depends upon the influence and the vulnerability of the pair of choosen nodes (state of the system).

A wide range of decision problems, say games, involve multiple agents, henceforth called ***players***, with conflicting interests. It is worth nothing to note that multiple agents who share the same interests should be regarded at as a single agent. Precisely, a ***game*** is "any social situation involving two or more agents" [61]. The social situation implies that the decision of all player influences the outcome of other players. In addition, this decision works against the interest of at least one other agent.

This chapter is about the description and the classification of discrete time games with numeric outcomes and finite sets of players and actions. When the numerical outcome represents a gain or a cost, we refer to it as a reward. More precisely, the ***reward*** of a player is his outcome at the end of a period of the game. When actions are taken repeatedly, the outcome may be more general and represent the overall satisfaction of the player at the end of the process. We refer to this overall satisfaction as ***utility***. The classification of a game model relies on different considerations including, but not only, the number of players, the reward structure, how the decision-making is scheduled, the influence of the system upon the outcome, the knowledge of players about the system and reward structure, the number of times the game is played.

3.2 Strategic-form Games

When each player reward is given at once, the game is said to be a ***one-stage*** one. The reward and the utility are the same thing in this case. Like in [61], throughout this thesis,

[1] Unlike in [67] and since the state takes decisions (probabilistic transitions), we consider that the MDP involves two decision makers, exactly one of whom is intelligent and rational. This decision maker is the agent. The word "player" usually refers to a conflictual situation. The system is considered as an actor who does not defend any interest. It is therefore assumed that it is made up of all the actors who are not players.

it is admitted that odd-numbered players are male and even-numbered players are female.

3.2.1 Definition and Representation

When no player knows the ***move*** (i.e., the action decided) of her opponent at the moment she makes her choice, the situation is the same as if all players act simultaneously, and the game is said to be ***simultaneous***; otherwise the game is said to be ***sequential***. [2] In the general case of n-player games with $n \geqslant 2$, a game is said to be sequential if players take their decisions in any order, provided at the time he makes his move, no player is informed of the decision of the players who decided before him. In case the state of the system is unique and known to all players, and all players know how exactly how the overall decision determine the players' outcomes, the simultaneous game is called ***normal- (or strategic-) form game***. A normal-form game can be modelled by a tuple

$$\mathcal{N} = \left(N, A, (r_i)_{i \in N}\right)$$

where:

- N is the finite set of players;
- For all player $i \in N$, A_i is the set of actions (or action space) of player i. All tuple $a = (a_i)_{i \in N}$ representing an action for each player is referred to as ***action profile***. $A = \prod_{i \in N} A_i$ is the set of action profiles.
- For all player $i \in N$, $r_i \colon A \longrightarrow \mathbb{R}$ is the ***reward function***, with $r_i(a)$ being the reward generated to player i by the system when action profile a was taken.

In the special case of two-player game, the set N of players can be brushed out of the notation. If, in addition, the game is a ***zero-sum***, i.e., the sum of players' rewards always takes the value 0, only the reward $r = r_1$ accounts in the notation.

The same tuple $\mathcal{N}$ is used to represent an extensive-form game. In this case, however, it should be supposed that player i takes his move before player $i + 1$.

Example 3.2.1. *Consider a network defender who, one week, will invest* $400,000$ *francs either on Saturday or Sunday to protect his network against an attack which will take place on one of these two days and will cost the attacker* $300,000$ *or* $500,000$ *francs respectively. Assume that the defense technique is efficient, and: a successful attack rewards* 15 *to the attacker and incurs a cost of* 8 *to the defender, while an abortive attack results in nothing for both players.*

This situation is a game with two players, including the defender and the attacker, each of them having $A_1 = A_2 = \{Saturday, Sunday\}$ *as action space.*

In case defender and attacker make their decisions simultaneously, they are involved in the strategic-form game represented by table 3.1.

[2] The sequential game is the subject of section 3.3.

		Attacker	
		Saturday	Sunday
Defender	Saturday	$(-4, 3)$	$(-12, 10)$
	Sunday	$(-12, 12)$	$(-4, -5)$

Table 3.1: Example of normal representation of normal-form game
The game is a matrix game.

3.2.2 Solution of a Game

A game is a joint optimization problem of several intelligent and rational actors, and the most relevant solution concepts rely on dominance. If one action dominates another, i.e., it gives a better result regardless of the actions played by the other players, there is no reason for the player to play a dominated action. Thus, each player will prune out his dominated strategies. Formally, we distinguish and define two levels of domination as follows:

Definition 3.2.1 (***dominated action***)**.** *Take any player $i \in \mathcal{N}$ and any actions $x_i, y_i \in A_i$ of player i.*

1. *Action x_i is said to* (weekly) dominate *action y_i if:*
   (a) *For all action profile $a_{-i} \in A_{-i}$ of the other players (i.e., players in $N \setminus \{i\}$), it holds: $r_i(a_{-i}, x_i) \geqslant r_i(a_{-i}, y_i)$;*
   (b) *The inequality is strict for at least one strategy profile of the other players, i.e., $\exists a_{-i} \in A_{-i}$ such that $r_i(a_{-i}, x_i) > r_i(a_{-i}, y_i)$.*
2. *Action x_i is said to* strongly dominate *action y_i if, the inequality is always strict for all action profile $a_{-i} \in A_{-i}$ of the other players (i.e., players in $N \setminus \{i\}$), it holds: $r_i(a_{-i}, x_i) > r_i(a_{-i}, y_i)$.*

Remark 3.2.1. *For each profile a (with indices in N), we use a_{-i} for the tuple $(a_i)_{i \in N}$ deprived of its i-th component. We also use A_{-i} to denote the cartesian product of the list $(A_i)_{i \in N}$ deprived of its i-th set.*

In the strategic-form game in example 3.2.1 (or table 3.1), the strategy *Saturday* of the defender outperforms his strategy *Sunday* if the attacker plays *Saturday*, while the reverse happens if the attacker plays *Sunday*. So, no defender strategy is dominated. One can note that no attacker strategy either is dominated.

In order to integrate the dependence of each player's outcome on the actions of all his opponents, the solution of the game must be considered in terms of an irrefutable consensus. The notion of dominance can therefore be extended to action profiles. It is said that an action profile dominates another when one of the players would benefit from unilaterally deviating from the second profile towards the first and would therefore not harm any other player. This notion is known as Pareto-dominance and can be formalized as follows:

Definition 3.2.2 (***Pareto-dominated action***). *An action profile a* Pareto-dominates *an action profile b if:*

1. $r_k(a) \geqslant r_k(b)$ *for all player* $k \in N$;
2. $r_i(a) > r_i(b)$ *for all some player* $i \in N$.

The search for a solution in which each player avoids causing harm to his opponents fits well with the intelligence of each player and the fact that this intelligence is known to all. However, although it aims for a consensus freely accepted by all players, it contradicts their rationality. Indeed, every intelligent and rational actor is selfish and accepts a consensus only if he cannot do better. The balance thus proposed by John Forbes Nash Jr aims to optimize each player's regarding his opponents' actions [62].

Definition 3.2.3 (***best response***). *Take any player $i \in \mathcal{N}$, any action $x_i \in A_i$ of player i and any action profile $a_{-i} \in A_{-i}$ of the other players.*
The action x_i is said to best respond*, and called* best response*, to action profile a_{-i} if*

$$r_i(a_{-i}, x_i) = \max_{y_i \in A_i} r_i(a_{-i}, y_i),$$

i.e., for all $y_i \in A_i$, it holds: $r_i(a_{-i}, x_i) \geqslant r_i(a_{-i}, y_i)$. The set of best responses of player i to a_{-i} is noted $\mathrm{BR}_i(a_{-i})$*. This is,*

$$\mathrm{BR}_i(a_{-i}) = \arg\max_{y_i \in A_i} r_i(a_{-i}, y_i). \tag{3.1}$$

Considering the intelligence and the rationality of all players, the solution of the game, if it exists, is an action profile in which each player is best responding to his opponent's actions. Such an action is referred to as a Nash equilibrium.

Definition 3.2.4 (***Nash equilibrium, NE***). *A* Nash equilibrium (NE) *is any action profile a^* such that for all player $i \in N$, it holds:*

$$a_i^* \in \mathrm{BR}_i\left(a_{-i}^*\right). \tag{3.2}$$

If the NE exists and is unique, no player can unilaterally benefit in deviating from it. Indeed, for all action $b_i \in A_i$, from equations (3.1) and (3.2), it comes:

$$r_i\left(a_{-i}^*, b_i\right) \leqslant r_i\left(a_{-i}^*, a_i^*\right), \tag{3.3}$$

i.e., playing any other action instead of a_i^* can never increase his reward while the other players play their actions a_{-i}^*.

An NE should be viewed as a consensus from which no player will deviate if adopted. This traduces the fact that all player is intelligent and rational and assumes that his opponents will act intelligently and rationally. Hence, clearly, the game solution, if some exists, is

necessarily an NE. However, it may be troublesome to have more than one NE. Indeed, two different players may be assuming two different consensus, and this may result not in an NE. Otherwise, in some situations like the strategic-form game in the above example, the NE may not exist. The following subsection shows an approach such that each player does not decide his actions, but rather states them from a probability distribution that best matches his opponents'.

3.2.3 Mixed Action

So far, we have considered that a player's decision consists in choosing an action among all his actions. In this case, a decision means an action and is also called ***pure action***. In the case of non-existence of a Nash equilibrium, it is assumed that each player tries to optimize his result in terms of mathematical expectation. In this case, a player's decision would not be to perform a specific action, but to establish a probability distribution over all of his actions, which should guide his choice. The decision is then called mixed action.

Definition 3.2.5 (***mixed action***, ***decision***)**.** *A* mixed action (or decision) *for player $i \in N$ is any application $\pi_i \in \Delta(A_i)$, where $\Delta(X)$ is the set of probability distributions over any set X.*

Note that a pure action a_i should be viewed as the mixed action $x_i \longmapsto \mathbb{1}_{x_i=a_i}$ that takes its values in $\{0,1\}$, where $\mathbb{1}_P$ is somehow the Kronecker delta for any proposition P, defined by $\mathbb{1}_P = \begin{cases} 1 & \text{if } P \text{ is true} \\ 0 & \text{if } P \text{ is false} \end{cases}$. That is, a pure action is a mixed action, and the NE can be extended for mixed actions as follows:

Definition 3.2.6 (NE, mixed-actions Nash equilibrium)**.**

- *A* **decision profile** *is any tuple $\pi = (\pi_i)_{i \in N}$ of decision for each player, i.e., $\pi_i \in \Delta(A_i)$.*
- *The expected reward of a player i when a decision profile π is played is also referred to as his reward. In other words, the* **reward** *of player i in the decision profile π is*

$$r_i(\pi) = \sum_{a \in A} \pi(a)\, r_i(a) \quad = \quad \sum_{a \in A} \prod_{j \in N} \pi_j(a_j)\, r_i(a) = \sum_{a_i \in A_i} \pi_i(a_i)\, r_i(\pi_{-i}, a_i), \quad (3.4)$$

  where $r_i(\pi_{-i}, a_i) = \sum_{a_{-i} \in A_{-i}} \pi_{-i}(a_{-i})\, r_i(a_{-i}, a_i)$ is the reward of player i playing action a_i against decision π_{-i} of his opponents.

- *A* **best response** *of a player i to his opponents' decisions π_{-i} is any decision d_i^* such that*

$$r_i(\pi_{-i}, d_i^*) = \max_{d_i \in \Delta(A_i)} r_i(\pi_{-i}, d_i). \quad (3.5)$$

  As with pure actions, the set of best responses is noted $\mathrm{BR}_i(\pi_{-i})$.

- *A decision profile π^* is said to be an* **NE** *if each player best responds to his opponents in π^*.*

3.3 Extensive-form Games

Unlike in the previous section, there exists some games where not only some players may have some information at the time they take their decision, in addition, each player's turn is unpredictable and is determined by the sequence of actions. Moreover, regarding some conditions, the rules of the game may give some players the possibility to act several times or impose to some others not to act at all during the process. It seems natural to represent these sequences of actions by a figure that captures the decision epoch of every player.

3.3.1 Representation

A game in which players act sequentially may not be a sequential game, but can be represented by a rooted tree such that each edge ending to the root node represents an action of the system, known as actor 0, each edge adjacent to an edge that represents an action of actor i represents an action of player $i+1$ – note that actors 1 to $|N|$ are players 1 to $|N|$. The leafs represent the outcome profiles and each other node represents an actor at the moment he is taking his decision is called ***decision node***. The path from the root node to any leaf node represents an action profile. See figure 3.1, as an illustration of possible scenarios related to example 3.2.1. The game is then said to be described in the extensive form and called ***extensive-form*** one. The tree is referred to as the ***extensive representation***. In this representation, for each player, two decision nodes at which the player has exactly the same information are said to belong to the same ***information set*** of the player.

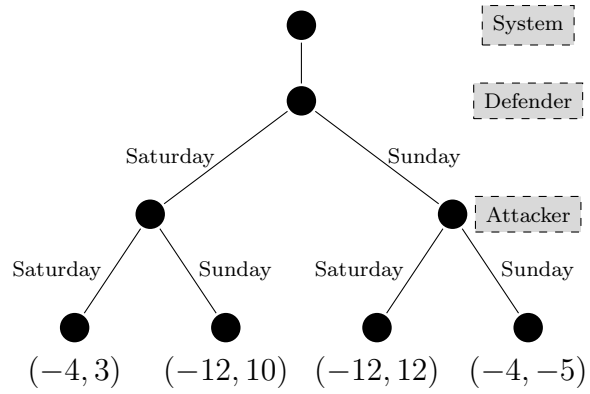
As depicted on figure 3.1 (compare for example 3.1a to 3.1b), the key difference between sequential and extensive-form games is the "quality" of the information available to players. In a sequential game, at the moment a player makes his decision (i.e., at the decision node), he perfectly knows all the moves that have been made before. In other words, two distinct decision nodes do not belong to the same information set. Hence, a sequential game is perfect information, perfect recall game, where:

- An extensive-form game is said to be of ***perfect information*** if the action of each agent is revealed to his followers before they make their decisions;
- An extensive-form game is said to be of ***perfect recall*** if each player remember all the information he has collected at the time he makes his decision.

A strategic-form game can be described as an extremely imperfect information extensive-form game with one move per player.

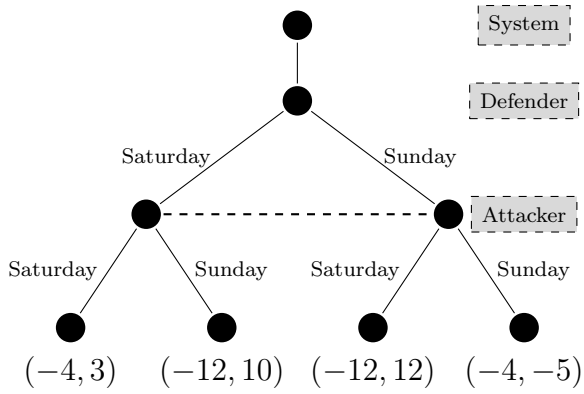
3.3.2 Game with Incomplete Information or Partial Observation

Perfect and imperfect information refer to the past moves of the actors. A player should also pay interest in the rewards-relevant data, that capture the players' preferences or the decision of the state. When all the players' preferences and the system state are common knowledge, the game is said to be of ***full observation***. Otherwise, the game is a ***Bayesian***



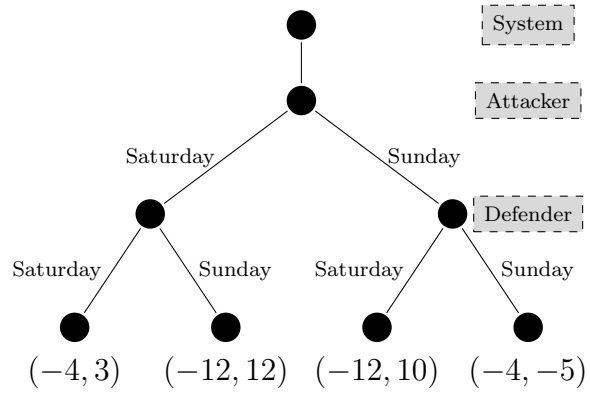
The defender acts first and his choice is revealed to the attacker

(a) The attacker decides second



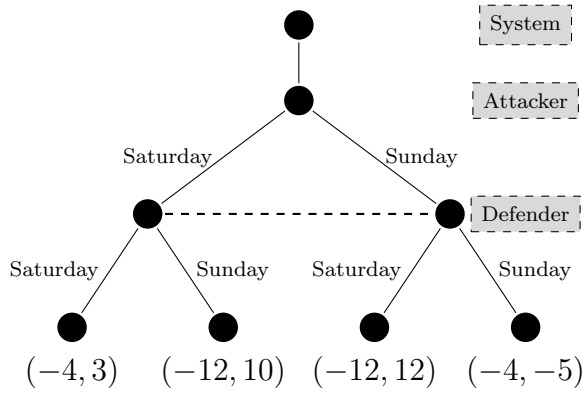
The defender acts first, and his choice is not revealed to the attacker

(b) The attacker decides second



The attacker acts first, and his choice is revealed to the defender

(c) The defender decides second



The attacker acts first, and his choice is not revealed to the defender

(d) The defender decides second

Figure 3.1: Example of extensive representation of a game

The dashed line indicates that at the moment the second makes her move, she does not know what action has been taken before, i.e., she does not know at what of the two possible nodes she is. So, figures (b)and (d) are extensive representations of a simultaneous game, while figures (a) and (c) represent a sequential game.

one [102]. When the state of the system is common knowledge, the game is said to be of ***full observation***.

3.3.3 Strategy

An intelligent and rational actor would certainly adapt his choices to the information at his disposal. This information includes the actions and the system states he was informed of and still know. In this section, we limit to player responses to the state of the system at the time he makes his move in the case of complete information games with several states, or to the history in the sequential games. We discuss the more general information in the next section, along with the stochastic games.

Definition 3.3.1 (strategy, Nash equilibrium strategy profile)**.** *1. A* **pure strategy** *for player i in strategic-form game (respectively, in sequential game) is any application $\psi\colon Z \longrightarrow A_i$ (respectively, $\psi\colon \prod_{j=0}^{i} A_j \longrightarrow A_i$), i.e., a rule that prescribes an action when the state of the system (respectively, the history of the other actors' move) is given.* [3]

2. A **mixed strategy** *for player i is any element of $\Delta(\Psi_i)$, i.e., a probability distribution over his pure strategies.*

3. A **(behavioral) strategy** *for player i in strategic-form game (respectively, in sequential game) is any application $\sigma\colon Z \longrightarrow \Delta(A_i)$ (respectively, $\sigma\colon \prod_{j=0}^{i} A_j \longrightarrow \Delta(A_i)$), i.e., a rule that prescribes a decision when the state of the system is given. The set of strategies for player i is noted Σ_i, and the set of strategy profiles is noted Σ.*

4. An **NE** *in any of the above strategy concepts (pure, mixed or behavioral strategies) is a strategy profile such that the strategy of each player is the best response to his opponents' strategies regarding the underlined strategy concept.*

Some decision concepts are generalizations of some others. Another example is given by the Kuhn's theorem [1], that states the equivalence between mixed and behavioral strategies. [4] All the generalizations are summarized in figure 3.2.

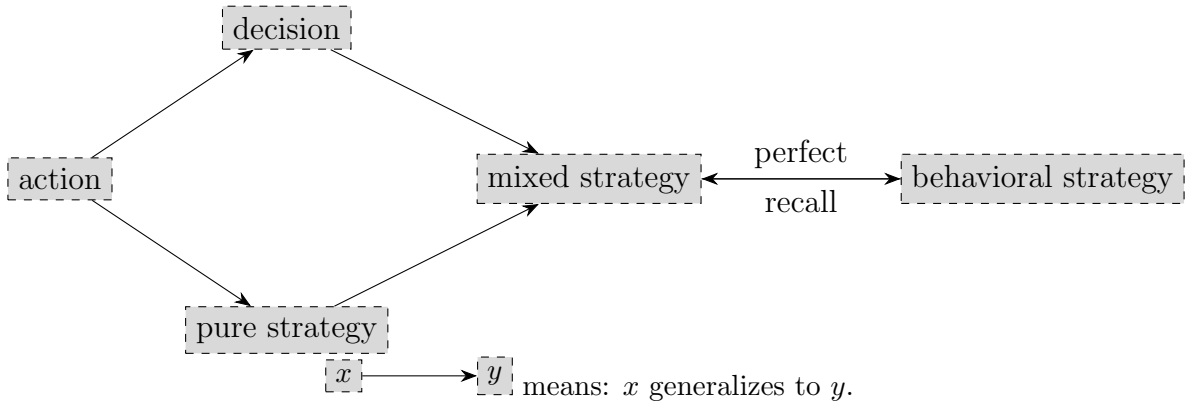


Figure 3.2: Decision concepts

[3] Z stands for the state pace of the system. The set of pure strategies for player i is noted Ψ_i, and the set of pure strategy profiles is noted Ψ.

[4] The Kuhn's theorem states that both strategies are equivalent if and only if the game is of perfect recall.

3.4 Complementary Study of Game Classification

3.4.1 Maximin Game

When the rewards of all players always sum to 0, the game is referred to be ***zero-sum*** one. In a two-player zero-sum game, the reward of each is exactly the opposite of his opponent's. For this reason, we define the ***reward of the game*** to be the reward of player 1. It turns out that player 1 aims at maximizing the reward, while player 2 aims at minimizing it. The two-player zero-sum game is therefore called ***maximin game***.

Here are some important properties on maximin games:

1. A maximin game admit at least one NE.
2. The reward of each player in a maximin game NE is independent on the NE.
3. In a maximin game, there cannot exist more than one NE in pure strategy.

3.4.2 Dynamic Game

When the rules of a game allow one or more players to act several times, the game is said to be ***dynamic***; otherwise, the game is said to be ***static***. Some dynamic games, referred to as ***repeated games*** consist in repetitions of some "base game" called ***stage game***. Each occurrence of the stage game is equally called ***decision epoch***, ***time***, ***time-slot***, ***instant*** or ***period***. [5] A repeated game may be of ***finite horizon*** $\mathcal{T}$, if the number of periods is bounded by $\mathcal{T}$, or of ***infinite horizon*** (or horizon ∞), otherwise.

3.4.3 Stochastic Game

In repeated games, it is assumed that the state of the system may change from one period to another, and this is called a ***transition***. When the system state transition is not predictable, it is assumed that the future state probabilistically depends on the current state and the players move, and the game is said to be ***stochastic***. Stochastic games constitute a bridge between game theory and Markov decision processes (MDP). Indeed, although being repeated games, they can be perceived as competitive MDPs in a situation of adversity [24].

3.4.4 Summary of Games Classification

Games classification is partially summarized in figure 3.3.

[5] Some authors also refer to each occurrence as stage. However, we reserve another meaning for this word (see subsection 5.2.3).

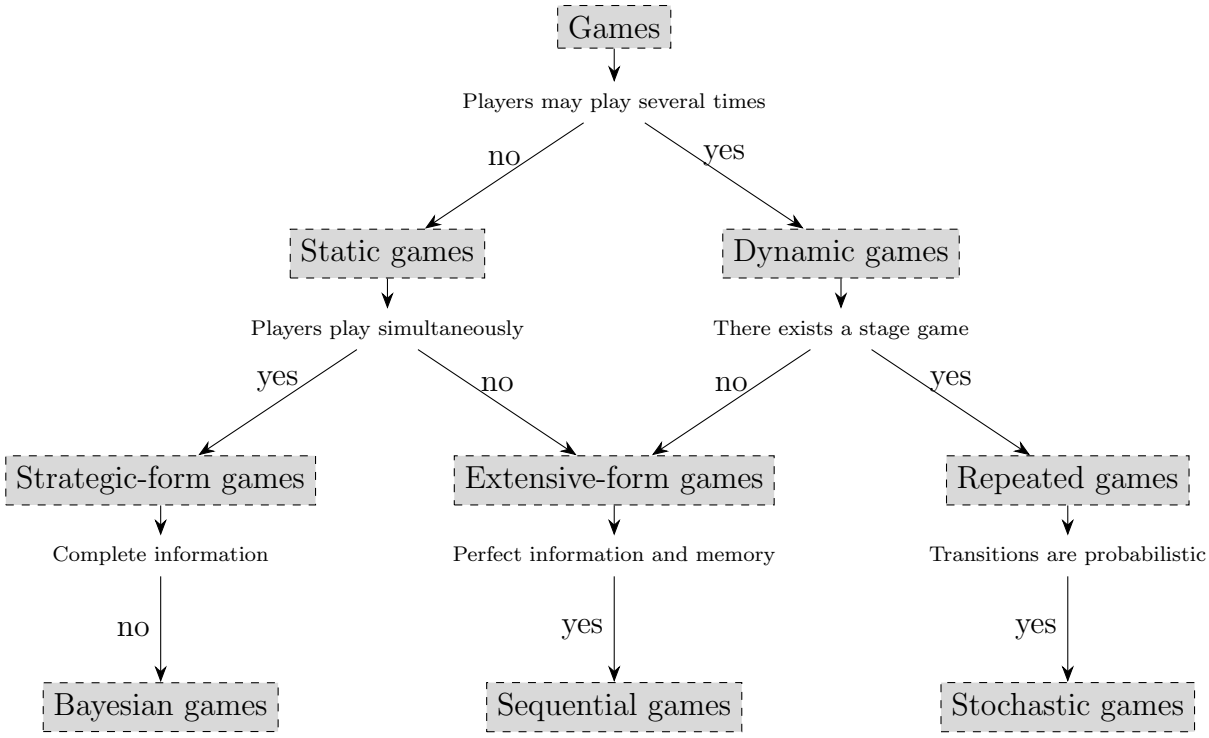


Figure 3.3: Typology of game models

3.5 Two-Player Zero-Sum Partially Observable Stochastic Games

When a POSG involves two players with opposite rewards, the game is a two-player zero-sum stochastic game (***SG***). In this section, we present the key components of a maximin perfect recall SGs in which player 1, only, may not know the state of the system, and no player observes the moves of his opponent. Such games are therefore two-player zero-sum partially observable stochastic games (***POSGs***). The latest statement makes the underlined models both-sided imperfect information games. We prove in chapter 5 that the result of the case of perfect information on the player 2's side hold in these frameworks. Finally, we make the assumption that the game is with infinite horizon. This generalizes the finite horizon with the condition that the reward is equal to 0 after the horizon.

3.5.1 Definition

A ***two-player zero-sum goal-POSG*** is any tuple

$$\mathcal{G} = \left(Z, O, A, T, r, b^0, Z_{\text{goals}}\right)$$

that, with the assumptions of common knowledge and rationality, represents the following scenario infinitely repeated in the time divided in periods: two intelligent and rational individuals, call them players 1 and 2, act simultaneously and independently upon a system.

Their actions induce opposite rewards to both players and probabilistically result in a transition of the system. However, for some particular state values called goal states, from the period the state of the system takes any of these values onward, no effective transition will happen and players will not be rewarded. When the set of goals is ignored, the goal-POSG is called ***two-player zero-sum POSG***. The state is always known to player 2 while player 1 may be unable to infer it, unless it is a goal state. In the above notation:

- Z is the finite set of possible states of the system; $Z_{\text{goals}} \subseteq Z$ is the set of goal states;
- A_i, $i = 1, 2$, is the finite action space of player i, and $A = A_1 \times A_2$ is the set of action profiles;
- O is the finite set of possible observations for player 1;
- $T\colon \begin{array}{ccc} Z \times A & \longrightarrow & \Delta(Z \times O) \\ (z,a) & \longmapsto & T(\cdot \,|\, z,a) \end{array}$ is the transition function, with $\Delta(Z \times O)$ denoting the set of probability distributions over the set $Z \times O$;
- $r\colon \begin{array}{ccc} Z \times A & \longrightarrow & (\mathbb{R}_-)^Z \\ (z,a) & \longmapsto & r(\cdot \,|\, z,a) \end{array}$ stands for the stage reward function for player 1; [6]
- $b^0 \in \Delta(Z)$ is the belief of player 1 over the state of the system at the beginning of the game.

At the beginning of every period $t \in \{1, 2, \ldots\}$, each player $i \in \{1, 2\}$ selects an action, which is consequently denoted $a_i^{[t]}$, from his action space A_i. As a result of the action profile $a^{[t]} = \left(a_1^{[t]}, a_2^{[t]}\right)$, taken in state $z^{[t]}$, the state of the system transitions into the state $z^{[t+1]}$ at the end of the period and generates the observation $o^{[t]}$ to player 1, according to a probability distribution P that satisfies: $P\left(z^{t+1} = z', o^{[t]} = o \middle| z^{[t]} = z, a^{[t]} = a\right) = T(z', o|z, a)$; the system also generates the reward $r^{[t]} = r\left(z^{[t+1]} \middle| z^{[t]}, a^{[t]}\right)$ to player 1 and the reward $\mu^{[t]} = -r^{[t]}$ to player 2. Alternatively, one may not care the exact reward $r(z'|z, a)$ resulting from the transition to state z' when action profile a was taken in state z, and replace the exact reward function r in the above tuple $\mathcal{G}$ by the expected reward function

$$\begin{array}{rccl} \mathcal{R}\colon & Z \times A & \longrightarrow & \mathbb{R} \\ & (z,a) & \longmapsto & \mathcal{R}(z,a) = \displaystyle\sum_{(z',o) \in Z \times O} T(z', o|z, a) \cdot r(z' \,|\, z, a)\,. \end{array}$$

Only one player, player 2, is always aware of the transition; player 1 only makes the observation $o^{[t]} \in O$ and consequently updates his belief from $b^{[t-1]}$ corresponding to the beginning of period t to $b^{[t]}$ corresponding to the end of the period. Without any loss of generality,

[6] In the case of POSSPGs (see subsection 3.5.2) with empty goal state [88], the assuption "$r(z' \,|\, z, a) \leqslant 0$" can be relaxed without any risk of error. In the case of POSMPGs, we relax the strong assumption "$r(z' \,|\, z, a) < 0$ until a goal state is reached" [85].

we assume that this scenario is repeated at infinite horizon. However, for all $z \in Z_{\text{goals}}$ and $z' \in Z$, the player 1's reward is $r(z'|z,a) = 0$ and for a fixed observation $o_{\text{reach}} \in O$, the transition probability is $T(z', o|z, a) = \begin{cases} 1 & \text{if } o = o_{\text{reach}} \text{ and } z' = z \\ 0 & \text{otherwise} \end{cases}$. This means that once a goal state has been reached: (1) the system remains in that state thenceforth and (2) no player no longer receive any reward.

The sequences $\mathsf{out}_1 = \left(r^{[t]}\right)_{t=1}^{\infty}$, $\mathsf{out}_2 = \left(\mu^{[t]}\right)_{t=1}^{\infty}$, $\theta = \left(z^{[t]}, a^{[t]}, o^{[t]}\right)_{t=1}^{\infty}$, $\theta_1 = \left(a_1^{[t]}, o^{[t]}\right)_{t=1}^{\infty}$ and $\theta_2 = \left(z^{[t]}, a_2^{[t]}\right)_{t=1}^{\infty}$ are random variables respectively equal to the players 1 and 2's outputs, the ***paths*** of the play and the players 1 and 2's ***paths***.

For all period $n \geqslant 2$, the prefixes $\theta^{[n]} = \left(\left(z^{[t]}, a^{[t]}, o^{[t]}\right)_{t=1}^{n-1}, z^{[n]}\right)$, $\theta_1^{[n]} = \left(a_1^{[t]}, o^{[t]}\right)_{t=1}^{n-1}$ and $\theta_2^{[n]} = \left(\left(z^{[t]}, a^{[t]}\right)_{t=1}^{n-1}, z^{[n]}\right)$ are random variables representing the ***history*** of the play, and the ***histories*** available to players 1 and 2 respectively at period n. At period $n = 1$, the history of player 1 is "no observation", that we denote $\theta_1^{[1]} = \phi$, and player 2's history is the state of the system, i.e., $\theta_2^{[1]} = z^{[1]}$. For all possible history $\hbar$ of the play or of a player at period $n \geqslant 1$ and all period $t \in \{1, \ldots, n\}$, wherever it exists, we use the notations $z^{[t,\hbar]}$, $a_i^{[t,\hbar]}$, $a^{[t,\hbar]}$, $o^{[t,\hbar]}$, $\theta_i^{[t,\hbar]}$ and $\theta^{[t,\hbar]}$ respectively to refer to the state, the player i action, the action profile, the player 1's observation, the player i history and the play history at period t if the history at period n is $\hbar$. The set of possible histories (respectively possible histories for players 1 and 2) at period n is denoted H^n (respectively H_1^n, H_2^n) while the overall set of possible histories is $H = \bigcup_{t=1}^{\infty} H^t$ (respectively $H_1 = \bigcup_{t=1}^{\infty} H_1^t$, $H_2 = \bigcup_{t=1}^{\infty} H_2^t$).

3.5.2 Minimum Threat and Shortest Path Games

The aim of each player is to realize the optimum output. However, the comparison criterion, named the ***utility***, widely depends on the situation. We discuss two criteria.

When the stage rewards of each player represent an involver, the utility is a particular stage reward of the process. For example, in the context of malware propagation ordered by player 2 trying to take control of the largest number of devices in a network, the reward of player 2 can be seen as the current number of devices under the control of the malware. In this context, the utility of the attacker's output, out_2, is the overall maximum level $\mathfrak{u}_2 = \max_{t \geqslant 1} \mu^{[t]}$ of the threat and the corresponding defender reward $\mathfrak{u}_1 = \min_{t \geqslant 1} r^{[t]}$ is the utility of the defender. That is, the utility functions are the applications

$$\mathfrak{u}_1 \colon (x_t)_{t=1}^{\infty} \longmapsto \inf_{t=1,\ldots,\infty} x_t \qquad \text{and} \qquad \mathfrak{u}_2 \colon (x_t)_{t=1}^{\infty} \longmapsto \sup_{t=1,\ldots,\infty} x_t.$$

Note that $\mu^{[t]} \geqslant 0$ and $r^{[t]} \leqslant 0$. We refer to such game as a ***partially observable stochastic minimum threat path game (POSMPG)***, and to $\mathfrak{u}_1$ and $\mathfrak{u}_2$ as ***extremum-utility*** functions.

When the rewards are additive, the rewards of each player are summed over the periods of the game. However, the value (i.e., the degree of importance) of a reward may decrease (or remain unchanged) over the time. To capture this consideration, we assume that some $\lambda \in (0,1]$, it is the same satisfaction for each player to be rewarded a quantity a at a given period or to be rewarded λa at the next period. So, for two consecutive periods at which he received rewards a and b respectively, the utility as it should be estimated at the former one is $a + \lambda b$. Therefore, the utility function is the application

$$\mathsf{u}_1 = \mathsf{u}_2 = \mathsf{u} \colon \ (x_t)_{t=1}^{\infty} \longmapsto \sum_{t=1}^{\infty} \lambda^{t-1} x_t,$$

with $\lambda \in (0,1]$, and the game is called ***partially observable stochastic shortest path game (POSSPG)***. If $0 < \lambda < 1$, the utility common function u is referred to as the ***λ-discounted sum*** of the rewards, and λ is the ***discount factor***. If $\lambda = 1$ and the sequence $\left(\sum_{t=1}^{n} x_t \right)_{n=1}^{\infty}$ converges, u is the sum of the rewards. The sum and the discounted sum are standard in the study MDPs and SGs.

3.5.3 Strategies

While studying stochastic games, we use the phrases ***stage strategy*** or ***decision rule*** to name a strategy in the stage game, while the phrase ***(behavioral) strategy*** refers to a strategy in the repeated game. Practically, at each period t, player 2 observes the state $z^{[t]}$ of the system and takes a decision $\pi_2^{[t]}\left(\cdot \middle| z^{[t]}\right) \in \Delta(A_2)$ that consists in a probability distribution over the possible actions; player 1 takes a decision (which is also a decision rule) $\pi_1^{[t]} \in \Delta(A_1)$. The set of possible stage strategies for player i is denoted Π_i. Note that $\Pi_1 = \Delta(A_1)$ and $\Pi_2 = \Delta(A_2)^Z$.

Players' behaviors can be smarter described. Each player i makes a decision regarding his history, i.e., his strategy is an application $\sigma_i : H_i \longrightarrow A_i$. The set of possible strategies for player i is Σ_i.

The behavioral strategy of any player may consist of a repetition, at all periods, of the same step strategy. As a result, a stage strategy for player 1 can be viewed as a constant behavioral strategy; a stage strategy for player 2 can be viewed as a behavioral strategy that takes into account only the current state. So, behavioral strategy generalizes stage strategy.

3.5.4 Nash Equilibrium

Considering initial system state z, the expected utilities of players 1 and 2 associated with the strategy profile $\sigma = (\sigma_1, \sigma_2)$ are respectively $U_\sigma(z) = \mathbb{E}_\sigma^z(\mathfrak{u}_1) = \mathbb{E}_\sigma\left(\mathfrak{u}_1 \middle| z^{[1]} = z\right)$ and $U_{2\,|\,\sigma}(z) = \mathbb{E}_\sigma^z(\mathfrak{u}_2) = \mathbb{E}_\sigma\left(\mathfrak{u}_2 \middle| z^{[1]} = z\right)$. Clearly, these utilities are opposite numbers, so U_σ

stands for the utility of strategy profile σ. Note that $U_\sigma(z) = \lim_{n\mapsto\infty} U_\sigma^{[n]}(z)$, i.e.,

$$U_\sigma(z) = \mathbb{E}_\sigma^z\left(\min_{t\geqslant 1} r^{[t]}\right) \tag{3.6a}$$

for the POSMPG, or

$$U_\sigma^{[n]}(z) = \mathbb{E}_\sigma^z\left(\sum_{t=1}^{\infty}\lambda^{t-1}r^{[t]}\right) \tag{3.6b}$$

for the POSSPG, where

$$U_\sigma^{[n]}(z) = \mathbb{E}_\sigma^z\left(\min_{t=1,\dots,n} r^{[t]}\right) \tag{3.7a}$$

for the POSMPG, or

$$U_\sigma^{[n]}(z) = \mathbb{E}_\sigma^z\left(\sum_{t=1}^{n}\lambda^{t-1}r^{[t]}\right) \tag{3.7b}$$

for the POSSPG. This definition of the utility function broadens to the belief of the defender over the network state as $U_\sigma^{[n]}(b) = \sum_{z\in Z} b(z)\, U_\sigma^{[n]}(z) = \left\langle b, U_\sigma^{[n]}\right\rangle$ and $U_\sigma(b) = \sum_{z\in Z} b(z)\, U_\sigma(z) = \langle b, U_\sigma\rangle$, where $\langle\cdot,\cdot\rangle$ is the scalar product.

The utility function is linear in the belief, i.e., $U_\sigma \in \mathrm{lin}_{\Delta(Z)}$, where lin_X is the set of linear functions over a linear space X.

The sets of actions A and states Z are finite. Then, for all action profile a and all system state z, $\mathcal{R}_{min} \leqslant \mathcal{R}(z,a) \leqslant \mathcal{R}_{max}$ where $\mathcal{R}_{min} = \min_{(z,a)\in Z\times A} \mathcal{R}(z,a)$ and $\mathcal{R}_{max} = \max_{(z,a)\in Z\times A} \mathcal{R}(z,a)$. Thus, the defender utility is also bounded, i.e., for strategy profile σ and all belief b:

$$\sum_{t=1}^{\infty}\gamma^{t-1}\mathcal{R}_{min} \leqslant U_\sigma\left(b^0\right) \leqslant \sum_{t=1}^{\infty}\gamma^{t-1}\mathcal{R}_{max}.$$

Hence, we obtain the following lemma that summarizes important results in order to build a resolution method.

Lemma 3.5.1. *The utility function U_σ is linear in the initial belief b^0 and is bounded between $U_{min} = \dfrac{\mathcal{R}_{min}}{1-\gamma}$ and $U_{max} = \dfrac{\mathcal{R}_{max}}{1-\gamma}$ for strategy profile σ and all belief b.*

If player 1 decides to play a strategy σ_1, player 2 will play a strategy σ_2 such that the expected utility is minimal. Therefore, this minimal expected utility,

$$\mathrm{val}_{\sigma_1}(b) = \min_{\sigma_2\in\Sigma_2} U_{(\sigma_1,\sigma_2)}(b) \tag{3.8}$$

is referred to as the ***value*** of the strategy σ_1 under the initial belief b, and val^{σ_1} is the ***value function*** of the strategy σ_1. So, the optimal utility, termed ***optimal value function***, for player 1 (and for the game) is

$$V^*(b) = \max_{\sigma_1\in\Sigma_1} \mathrm{val}_{\sigma_1}(b) = \max_{\sigma_1\in\Sigma_1}\min_{\sigma_2\in\Sigma_2} U_{(\sigma_1,\sigma_2)}(b) = \min_{\sigma_2\in\Sigma_2}\max_{\sigma_1\in\Sigma_1} U_{(\sigma_1,\sigma_2)}(b)\,. \tag{3.9}$$

The objective is to find out a Nash equilibrium (NE) strategy profile, i.e., a player 1's strategy σ_1 that maximizes val_{σ_1} and a player 2's best response to it. An NE is any strategy profile σ such that $U_\sigma = V^*$.

3.5.5 Belief update

Player 1 with belief b at period t at which he took action a_1 and received observation o, knowing that player 2 ruled stage strategy π_2 will update his belief it at period $t+1$ to the belief $\tau\left(b \,|\, \pi_2, a_1, o\right)$ that satisfies the following proposition:

Proposition 3.5.1 (belief Update)**.** *At any period t, if player 1 has chosen an action a_1 under a belief b, if, also, player 2 has played a strategy π_2 and afterwards the defender gets an observation vector o, then the component of his updated belief at any state z' is given by:*

$$\tau\left(b|a_1, \pi_2, o\right)\left(z'\right) = \frac{1}{\mathbb{P}^b_{\pi_2}\left(o \,|\, a_1\right)} \sum_{(a_2, z) \in A_2 \times Z} T\left(z', o \,|\, z,\, a_1, a_2\right) b\left(z\right) \pi_2\left(a_2 \,|\, z\right), \tag{3.10}$$

where, given that player had the belief b upon the system state and know that player 2 ruled stage strategy π_2,

$$\mathbb{P}^b_{\pi_2}\left(o \,|\, a_1\right) = \sum_{z' \in Z} \sum_{(a_2, z) \in A_2 \times Z} T\left(z', o \,|\, z,\, a_1, a_2\right) b\left(z\right) \pi_2\left(a_2 \,|\, z\right) \tag{3.11}$$

is the probability (i.e., player 1's belief) that the system generates observation o when player 1 played action a_1.

Proof. For all $z' \in Z$:

$$\begin{aligned}
b'\left(z'\right) &= \mathbb{P}^b_{\pi_2}\left(z^t = z' \,\middle|\, o^{t-1} = o,\, a_1^{t-1} = a_1\right) \\
&= \frac{1}{\mathbb{P}^b_{\pi_2}\left(o^{t-1} = o,\, a_1^{t-1} = a_1\right)} \mathbb{P}^b_{\pi_2}\left(z^t = z',\ o^{t-1} = o,\, a_1^{t-1} = a_1\right) \\
&= \frac{\displaystyle\sum_{(a_2, z) \in A_2 \times Z} \mathbb{P}^b_{\pi_2}\left(z^t = z',\ o^{t-1} = o,\, z^{t-1} = z,\, a_1^{t-1} = a_1,\, a_2^{t-1} = a_2\right)}{\mathbb{P}^b_{\pi_2}\left(o^{t-1} = o,\, a_1^{t-1} = a_1\right)}.
\end{aligned}$$

On the one hand:

$$\begin{aligned}
&\mathbb{P}^b_{\pi_2}\left(z^t = z',\ o^{t-1} = o,\, z^{t-1} = z,\, a_1^{t-1} = a_1,\, a_2^{t-1} = a_2\right) \\
&\qquad = \mathbb{P}^b_{\pi_2}\left(z^t = z',\, o^{t-1} = o \,\middle|\, z^{t-1} = z,\, a_1^{t-1} = a_1,\, a_2^{t-1} = a_2\right) \times \\
&\qquad\qquad \times \mathbb{P}^b_{\pi_2}\left(a_1^{t-1} = a_1\right) \mathbb{P}^b_{\pi_2}\left(z^{t-1} = z\right) \mathbb{P}^b_{\pi_2}\left(a_2^{t-1} = a_2 \,\middle|\, z^{t-1} = z\right),
\end{aligned}$$

and clearly

$$\mathbb{P}^b_{\pi_2}\left(z^t = z',\, o^{t-1} = o \,\middle|\, z^{t-1} = z,\, a_1^{t-1} = a_1,\, a_2^{t-1} = a_2\right) = T\left(z', o \,|\, z,\, a_1, a_2\right).$$

On the other hand: $\mathbb{P}^b_{\pi_2}\left(o^{t-1}=o,\, a_1^{t-1}=a_1\right)=\mathbb{P}^b_{\pi_2}\left(o^{t-1}=o,\,\middle|\, a_1^{t-1}=a_1\right)\mathbb{P}^b_{\pi_2}\left(a_1^{t-1}=a_1\right)$. Then:

$$b'\left(z'\right)=\frac{\sum\limits_{(a_2,z)\in A_2\times Z} T\left(z',o\,|\,z,\,a_1,a_2\right)b\left(z\right)\pi_2\left(a_2\,|\,z\right)}{\mathbb{P}^b_{\pi_2}\left(o^{t-1}=o,\,\middle|\, a_1^{t-1}=a_1\right)}.$$

Now as $\sum\limits_{z'\in Z} b'\left(z'\right)=1$, we obtain:

$$\mathbb{P}^b_{\pi_2}\left(o^{t-1}=o,\,\middle|\, a_1^{t-1}=a_1\right)=\sum_{(a_2,z)\in A_2\times Z} T\left(z',o\,|\,z,\,a_1,a_2\right)b\left(z\right)\pi_2\left(a_2\,|\,z\right). \tag{3.12}$$

□

Note from equation (3.12) that the probability $\mathbb{P}^b_{\pi_2}\left(o\,|\,a_1\right)$ defined in the previous proposition is a linear function in $b\in\Delta\left(Z\right)$.

Player 1 also upgrades his belief over the history of the play as follows:

Proposition 3.5.2 (belief over a history)**.** *For any period $n\geqslant 2$ and any game history $\hbar$, the player 1's belief over the history $\hbar$ over the commonly known strategy σ satisfies:*

$$b^0\left(\hbar\right)=\mathbb{P}^{b^0}_{\sigma}\left(\theta^{[n]}=\hbar\right)=b^0\left(z^{[1,\hbar]}\right)\prod_{t=2}^{n}T\left(z^{[t,\hbar]},o^{[t-1,\hbar]}\,\middle|\, z^{[t-1,\hbar]},a^{[t-1,\hbar]}\right)\sigma\left(a^{[t-1,\hbar]}\,\middle|\,\theta^{[t-1,\hbar]}\right).$$

Proof (by induction). If $n=2$,

$$\begin{aligned}
\mathbb{P}^{b^0}_{\sigma}\left(\theta^{[n]}=\hbar\right) &= \mathbb{P}^{b^0}_{\sigma}\left(z^{[1]}=z^{[1,\hbar]},a^{[1]}=a^{[1,\hbar]},o^{[1]}=o^{[1,\hbar]},z^{[2]}=z^{[2,\hbar]}\right)\\
&= \mathbb{P}^{b^0}_{\sigma}\left(z^{[2]}=z^{[2,\hbar]},o^{[1]}=o^{[1,\hbar]}\,\middle|\, z^{[1]}=z^{[1,\hbar]},a^{[1]}=a^{[1,\hbar]}\right)\times\\
&\quad\times\mathbb{P}^{b^0}_{\sigma}\left(a^{[1]}=a^{[1,\hbar]}\,\middle|\, z^{[1]}=z^{[1,\hbar]}\right)\mathbb{P}^{b^0}_{\sigma}\left(z^{[1]}=z^{[1,\hbar]}\right)\\
&= b^0\left(z^{[1,\hbar]}\right)T\left(z^{[2,\hbar]},o^{[1,\hbar]}\,\middle|\, z^{[1,\hbar]},a^{[1,\hbar]}\right)\sigma\left(a^{[1,\hbar]}\,\middle|\,\theta^{[1,\hbar]}\right).
\end{aligned}$$

The equality holds.

Suppose it holds also for some $n\geqslant 2$, then take any period $n+1$ history $\hbar$ consider the period 2 history $\hbar'=\theta^{[n,\hbar]}$. We get:

$$\begin{aligned}
\mathbb{P}^{b^0}_{\sigma}\left(\theta^{[n]}=\hbar\right) &= \mathbb{P}^{b^0}_{\sigma}\left(\theta^{[n]}=\theta^{[n,\hbar']},a^{[n]}=a^{[1,\hbar]},o^{[n]}=o^{[n,\hbar]},z^{[n+1]}=z^{[n+1,\hbar]}\right)\\
&= \mathbb{P}^{b^0}_{\sigma}\left(z^{[n+1]}=z^{[n+1,\hbar]},o^{[n]}=o^{[n,\hbar]}\,\middle|\,\theta^{[n]}=\theta^{[n,\hbar]},a^{[n]}=a^{[n,\hbar]}\right)\times\\
&\quad\times\mathbb{P}^{b^0}_{\sigma}\left(a^{[n]}=a^{[1,\hbar]}\,\middle|\,\theta^{[n]}=\theta^{[n,\hbar']}\right)\mathbb{P}^{b^0}_{\sigma}\left(\theta^{[n]}=\theta^{[n,\hbar']}\right),
\end{aligned}$$

and result falls from the hypothesis. □

Chapter 4

Cyber Security and Cyber Deception

Contents

4.1 Introduction

As mentioned in [81], the "Common Vulnerabilities and Exposure (***CVE***) published 16,555 vulnerabilities between January 1, 2018, and December 31, 2018" Cyber threats are a big deal. Cyber attacks can cause electrical blackouts, failure of military equipment, and breaches of national security secrets. They can result in the theft of valuable, sensitive data like medical records. They can disrupt phone and computer networks or paralyze systems, making data unavailable. The most important consequences encompass the financial looses, among which the often underestimated Below-the-surface costs and the fines, the theft, and, importantly, the reputation damage.

4.2 The Basics of Cyber Security

4.2.1 Definition

Cyber Security

"Cyber security is the application of technologies, processes, and controls to protect systems, networks, programs, devices and data from cyber attacks. It aims to reduce the risk of cyber attacks and protect against the unauthorized exploitation of systems, networks, and technologies." [1] Cyber security includes:

- *Critical infrastructure cyber security.* Security of physical and cyber systems and assets that are so critical that their incapacity or destruction would have a debilitating impact on physical or economic security or public health or safety.
- *Network security.* Address the vulnerabilities affecting operating systems and network architecture, including servers and hosts, firewalls and wireless access points, and network protocols.
- *Cloud computing security.* Secure data, applications, and infrastructure in the Cloud.
- *IoT (Internet of Things) security.* IoT is "an open and comprehensive network of intelligent objects that have the capacity to auto-organize, share information, data and resources, reacting and acting in face of situations and changes in the environment". [56] The aim of IoT security is to ensure the availability, integrity, and confidentiality of IoT solutions.
- *Application security.* Address the vulnerabilities resulting from insecure development processes in designing, coding, and publishing software or a website.

Each cyber security activity should aim at the following goals, referred to as the CIA security triad :

1. *Confidentiality*: only authorized persons can read the information.
2. *Integrity*: only authorized persons can modify the information.
3. *Availability*: only authorized persons can denial the use of the information.

Cyber Attack

Under the following definition, as "any kind of malicious activity that attempts to collect, disrupt, deny, degrade, or destroy information system resources or the information itself", [2] cyber attacks history dates back to the seventies, before most people even had a computer. These activities encompasses for example:

[1] `https://www.itgovernance.co.uk/what-is-cybersecurity`
[2] See `https://csrc.nist.gov/glossary/term/cyber_attack`.

- Ransomware or the theft of intellectual property,
- Attacks on the cloud,
- The use of viruses and malware to illegally access emails, among other purposes,
- Hacking into military computers,
- The phone phreaking,

For the purpose of this thesis however, we consider the following definition of a ***cyber attack***: "an attack, via cyberspace, targeting an enterprise's use of cyberspace for the purpose of disrupting, disabling, destroying, or maliciously controlling a computing environment/infrastructure; or destroying the integrity of the data or stealing controlled information"[2] [45].

Cyber attacks may be perpetrated by isolated individuals as by very structured organizations as well, including criminal organizations and sovereign states. The hackers are so organized that they form a large community that benefits from their criminal activities. Authors in [37] distinguish six types of hackers organizations:

1. ***Aggregates***, loosely organized groups of hackers engaged only in temporary collaboration;
2. ***Swarms***, constituted of hackers who collaborate without a real chain of command;
3. ***Hubs***, that consists in core groups of hackers working with associates;
4. ***Clustered hybrids***, hubs that combine online and offline activities and that focus on specific activities or methods;
5. ***Extended hybrids***, clustered hybrids with many associates and subgroups;
6. ***Hyrarchies***, tradition organizations and criminal groups that advantage of online technology to facilitate activities.

The activities of an organized group are minimally oriented toward:

1. Operations risk and cost minimization alongside with revenue maximization, which involves a good manage of the distribution process, an adequate selection of valuable targets and strategies to hide from authorities in case necessary;
2. Human resource, i.e., training and recruitment of hackers;
3. Marketing and delivery, that aims for the marketplace, the reputation, the evaluation of the value of vulnerabilities and the money laundering.

4.2.2 The Cyber Attack Process

A cyber attack is a sophisticated process prepared step by step by the attacker, that ranges from simple reconnaissance activities to a harm upon a victim. The process can be analyzed in three stages:

The Survey

The prerequisite for any cyber attack is the existence of at least one ***vulnerability*** of the system, i.e., a "weakness in the information system, system security procedures, internal controls, or implementation that could be exploited or triggered by a threat source". [3] Vulnerabilities include but are not limited to:

- Misconfiguration of application security tools,
- Unsecured application programming interfaces (***APIs***),
- outdated or unpatched software, while software updates containing valuable and important security measures is released by vendor,
- Zero-day vulnerabilities, that is unknown to the enterprise and software vendor,
- Weak or stolen user credentials, when the user fails to create a unique and strong password,
- Access control or unauthorized access, when for example members of a compagny are granted more access than needed to perform their job functions.

When a vulnerability meets the requirements of operations risk, cost minimization and revenue maximization, it is worth nothing to mention that cyber attackers will exploit it one day or another. Indeed, even user errors can reveal information that can be used in attacks. More inclusively, attackers use any mean to collect and assess any information about your organisation's computers, security systems and personnel. The means they use include:

- Open source information such as LinkedIn and Facebook, domain name management and search services, and social media;
- Commodity toolkits and techniques, and standard network scanning tools. [4]

Once they have discovered the vulnerabilities, the hackers develop a program to exploit it and force the system to behave in an unintended way.

[3] See `https://csrc.nist.gov/glossary/term/vulnerability`.
[4] See `https://www.ncsc.gov.uk/information/how-cyber-attacks-work`.

The Delivery

After the survey stage, the attackers look to get into a position where they can deliver the exploitative program they have developed. From this position, they aim at exploiting the vulnerabilities and gain some form of unauthorized access. One of the following mechanisms support this delivery:

Physical infection The infection may be performed through the connection of a physical medium, like the universal serial bus (***USB***) drives or port, to a targeted or intermediate host.

Direct delivery The program may be downloaded through digital channels like SMS, and emails.

Via Download While navigating on the web, a user may be driven to a compromised website, from which he will be redirected to a specifically designed page. From this page, a downloader will be installed on the machine of the user, that will establish a contact with the Command and Control Server.

Software Distribution "An insider in any organization can pose threat if compromised. There are cases when an employee was bribed to obtain privileged network information" [83]. This just goes to show at what extent an organization is susceptible to software distribution. Technically, a malicious code will be attached to the software of a target device in the enterprise. Once the adulterated software runs, the malicious code is executed to download the exploitative program.

The Attack Itself

At this stage, the attackers have injected an exploitative program that allow them to achieve their goal while they remain undetected. The goal may be for example to illegitimately retrieve information, to make changing for their own benefit, or even to delete the operating system. The list is not exhaustive. The attack may also be a multi-stage, i.e., attacking the victim machine is actually an open door for subsequent attacks. This second type includes: identifying other vulnerabilities to gain privileged access; controlling many devices to overwhelm a target device and thereby escape control of a prohibited activity.

4.2.3 Classification of Cyber Attacks

Authors in [90] listed 5 classification modes of cyber attacks:

1. Legal Classification (see figure 4.1):

(a) *Cyber War*: War conducted in and from computers and the networks connecting them, waged by states or their proxies against other states. The first cyber war probably took place in Estonia, from late April to mid-May 2007, during which a series of cyber attacks were launched against government, media, banking and political party websites [40].

(b) *Cyber Terrorism*: The convergence of cyberspace and terrorism, unlawful attacks and threats of attacks against computers, networks and the information stored therein when done to intimidate or coerce a government or its people in furtherance of political or social objectives [97].

(c) *Cyber Espionage*: Use of computer networks to gain unauthorized access to confidential information. Cyber espionage activities can be planned and carried out by individuals, organizations or nations to gain advantage through illicit access to confidential information. They can cause a number of damages including information leakage, financial and privacy losses, deterioration of system functionality, destruction of facilities, legal confrontations, even wars and others [70].

(d) *Cyber Crime*: There are many definitions of the cyber crime, among which: Offenses against the confidentiality, integrity, and availability of computer data and system [63]. Cyber crime includes: [5]

- Offences against the confidentiality, integrity and availability of computer data and systems:
  - Illegal access, that results from:
    - Inadequate and incomplete protection of computer systems
    - Development of software tools to automate attacks
    - The growing role of private computers as a target of hacking attacks
  - Illegal data acquisition (illegal interception, data interference, system interference)
- Content-related offenses, including but not limited to:
  - Child pornography
  - Erotic or pornographic material
  - Racism, hate speech, glorification of violence
  - Religious offenses
  - Illegal gambling and online games
  - Libel and false information
  - Spam and related threats
- Copyright- and trademark-related offenses
- Computer-related offenses:
  - Fraud and computer-related fraud (the most common offenses include online auction fraud and advance fee fraud)

[5] See `https://www.itu.int/ITU-D/cyb/cybersecurity/legislation.html`.

- Computer-related forgery
- Identity theft. The most relevant data are social security and passport numbers, date of birth, address and phone numbers, and passwords
- Misuse of devices

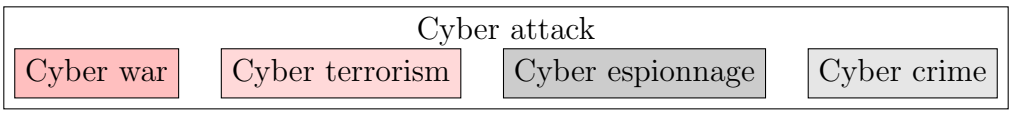


Figure 4.1: Legal classification of cyber attacks

2. Classification Based on the Purpose (see figure 4.2):

   (a) *Reconnaissance Attack*. Attempt to gain information about an organization's systems and networks without the explicit permission of the organization. Some common examples of reconnaissance attacks include packet sniffing, ping sweeps, port scanning, phishing, social engineering, and internet information queries. The reconnaissance phase takes place in two stages:

   - Pre-exploitation reconnaissance.
     - Gathering information about the target infrastructure on the target systems (active or passive reconnaissance)
     - Vulnerability discovery (through enumeration of specific details about a particular system)
     - Gathering information about the human targets selected for the initial compromise phase
   - Post-exploitation reconnaissance, that takes place after an initial foothold on a target system has been established and further information has to be collected in order to discover valuable assets by moving laterally within the target network

   (b) *Access Attack*. The attacker illegally procures ingress to a machine with the intent to manipulate information [42]. Access attacks include:

   - *Password attack*, i.e., attempt to access a file, folder, account, or computer secured with a password. Authors in [69] count 7 types of password attack:
     - *Brute force attack*, that consists in making numerous hit-or-miss attempts to gain access
     - *Dictionary attack*. Attempt to login to accounts by trying all possible passwords, until they find the correct one [92].
     - *Shoulder Surfing*. "The attacker observes the user, how he enters the password, i.e., what keys of keyboard the user has pressed"

- *Replay attack*. The attacker sends a message packet that been received by the target host to spoof the system. The basic principle is to send the previously eavesdropped data to the receiver without doing any change [51, 103].
- *Phishing attack*. Social engineering technique to steal victims' sensitive data, such as login credentials, personal details, and credit card numbers. [6]
- *Key logger attack*. A spyware, the keylogger, records the user's activity by logging keyboard strokes.
- *Video recording attack*. Using a computer vision algorithm, the attacker tracks fingertip motions from a video taken with a smartphone. It then matches the locations of the tracked fingertips to a few candidate patterns to test on the target device [100].

- *Trust exploitation attack*. An individual takes advantage of a trust relationship within a network. In social networks for example, "even if you have a high security setting, if you share your content with friends, and one of your friends is hacked, your data will be exposed to the attacker. Thus, the strength of your security level is as weak as that of your friend with the lowest level of security." [98]
- *Port redirection attack*. Use of a compromised host to gain access through a firewall that would otherwise be blocked. This attack is based on trust exploitation.
- *Man-in-the middle attack*. The attacker makes independent connection with victims and relays messages between them making them believe that they are in contact privately.

(c) *Denial of service (***DoS***) Attack*. The attacker seeks to make a machine or network resource unavailable to its intended users by disrupting services of a host connected to a network. Authors in [68] group DoS attacks and representatives regarding many criteria.

- Regarding the vulnerability exploited:
  - *Bug exploitation attack*. Exploitation of bugs on a technical equipment or software:
    - *Attack from inside*. The attacker exploits a bug in some system to get some kind of system control and then uses that control to affect the desired service
    - *Attack from outside*. The bug directly affects the targeted service
  - *Resource depletion attack*. The attacker sends numerous queries to the system, knowing that the system will allocate relevant resources to proceed each query. The system thereby stops working properly.

[6] `https://easydmarc.com/blog/12-types-of-phishing-attacks-and-how-to-identify-them/`.

        * Classification based on the type of resource exhausted:
            · *Memory depletion attack.* The attacker creates a situation where all available memory is allocated for queries and there is not enough memory for the new ones.
            · *Central processing unit (CPU) work depletion attack.* The attacker supplies incoming data which requires even more CPU work to analyze or process the query, then CPU gives excessive attention to this one job and is unable to do other jobs.
        * Classification based on the necessity to modify the packet:
            · *Semantic resource depletion attack*, when systems use more resources if incoming packets are modified.
            · *Brute-force resource depletion attack.* The number of incoming queries is sufficiently large the service will be denied.
    - *Bandwidth exhaustion attack.* The attacker exhausts the bandwidth by sending a huge data stream.
- Regarding the attacker size:
    - *(Single source) DoS attack.* The attack is launched from a single machine.
    - *Distributed denial of service (***DDoS***) attack.* Before launching the attack, the attacker compromises numerous devices, making them zombie machines or ***bots***. The recruitment process ends with a network of machines under the orders of the attacker, the ***botnet***, which constitutes the army whose actions will be coordinated to deny the service. There are four categories of army formation:
        * *Manual agent army formation.* The attacker finds vulnerable machines and installs malicious code into them himself.
        * *Semi-automatic agent army formation.*
        * *Automatic agent army formation.* The formation equires human action just to launch the tool which will do all necessary tasks to form the agent army.
        * *Takeover of an existing agent army.* "There are agent armies which can be borrowed, rented or even stolen."

3. Classification Based on the Scope:
    (a) Malicious Large Scale
    (b) Non-Malicious Small Scale
4. Classification Based on the Severity of Involvement:
    (a) *Passive Attacks.* An eavesdropper attempts to obtain useful information by spying on transmissions, without altering the information.

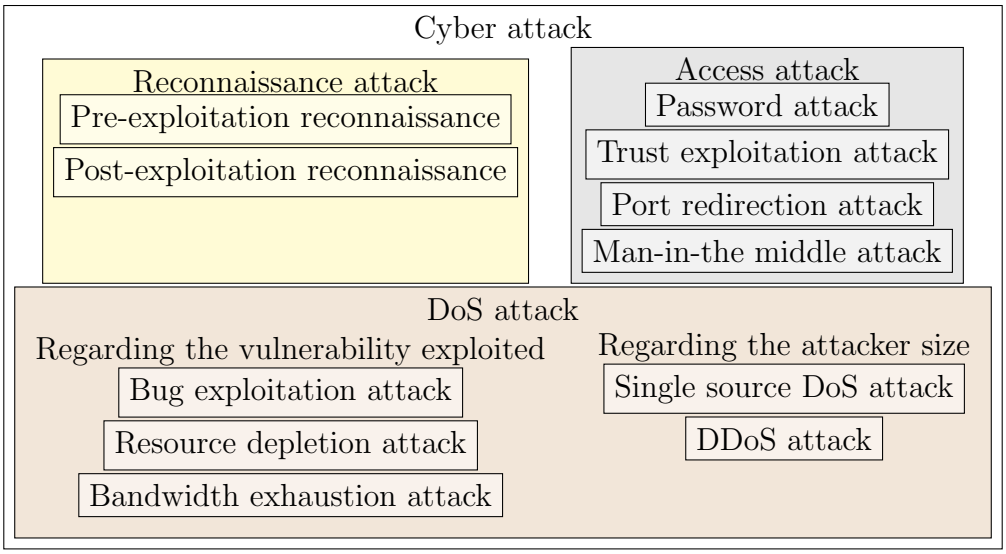


Figure 4.2: Classification of cyber attacks based on the purpose

- *Wiretap attack*. The eavesdropper tries to resolve the source message by wiretapping transmissions.
- *Traffic Analysis Attack*. The eavesdropper aims to extract additional information, such as the identity and location of communicators, by observing and analyzing traffic patterns. [54]

(b) *Active Attacks*. Hostile activities of an enemy to destroy or disturb normal communications, such as masquerading as an authorized entity to access the network, exhausting network resources by forwarding outdated packets, falsifying the content of packets, or injecting polluted packets into networks, etc.

5. Classification Based on the Network Type:

(a) Attacks in a Mobile Ad-Hoc Network (MANET). The common attacks can be grouped regarding the protocol layer [59]:

- Physical layer: eavesdropping, jamming, interception
- Data link layer: monitoring, traffic analysis
- Network layer: Black hole, Wormhole, Gray hole, Byzantine, message tempering, resource consumption, Flooding, location disclosure attacks
- Transport layer: SYN Flooding, Session hijacking
- multiple layers: man-in-the-middle attack, DoS

(b) Attacks in a Wireless Sensor Network (WSN). Authors in [77] enumerate the attacks on each protocol layer:

- Physical layer: Jamming, Tampering
- Data link layer: Collision, Exhaustion

- Network layer: Route Inform Manipulating, Selective, Forwarding, Sybil Attack, Sinkhole, Wormhole, Hello flood
- Transport layer: Flooding
- Application layer: Clone attack

4.2.4 The Pillars of Cyber Security

To be truly effective, it must take into account the strengths and weaknesses of the adversary and offensive techniques, and prevent them upstream while countering them downstream. Preventive measures include educating administrators and users, both on current regulations and on the urgency of synergy in cyberspace, which is in fact a global village. These measures must also concern the constitution of each network, from the choice of equipment to the data exploitation protocols, and even anticipate intrusions into the system or manipulations of the data. Like security in general, cybersecurity is based on three pillars: protection, detection and response [21].

Protection

The efficiency of a cyber attack relies on it capability to illegitimately exploit or affect systems, networks, programs, devices. The first principle for cyber security is therefore the protection of these assets. Since human work cannot be perfect, the system to be protected will always have vulnerabilities. It is therefore obvious that the best protection inevitably includes a vulnerability management policy. The vulnerability management policy is composed of six phases:

1. Asset inventory. Perhaps it is the forgotten asset that the attacker will target. It might then be unfortunate to be surprised by its vulnerability and value.

2. Information management. Verify protocols for gathering, storing, distributing, and deleting information.

3. Risk assessment. "Prioritize some vulnerabilities over others and allocate resources to mitigate against them." To this end:

   (a) Identify the scope, i.e., the area to cover and the area not to cover,

   (b) Collect data about the existing policies and procedures that are in place to safeguard the networks, applications, and systems that are covered in the scope.

   (c) Verify that policies and procedures comply with user and administrator tasks.

   (d) Evaluate the effectiveness of protective measures.

   (e) Identify and classify threats based on the motivations and capabilities of potential attackers.

(f) Identify the procedures and security mechanisms that are still inadequate and take corrective actions to update and improve them until they are adequate. This means determining the recommended standards that the safeguards must meet. Over time, analyze the acceptable risks and, if necessary, update the security standards, until these risks no longer pose a threat.

Detection

One of the most effective and valuable security mechanisms for detecting malicious network behavior is an Intrusion Detection System (IDS). Authors in [53,75] distinguish four classes of IDS: host-based IDS (detecting suspicious activity based on single host monitoring), network-based IDS (which is based on the detection of anomalies in the network and application protocols), wireless-based IDS (which consists of monitoring the networking protocols in the wireless traffic), network behavior analysis (which consists in ensuring the examination of network traffic to identify threats that generate unusual traffic flows).

4.3 Cyber Deception

4.3.1 The Need of Cyber Deception

As seen in the previous section, cyber security is a sustained confrontation between cyber attackers and cyber defenders, i.e., a warfare in the cyber space. As says the author in [89], "all warfare is based on deception". Following his advice, the strategic cyber defender must "hold out baits to entice" the cyber attacker, "feign disorder, and crush him". Therefore, as well as for military use, psychology, criminology and privacy advocacy, deception is a must to achieve cyber security goals. In deceive, we should understand: "intentionally cause another person to acquire or continue to have a false belief, or to be prevented from acquiring or cease to have a true belief" [66]. However, to fit with the battlefield of cyber security, cyber deception should be "deception in and using cyberspace" [73]. That is, practically, defensive cyber deception is an adjustment of deception techniques in the cyber space for the cyber security purpose. Since the same techniques are also investigated in the offensive side, cyber deception involves a confrontation between rational, intelligent actors, resulting in the need to be associated with game theory.

Knowledge of cyber deception is essential to be able to thwart the subterfuges used by a potential enemy to attack the system. However, the knowledge of cyber tricks for defensive purposes is not limited to this increase in the capacity of anticipation and deduction. Indeed, in terms of protection, it can be used to divert the enemy's attention from the weaknesses he may use. It can also be used to push potential undetected attackers to reveal themselves or at least their true intentions. Finally, in response to an attack, defensive cyber deception can induce losses in the adversary that, if it had been considered before the offensive was launched, would have discouraged the action in progress. This list is obviously not exhaustive.

4.3.2 Some Deception Techniques

The techniques are listed in [66].

Impersonation

Roughly, impersonation consists in operating with the permission of a different user account. It happens in the cyber space when the attacker pretends to be someone else or impersonates a legitimate user (or group of users), to gain access to information they are not authorized to have. Via email, for example, the attacker impersonates a legitimate sender in order to trick the recipient into clicking on a malicious link or attachment.

Delay

Making someone or a software take more time than necessary to accomplish a particular task.

Fakes

Use an artifact to lead the adversary to a wrong conclusion upon the existence or some property of something. For example: fake vulnerability (the subject appears to be vulnerable, while it is not).

Anti-forensic

"Attempt to thwart intelligence gathering by forensic techniques" [73]. Its meyhods include [20]: data fabrication, contraception, hiding or destruction, program packers, encryption, data trial obfuscation and file system attack.

Camouflage

Camouflage is the concealment of the real. The technique can be used for example to:

- Let a patch figure out of the vulnerabilities,
- Fool the attacker on the real version of the operating system or on the operating system itself,
- Secretly move information or keep its existence secret

False Excuses

"A false excuse is a deception disavowing wrongdoing so as to avoid harm to the self." [36]

Social Engineering

Social engineering is the psychological manipulation of people into performing actions or divulging confidential information. The aim of the cyber attacker is to gain access to private personal and financial information for the purpose of financial reward, or for reconnaissance purposes to gather more detailed intelligence on a target organization. The stratagems used for social engineering [4] include:

- Phishing. Promote a sense of seriousness, necessity, strangeness or panic in the targeted person. When the fishing process encompass the selection of specific persons or groups, it is referred to as spear phishing.
- Baiting. The attacker lures the victim into a trap by promising an attractive, hard-to-refuse offer.
- Scareware. The attacker causes shock, anxiety, or the false perception of a threat.
- Pretexting. The attacker begins by developing as a co-worker, police, tax officials who have the authority to know things.

4.3.3 Cyber Deception Taxonomy

The principle

To classify cyber deception techniques used in conflict situations, [66] examines them in four modalities: the general attitude of the deceiver (telling the false or hiding the true), the actors involved, the duration of the process, and the actions taken.

The Duration A cyber deception technique is either static or dynamic.

The Private Information A cryptic deception consists in hiding the truth, while a dynamic deception consists in telling the false.

The Actors A deception is intensive if the deceiver hides the subject by modifying it. If the deceiver uses an outsider actor to hide the subject, the deception is extensive.

The Actions If the deceiver manipulated the data received about the property of the subject, the deception is informational. Otherwise, i.e., if he modifies the properties of the subjet over time, or if he realizes the property from a random variable, the deception is a motive.

Perturbation

Cryptic, intensive and informational cyber deception technique, it consists in introducing a noise in the data to limit the leakage of information.

Obfuscation

Cryptic, extensive and informational cyber deception technique, it consists in using a decoy to make the adversary have false value about the data.

Mixing

"Mixing strategies use exchange systems to prevent direct linkage between systems" [18]. Cryptic, extensive and motive cyber deception technique, it consists in

Moving target

Cryptic, intensive and motive cyber deception technique, it consists in "continuously changing a system's attack surface through adaptation, thereby increasing the uncertainty, complexity, and costs for the attacker" [94].

Honey-X

Mimetic and static cyber deception technique, it consists in making the attacker acquire false data in a single-stage interaction.

Attacker engagement

Mimetic and dynamic cyber deception technique, it consists in making the attacker acquire false data in a multi-stage interaction.

4.4 Conclusion

The cybersecurity context is very much a war context in which enemies can make strategic alliances. Using any means necessary, they flout privacy rules and access sensitive information. To deal with them effectively, administrators must develop cyber deception techniques that the attackers themselves use. The different categories of cyber deception were presented, taking into account the practical context of implementation. In the following chapter, we present a context in which attackers use cyber deception and propose a proportionate response to it.

Chapter 5

Game Theoretic Modeling of Network Epidemics

Contents

5.1 Introduction

The framework considered in this work is a network epidemic problem in which an attacker is trying to compromise computers in a network using a cyber-attack that propagates into the network following an epidemic process. Such an attack might be a threat inside an Internet of Things (IoT) network or any other network of devices, a rumor in a social network, etc.

Worms propagate as an epidemic in a computer network whereas virus (or e-virus) does not necessarily propagate. They can stay in the same computer or advance to a target as a lateral movement [99]. Then our framework deals with worms into a computer network but, by abuse of language, we use the term "virus" as our propagation model is based on an epidemiological framework. Epidemic processes can be modeled using well-known compartmental framework like the Susceptible-Infected-Resistant (SIR) framework [64]. In this framework, each agent can be in one of the three states at any time: vulnerable non-infected (or susceptible), infected, or removed (that are out of the reach of the threat). Unlike in general problems in epidemiology, in our framework, the spread of an attack is controlled by a rational player who is called the attacker. Cyber-attacks have various goals encompassing distributed denial of services (DDoSs) [47], which are malicious [1] attempts to disrupt the normal traffic of a targeted server, service or network by overwhelming a target or its surrounding infrastructure with a flood of traffic. One of the best defense mechanisms against such a distributed attack is to consider decentralized techniques like the uniform defense strategy as suggested in [86]. However, in a network epidemic attack scenario, agents will not rigorously evaluate their strategies and may not even take the best action. To better mitigate the attacker's actions, one should envision the intervention of a rational, intelligent defender. Game theory proposes the best mathematical tools to study the defender's strategy in a conflict against an attacker in a scenario of cyber-attacks over a network. Our framework is based on a zero-sum game between an attacker and a defender in order to control an epidemic process through a network of devices.

Some important features have to be considered when designing the game theoretic framework that models the competitive scenario between intelligent and rational players for controlling epidemic spreads over a network. This is true particularly in terms of information available to each player. First, unlike the attacker, the defender cannot illegally brute force or use any other unauthorized techniques of scanning in order to get the state of each device on the network. So, he may ignore the state of some nodes and should be assumed to have partial observability over the overall network state at each time. Second, even though nodes are not intelligent and rational, they may take some actions based on their state. That is, no player can monitor the state transitions; the game framework is by construction partially observable. Moreover, a cyber-attacker may deploy probing techniques in order to get the state of each agent on the network. Then, the partial observation assumption is assumed to be true only for the defender. Finally, the goal of the defender corresponds to the opposite of the attacker's; and then our game is zero-sum. To summarize, the problem considered in this thesis of controlling an epidemic over a network corresponds to a one-sided partially observable stochastic zero-sum game (OS-POSG) which is a high complexity problem [6].

The resolution of OS-POSGs has been addressed recently in the literature for special cases of information systems [12, 33]. The solution is generally intractable for more than 5 nodes but the tractability is improved using the value iteration (VI) algorithm as proposed

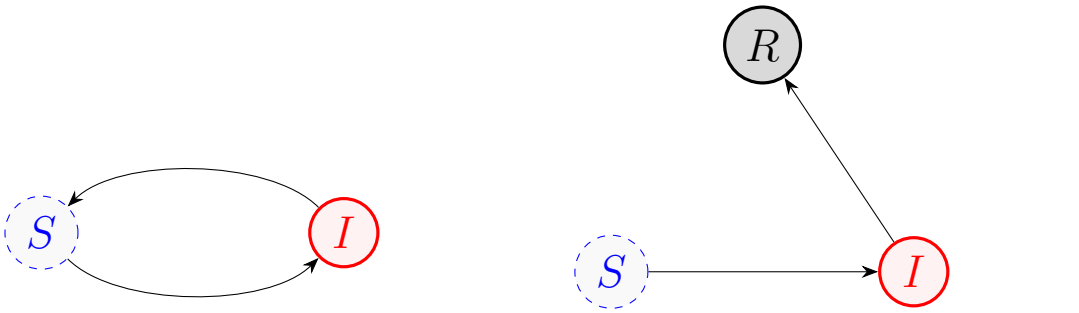
[1] The adjective "malicious" does not determine the attacker, but defines the a priori judgment of the attempt to disrupt the flow of traffic in the network. The same adjective is used to describe the code used by the attacker to control the devices to ensure the success of his endeavor.

in [35]. The proof of the convergence of this algorithm is based on the assumption that the attacker observes all the actions of the defender, i.e., the defender has a partial observability (imperfect information) and cannot infer the attacker's moves (incomplete information), while the attacker has complete and perfect information. This assumption may not hold in our context, where the defender makes use of intrusion detection systems like an IPS that reveals transmissions of the virus between nodes. An IPS is a computer security mechanism set to detect, deflect, or, in some manner, counteract attempts at unauthorized use of information systems. The detection of a transmission implies a cure of the infected nodes and the attacker cannot infer such action from the defender. The fact that the attacker is not aware of defender's action presents an important challenge within our POSG framework compared to the classical one. Thus, our paper aims to answer how to solve this more complicated OS-POSG framework with incomplete information assumption for the player with full observation. To the best of our knowledge, [88] is the first work that tackle this issue.

5.2 A New Mathematical Model for the Controlling of Active Spread of Epidemics

5.2.1 Compartmental Study of Epidemics

From epidemiological perspective, the population is divided into groups, called ***compartments***. The group compartment uses the same label with the state and is supposed not to contain the same individuals throughout the spreading process. For example, in a population subject to malaria [46], there are two compartments, I and S, respectively corresponding to states ***infectious***, of individuals who carry and can propagate the disease, and ***susceptible***, of individuals who do not carry the disease. All susceptible individual may contract the disease and become infectious: we denote such an individual state transition $S \to I$ (see figure 5.1a). When some individual is infected, he may recover and become susceptible again, and



(a) SIS epidemic model
E.g., malaria. The possible individual state transitions are $S \to I$ and $I \to S$.

(b) Basic SIR epidemic model
Possible individual state transitions: $S \to I$ and $I \to R$.

Figure 5.1: Dynamics in SIS and SIR epidemics

the individual state transition is denoted $I \to S$. The epidemic is therefore said to be an

SIS (susceptible-infected-susceptible) one. More generally, ***individual state transition*** from any values X to Y is denoted $X \to Y$, and the name the epidemic model relies on the individual states and the dynamics involved. The dynamics make a difference between the above SIS model and the SI model, in which only the transition $S \to I$ is possible.

The existence of an R (***removed***) compartment of individuals who cannot carry the disease, either because they have been vaccinated or because they are dead, suggests the basic SIR epidemic model (figure 5.1b) introduced in 1927 [84]. For instance, in the case of tuberculosis without vaccination [8], there is no transition between the compartment S and R, while an infected individual who recovers is longer susceptible. Accounting the compartment S, I and R, the statements:

1. No state is isolated, i.e., between two different states A and B, at least one of the transitions $A \to B$ or $B \to A$ is possible
2. The transition $R \to I$ is impossible, from the definition of removed individual
3. The transition $S \to I$ is possible

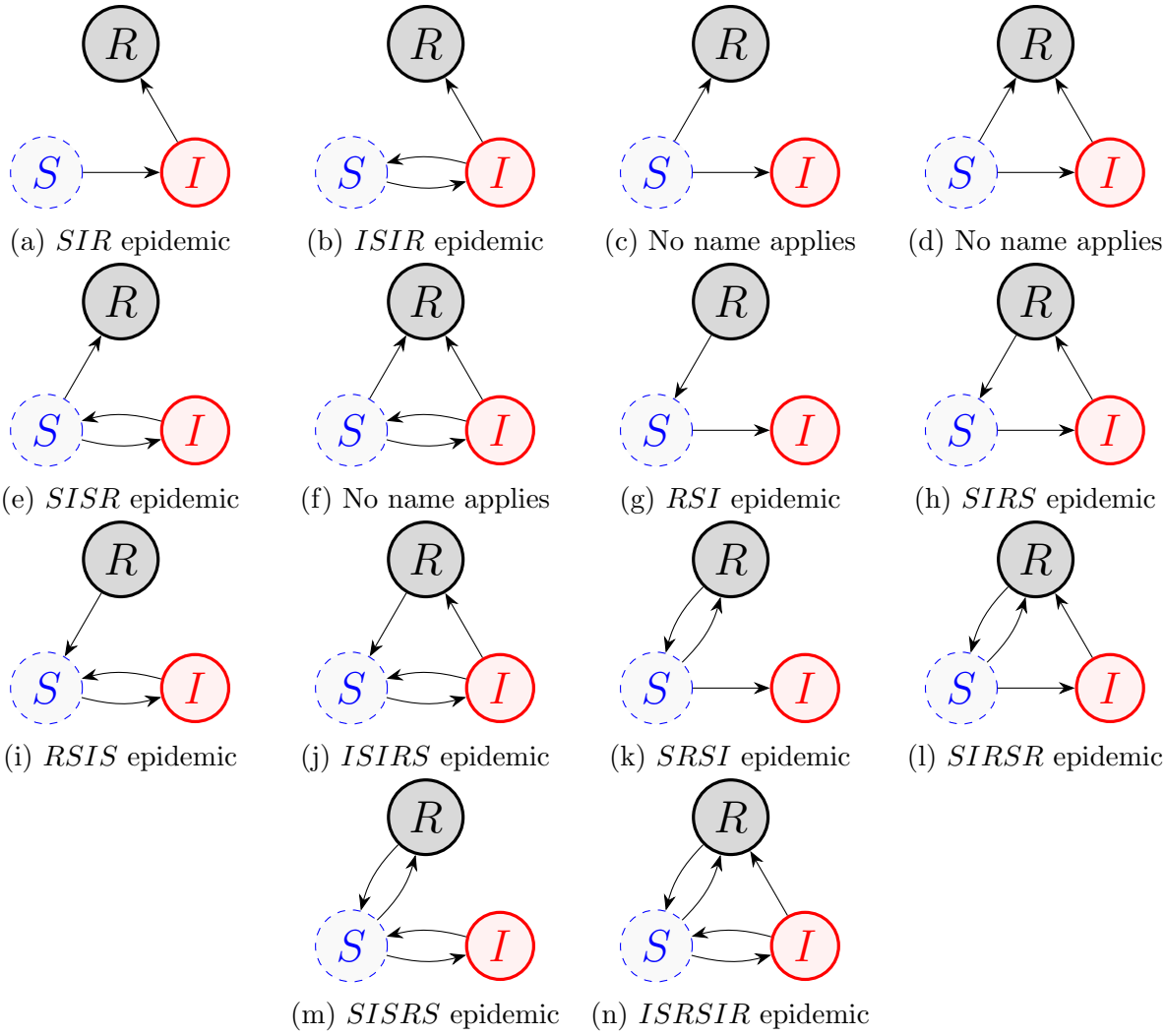
induce the number of possible transitions between two states and further on permit to list all possible epidemic models with these compartments. Indeed, in such an epidemic, we should envision four hypotheses about the transitions between the states S and R: either there is no possible transition, or only one transition is possible, or both transitions are possible. Similarly, with statements 2 and 3, we should envision two hypotheses about the transitions between the states S and I ($S \to I$ or both), and two hypotheses about the transitions between the states I and R ($I \to R$ or none). This roughly yields $4 \times 2 \times 2 = 16$ models. Now, from statement 1, we remove the two models that isolate the state R and we obtain the 14 models of figure 5.2. Note that in some cases, it is difficult to strictly account all the possible transitions. That is, all these models are referred to under the same basic SIR name.

Variants of SIR epidemic model are obtained by considering subclasses of its compartments. For example, the $SIRV$ model distinguishes between the vaccinated individuals (V compartment) to susceptible and recovered (R) dead individuals [39]. One may also admit a maternally derived immunity (compartment M) or distinguish individuals who carry the disease but cannot transmit it (C for *carrier* or *infected*). These considerations yield models like $SIRD$, $MSIR$, $SEIR$, $SEIS$, $MSEIR$, $MSEIRS$ for example.

5.2.2 Devices Transitions in Network *SIR* Epidemics

Rumor in Social Networks

The epidemics presented on figure 5.2 can also affect devices (or users) in network systems. A rumor in social networks, for example, is viewed as an SIR epidemic. The S compartment represents the individuals who may believe the rumor if they are aware of it, while the R compartment represents those who cannot believe it, and the I compartment consists of individuals who believe the rumor. In [52], the authors survey models of rumor diffusion



(a) SIR epidemic (b) $ISIR$ epidemic (c) No name applies (d) No name applies

(e) $SISR$ epidemic (f) No name applies (g) RSI epidemic (h) $SIRS$ epidemic

(i) $RSIS$ epidemic (j) $ISIRS$ epidemic (k) $SRSI$ epidemic (l) $SIRSR$ epidemic

(m) $SISRS$ epidemic (n) $ISRSIR$ epidemic

Figure 5.2: Epidemics with compartments S, I and R

The names rely on the available compartment and the possible individual transitions. All these models are also referred to as SIR epidemics. The transition $R \to S$ is a loss of immunity, while the transition $I \to S$ may happen in case of incomplete recovery. No name applies in three cases.

that extend the variants of SIR epidemic model presented on figures 5.2a, 5.2d and 5.2j, or trim down the SIS (5.1a) to the SI model. These models are $SEIR$ (Susceptible-Exposed-Infected-Removed), S-$SEIR$ (single layer-$SEIR$), $SCIR$ (SIR with "contacted" status), $irSIR$ (infection recovery SIR), $FSIR$ (fractional SIR [23]), and $ESIS$ (emotion-based SIS, [95]).

Controlling the Network through Epidemic

Another type of epidemic is spread by an attacker aiming to take control of many devices. In this category, ranges the Mirai botnet, an epidemic that was first found in 2016 and has been used in some of the largest and most distributed denial of service (DDoS) [3].

Mirai has many variants that follow the same infection strategy. With the aim to take control of numerous IoT devices, the strategy consists in injecting a malicious code from any infected device to one or more vulnerable neighbors. The vulnerability of a device is due to the fact that it still uses the default password, so it can be accessed through attempts from a limited list of possible default passwords and get injected a code that makes it a bot remotely controlled by some intruder. This is the transition $S \to I$, where susceptibility is vulnerability and the bot is an infected device, that can infect others, by executing the instructions of the intruder. The remove (or ***resistant***) compartment is that of devices with customized password, that does not lie in the list held by the intruder. The transition $I \to R$ from infected to this state is impossible because the default password is no longer working, and the IoT user cannot infer the current one. However, when the device is not infected, the transition to removed state is possible because since the attacker cannot control the device activities, the IoT user has the power to change the password. Furthermore, once the credentials have been customized, the IoT user can no longer return to the default one. This disables the transition $R \to S$. Finally, without any defense tool, the transition $S \to I$, is irreversible, and the dynamics are that of figures 5.3 and 5.2c.

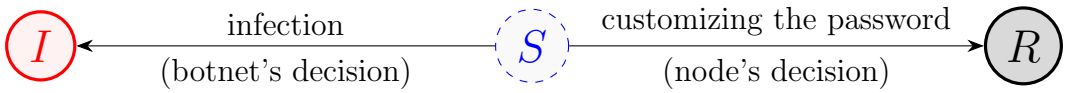


Figure 5.3: Possible state transitions in botnet epidemics without protection

Adversarial study of SIR epidemics

To mitigate such an intelligent, rational and active epidemic propagation in the network, we formulate the following adversarial scenario. An attacker is trying to take control of a large number of devices of a network and make it a foothold to launch a fatal attack. This attack may be for example to overload a server with a very large number of requests. Her strategy consists in silently spreading over the network a worm that ensures her the control of devices. She will propagate the worm until she has taken control of the desired number of devices. She frequently makes a probe over the network and then knows which nodes are vulnerable, which nodes are infected (and which nodes are resistant). To mitigate this spread, a defender combines two solutions:

1. He offers patches for infected devices and incites them to accept it. He also incites vulnerable, non-infected devices to customize their passwords and therefore become resistant against any attack. However, the result of this initiative is not predictable,

and the attacker knows it. Nevertheless, the defender knows the decision of any device, i.e., he knows whenever a device has got patched or has customized its password.

2. The defender has at his disposal a fixed number of IPSs that he can deploy on edges of the network. The validity of each IPS is one period. Note that the attacker does not have the IPSs' localization knowledge. The IPS can detect any attempt to use a default password. So it detects any virus propagation that traverses the edge, and then the defender strongly incites the device and the newly infected nodes to patch.

This scenario is repeated until the attacker has reached the targeted number of infected devices [2] or there is no infected device left in the network (the latter is an absorbing state of the system, as the threat has totally disappeared). The dynamics are summarized on figures 5.4 and 5.2e, and do not include the transitions $R \to S$ and $I \to R$ for the following

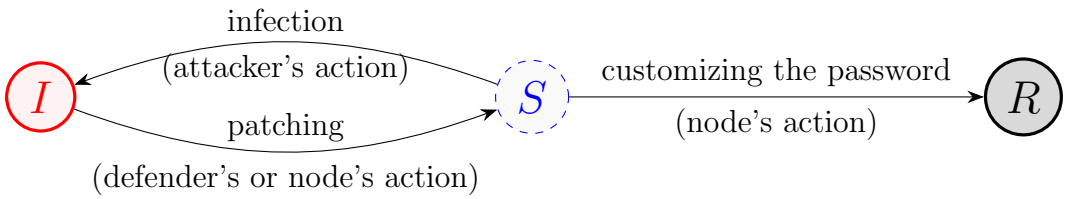


Figure 5.4: Possible state transitions in botnet epidemics with protection

reasons:

1. Once a user has customized his password, he will very unlikely set back the default password;

2. Once a device is infected, the attacker, who henceforth controls it, will give no chance to actually customize the password, so that for the device to get resistant, a non-circumventable way is to get patched, i.e., to get the malicious code removed.

Thus, notwithstanding this limitation in the number of transitions, this scenario is realistic enough to include important cases of network protection against epidemic propagation.

Cyber Deception in the Actors' Moves

As mentioned in [47], in the context of botnet propagation, the attacker's action is to inject malicious code that makes the device a zombie controlled by and communicating with the command and control server without being detected by the system or the user. This is the reason the user may not be interested in applying the corrective action proposed by the defender, which by the way is not proven. In other words, the attacker implants a code

[2] The attacker has no incentive to continue the combat one she has obtained a foothold to launch the attack she was preparing. However, this cannot be considered as a stopping criterion for the defender who does not know its value. Among other contributions, the current chapter is a step, exploited in the next one to circumvent this lack of information.

into the device and makes sure that it is not detected. This is probably a cryptic, extensive and informational cyber deception maneuver. It is: cryptic because the code is hidden, extensive because the malicious code blends into the system, and informational because the properties of the data returned by the code must be modified to better hide it. We deduce that the attacker practices an obfuscation against the users, from which the defender suffers the consequences.

This amplifies the asymmetric character of this conflict, in which the defender cannot afford, like the attacker, to scan the devices to know their status.

To reduce this asymmetry, we propose that the defender also make use of cyber deception. An IDS on a link would be sufficient to detect the transmission of the code and revoke the transmission. This would result in revealing to the attacker the defensive action taken. To hide this defensive action, we propose that the defender let the code reach its target, just until the attacker knows it. Then, before the attacker's next probe, the defender disinfects the affected device. The tool thus used is called ***intrusion-proof system (IPS)*** in the following. To kill two birds with one stone, the IPS also cleans the device through which the malicious code was transmitted. Since every infected device is likely to accept the patch, the attacker will see the transition, but cannot infer the cause. So the defender hides his action by using the action itself and taking the time to reverse the truth received by the attacker. This is a cryptic, intensive and motive cyber deception, i.e., a mixing.

5.2.3 A Game Theoretical Approach for the Mitigation of the Epidemic Spread

Summary of the Game Theoretic Framework

The above scenario involves two intelligent and rational agents (a defender and an attacker) acting on a system (the network) whose transitions they cannot control (because they are determined by the non-rational nodes), with one of the two actors aiming exactly at preventing the other from achieving its goal. Thus, we divide the time in periods, and we model the dynamics with zero-sum stochastic game (SG) framework. Such a model is neither a classical POSG [35], in which the attacker observes the defender's actions, nor a private information POSG, in which no player can observe the private state of another player [41]. Indeed:

1. The defender has partial observability of the network state, while the attacker knows the state of each node, so the SG is one-sided partial observability (POSG);

2. However, the attacker cannot infer the actions of the defender.

Moreover, the epidemic aspect puts forward two additional parameters, the endogenous probabilities of the transitions $S \to R$ and $I \to S$.

This results in the zero-sum one-sided POSG

$$\mathcal{G} = \left(Z,\, A,\, O,\, T,\, \mathcal{R},\, b^0,\, Z_{\text{goals}}\right),$$

where:

- Players 1 and 2 are respectively the defender and the attacker;
- The partial observability is on the side of player 2;
- Z, A_i, A, O, T, $\mathcal{R}$ and b^0 respectively stand for the sets of states, player i actions, action profiles, observations, the transition, reward functions, and belief of player 2 at the beginning of the first period;
- The state is a goal one if and only if no node is infected.

These remainder of this subsection is dedicated to the description of the components of this game model.

The System

The system is the network $G = (V, E)$, a finite and non-directed graph with the set V of nodes (devices) and the set E of edges. To simplify, we say that nodes are indexed by $1, 2, \ldots, i, \ldots, |V|$. An edge u is any pair $\{i, j\}$ of connected nodes. A node state can be viewed either as a subset of V (e.g., R is the set of all resistant nodes) or as a label S, I or R. The state of each of the nodes define the global state, $z = (z_i)_{i=1}^{|V|}$, of the network. Thus, the state space of the system is $Z = \{S, I, R\}^V$.

The Actions

The period is divided into two stages: the first stage consists of the players' strategic actions, and the second stage consists of the node's probabilistic actions (see figure 5.5).

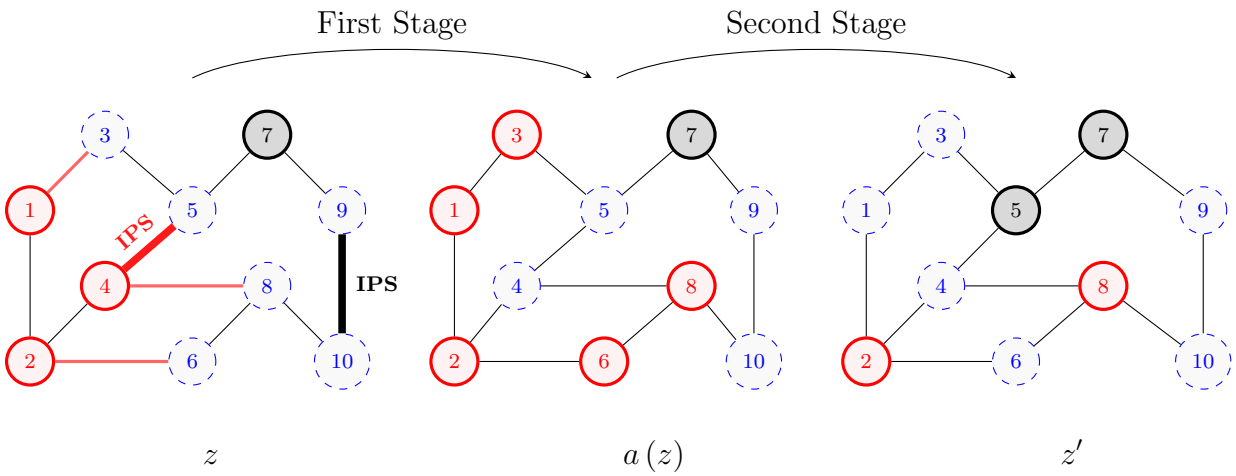
The Stake The defender's action (figure 5.6a) consists in deploying IPSs on edges of his choice, while the attacker's action (figure 5.6b) consists in choosing through which edges she will propagate the worm. Strategically, each of the underlined edges should link an infected to a susceptible node. In other words, since edges are pairs of nodes, the edge u selected by the attacker should contain an infected node, i.e., $u \cap I \neq \emptyset$, and a susceptible node, i.e., $u \cap S \neq \emptyset$. It turns out that if they know the state z of the system, both players should consider the set

$$\mathbb{S}_z = \left\{ u \in E \;\middle|\; \begin{array}{l} u \cap I \neq \emptyset \\ u \cap S \neq \emptyset \end{array} \right\}, \tag{5.1a}$$

that we refer to as the ***stake*** generated by the state z. However, the defender may not know this state and should instead consider the ***feasible stake***

$$\mathbb{S}_b = \bigcup_{\substack{z \in Z \\ b(z) \neq 0}} \mathbb{S}_z \tag{5.1b}$$

consisting of all possible edge members of the stake when the defender's belief upon the network state is b.



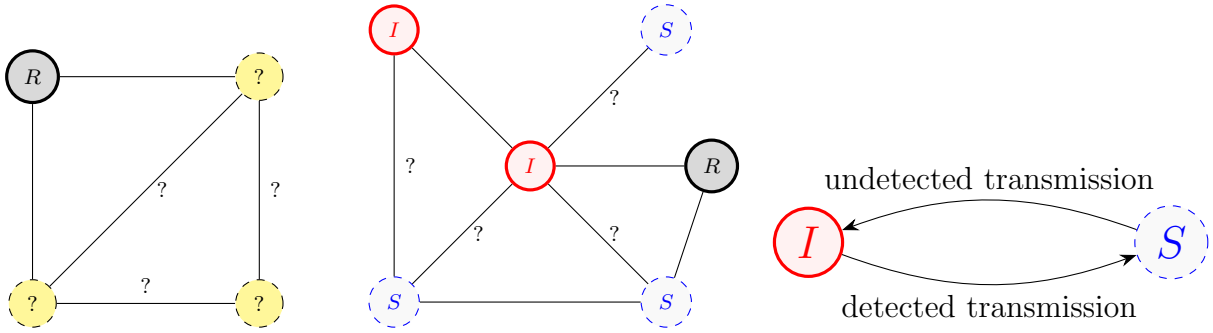
The defender places 2 IPSs, on edges $4 \leftrightarrow 5$ and $9 \leftrightarrow 10$; the attacker launches contamination $1 \to 3$, $2 \to 6$, $4 \to 5$ and $4 \to 8$; the first three ones are undetected while the last one is detected.

The propagation over nodes 3, 6 and 8 is not detected and therefore results in three new infected nodes, whereas the defender has intercepted an infection crossing the edge $4 \leftrightarrow 5$. Then, nodes 4 and 5 states becomes susceptible.

With respect to probabilities α and ρ, nodes 1, 3 and 6 transit $I \to S$, node 5 transits $S \to R$ and the other nodes do not change their states.

i = *susceptible node;* *i* = *infected node;* *i* = *resistant node;* ——— = *edge;* ——— = *edge chosen by attacker;* IPS = *IPS edge.*

Figure 5.5: One period of the game: a possible scenario with 10 nodes.



(a) Defender's action (b) Attacker's action (c) Result in each node state

S = *susceptible node (figure b) or state (figure c);* *R* = *resistant node;*
I = *infected node (figure b) or state (figure c);* *?* = *node in unknown state.*

Figure 5.6: The first stage of a period

Players Actions Formally, an action of the defender is any subset

$$a_1 = W \in \mathcal{P}_{\leqslant h}(\mathbb{S}_b) \tag{5.2a}$$

of at most h edge members of the feasible stake, where $\mathcal{P}_{\leqslant x}(X)$ denotes the set of all subsets of at most x elements of a set X and h is the maximum number of IPSs the defender can deploy at each period. An action for the attacker is a selection of a set of edges to susceptible neighbor nodes for each infected node. Denoting by Tar_i the set of edges through which the attacker propagates the malicious code from node i, an action for the attacker is any sequence

$$a_2 = \mathrm{Tar} = (\mathrm{Tar}_i)_{i \in I} \tag{5.2b}$$

such that:

1. $i \in u$ and $u \in \mathbb{S}_z$ for all infected node i and all target edge $u \in \mathrm{Tar}_i$;
2. $u \cap v = \emptyset$ whenever $u \in \mathrm{Tar}_i$ and $v \in \mathrm{Tar}_j$, for all distinct infected nodes i and j. This is, the attacker never takes the counterproductive decision to infect one node from two distinct sources.

The defender places the IPSs before the attacker sets her action. However, the attacker does not know the IPSs are allocated, and the game is therefore a simultaneous one.

The Transition Probability

The transition of nodes in one period happens in two steps:

First Intermediate Stage The players' action profile $a = (a_1, a_2)$ constitutes the first stage of the period, which is therefore referred to as the strategic stage. These actions result in a network state transition from $z = (z_i)_{i \in V}$, at the beginning of the period, to

$$a(z) = (a(z)_i)_{i \in V}, \tag{5.3}$$

at the end of the stage, that uniquely relies on the action profile a (figure 5.6c).
In case a node i is susceptible, its state changes (to infected) if and only if the attacker launches an undetected transmission from an infected node to it. In case a node i is infected, its state changes (to susceptible) if and only if the defender detects a transmission launched from its position through an IPS edge to a susceptible node. Remember that resistant nodes remain resistant. We introduce, for any collection X of sets, the union $\mathcal{U}(X) = \bigcup_{\omega \in X} \omega$ of all sets in the collection X. Take also $\partial a_1 = a_1$, $\partial a_2 = \bigcup_{i \in I} \mathrm{Tar}_i$ and $\partial a = \partial a_1 \cap \partial a_2$ respectively as the ***footprints*** of respectively the defender and the attacker actions, and ***public footprint*** of the action profile. Note that the footprint of player i's action can be partitioned into a ***private footprint***, $\partial a_i \setminus \partial a$, and the public footprint. Now refer to $\mathcal{U}(\partial a_i)$ and $\mathcal{U}(\partial a)$ as the ***borders*** of the player i action and the ***public border*** of the action profile, respectively, and pose $\mathcal{U}(\partial a_i) = \nabla a_i$ and $\mathcal{U}(\partial a) = \nabla a$. As for the footprint, the border of player i's action can be partitioned into a ***private border***, $\nabla a_i \setminus \nabla a$, and the public footprint. Note that:

- a node i is a side of an infection (either the side propagating or receiving) if and only if $i \in \nabla a_2$, i.e., i is in the border of the attacker's action;
- node i is a side of a detected infection if and only if $i \in \nabla a$, i.e., i is in the public border of the action profile;
- node i is a side of an undetected infection if and only if $i \in \nabla a_2 \setminus \nabla a$, i.e., i is in the private border of the attacker's action.

The transition $z \to a(z)$ due to players' action profile a can be explained as follows:

- the state of a susceptible node effectively transitions if and only if the node is in the private border of the attacker's action;
- the state of an infected node transitions if and only if the node is in the border of the action profile.

In other words, for all node i:

$$z_i = S \implies \begin{cases} a(z)_i = I & \iff \quad i \in \nabla a_2 \setminus \nabla a \\ a(z)_i = S & \iff \quad i \notin \nabla a_2 \setminus \nabla a \end{cases},$$

$$z_i = I \implies \begin{cases} a(z)_i = I & \iff \quad i \notin \nabla a \\ a(z)_i = S & \iff \quad i \in \nabla a \end{cases},$$

$$z_i = R \implies a(z)_i = R.$$

Second Intermediate Stage At the second stage, as summarized in figure 5.7 and table 5.1, each infected node performs the transition $I \to S$ at probability α while each

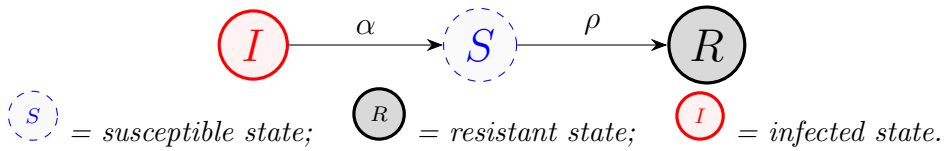


Figure 5.7: The second stage of a period

		z_i'		
		S	I	R
$a(z)_i$	S	$1-\rho$	0	ρ
	I	α	$1-\alpha$	0
	R	0	0	1

Table 5.1: Probabilities of nodes' probabilistic state transitions

susceptible node performs the transition $S \to R$ at probability ρ. The probabilities α and ρ define the conditional probability $\mathbb{P}\left(z' \,|\, a\left(z\right)\right)$ of final transition to any state z' when the network transitioned to the intermediate state $a\left(z\right)$. To easily express the transition probability $\mathbb{P}$, take the node states as the column matrices $S = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $I = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $R = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, and define the nodes ***transition matrix***

$$\mathbf{T} = \begin{pmatrix} 1-\rho & 0 & \rho \\ \alpha & 1-\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.4}$$

Clearly, $\mathbb{P}\left(z'_i \,|\, a\left(z\right)_i\right) = \left(z'_i\right)^{\mathrm{T}} \cdot \mathbf{T} \cdot a\left(z\right)_i$, where $\left(M\right)^{\mathrm{T}}$ stands for the transpose of a matrix M. Assuming the independence of nodes' transitions to each other, it comes that

$$\mathbb{P}\left(z' \,|\, a\left(z\right)\right) = \prod_{i \in V} \mathbb{P}\left(z'_i \,|\, a\left(z\right)_i\right) = \prod_{i \in V} \left(z'_i\right)^{\mathrm{T}} \cdot \mathbf{T} \cdot a\left(z\right)_i . \tag{5.5}$$

The Observations

The defender observation is composed of his knowledge on each node state at the end of each period. In the general case, if the defender is aware of the final state z'_i, the observation is $o_i = z'_i$. Otherwise, there is no actual observation for the defender and we denote $o_i := \mathfrak{X}$. As shown in table 5.2, 13 case exist and the observation o_i generated about a node i relies on the action profile a and the state transition $z_i \to z'_i$ of the node. The difference between cases 5 and 9 for example resides only in the action involving the nodes. For a susceptible node

	1, $1'$	2, $2'$	3	4	5	6	7	8	9	10	11
z_i	I	I	I	I	S	S	S	S	S	S	R
a	NT or UT		DT	DT	NT	NT	UT	UT	DT	DT	NT
$a\left(z\right)_i$	I	I	S	S	S	S	I	I	S	S	R
z'_i	I	S	S	R	S	R	I	S	S	R	R
o_i	$\mathfrak{X}$	S	S	R	$\mathfrak{X}$	R	$\mathfrak{X}$	S	S	R	R

Table 5.2: Observation of the defender at the end of the period
UT = the node is involved in an undetected transmission and is not involved in any detected transmission, DT = the node is involved in a detected transmission, NT = the node is not involved in any detected transmission, $\mathfrak{X}$ = the defender cannot infer the node's state

whom state did not changed the defender observes the final state only in case of trapped transmission. Case 5 is that of nodes 9 and 10 of figure 5.5, while case number 9 is the one of node 5. In cases 1 and $1'$, the node state remains infected during the period (nodes 1 and 2) and the defender cannot infer the final state. Finally, the observation space is $O = \{S, R, \mathfrak{X}\}^V$, and we note

$$\omega\left(z' | z, a\right) = \left(\omega\left(z' | z, a\right)_i\right)_{i \in V}, \tag{5.6}$$

the observation generated to the defender at the end of a period at which the action profile a was taken in state z and the network transitioned to state z'. In table 5.2, the defender does not make the observation only in cases 1, 1′, 5 and 7. It turns out that

$$\omega\left(z'|z,a\right)_i = \begin{cases} \mathfrak{X} & \text{if } z'_i = I \text{ or node } i \text{ absolutely remains susceptible} \\ z_i & \text{otherwise} \end{cases}, \tag{5.7}$$

remaining susceptible mining that node i was already susceptible at the beginning of the period and did not benefit on the IPS deception.

The transition function

The transition function can be explicitly defined by

$$T\left(z', o \,|\, z, a\right) = \begin{cases} \mathbb{P}\left(z' \,|\, a\left(z\right)\right) & \text{if } o = \omega\left(z'|z,a\right) \\ 0 & \text{otherwise} \end{cases}. \tag{5.8a}$$

Then, by equation (5.5), it turns out that, if $o = \omega\left(z'|z,a\right)$, then

$$T\left(z', o \,|\, z, a\right) = \prod_{i \in V} \left(z'_i\right)^{\mathrm{T}} \cdot \mathbf{T} \cdot a\left(z\right)_i . \tag{5.8b}$$

The Rewards

The defender's and the attacker's rewards are opposite to each other, and, for the attacker, it measures the increase of the satisfaction of progressing towards her objectives. This means that the utility function is the sum (or the discounted sum). Other reward models are discussed in chapters 7 and 6. From a general perspective, we define in the following paragraphs the exact reward r of the defender as an aggregation through a function w of marginal rewards μ resulting in nodes' individual transitions.

The marginal reward associated with any node's individual transition from a state A at the beginning of the period to a state B at the end of the period uniquely relies on the states A and B, and is therefore noted $\mu\left(A, B\right)$. Note that $\mu\colon \{S, I, R\}^2 \longrightarrow \mathbb{R}$ is a fixed function, i.e., μ is node and period independent. This is equivalent to associating the nodes' state transitions with 6 parameters (see table 5.3). In matrix notation, the function μ can be rewritten $\mu\left(z, z'\right) = z^{\mathrm{T}}\mathbf{R}z'$, where

$$\mathbf{R} = \begin{pmatrix} \mu_0 & -\mu'_3 & \mu_4 \\ \mu_3 & \mu_1 & 0 \\ 0 & 0 & \mu_2 \end{pmatrix} \tag{5.9}$$

is referred to as the ***reward matrix***. Because they reflect an advantage of the defender, the six parameters μ_0 to μ_4 and μ'_3 must be assumed positive. In other words,

$$\mu_0 \geqslant 0, \ \ldots, \ \mu_4 \geqslant 0, \ \mu'_3 \geqslant 0. \tag{5.10}$$

		End of period, state z_i'		
		S	I	R
Beginnig of period, state z_i	S	μ_0	$-\mu_3'$	μ_4
	I	μ_3	$-\mu_1$	//
	R	//	//	μ_2

Table 5.3: Marginal rewards associated with a node's state transition
The transitions $R \to S$, $R \to I$ and $I \to R$ are impossible, so it is irrelevant to associate them with any reward. It sounds intuitive that the five parameters μ_0 to μ_4 are non-negative.

The global reward associated with the network transition from state z to state z' is a w-aggregation of the marginal rewards $\mu(z_i, z_i')$, where w is a function from $\mathbb{R}^V$ to $\mathbb{R}$. Which means that

$$r(z, z') = w(\mu(z, z')) = w \circ \mu(z, z'), \tag{5.11}$$

where $\mu(z, z') = (\mu(z_i, z_i'))_{i \in V}$. We assume that the more important the marginals are, the most important the total is, i.e., w is an increasing function. In other words, for all p and q in $\mathbb{R}^V$,

$$p_i \leqslant q_i \quad \forall i \in \mathbb{R} \qquad \Longrightarrow \qquad w(p) \leqslant w(q). \tag{5.12}$$

Example 5.2.1. *If the reward matrix is* $\mathbf{R} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ *and w is the sum, then $r(z, z')$ is the increase in the number of infected nodes when the system transitions from states z to z'.*

Since he does not know the state of the network, the defender does not see this reward. Instead of calculating the exact sum of the marginal rewards, the study of stochastic games very often focuses on its mathematical expectation when the action profile and the state are known. The defender reward associated with the action profile a taken in a network state z is the expected value of the exact reward $-r(z, z')$ generated if the network transitions to state z'. That is,

$$\mathcal{R}(z, a) = \sum_{z' \in Z} \mathbb{P}(z' \mid a(z)) \cdot r(z, z'). \tag{5.13}$$

5.2.4 The Utility and the Value Function

In example 5.2.1, the negative sum of the rewards represents the total number of infected nodes at the end of the game. Thus, defining the situation in which a sufficiently large number of nodes are infected as a goal state ensures that the utility, seen as the sum of the rewards, is, for the attacker, the peak of the epidemic. This is quite realistic. However, solving a POSG with goal state relies on the convergence of the value iteration algorithm of a POSG with discounted sum. This convergence was proved in [32] under the assumption of perfect information of the attacker, an assumption that is not allowed in our model, since

the attacker cannot infer the actions of the defender. We therefore show that the value iteration algorithm remains convergent when we remove the perfect information assumption from the perfect information player side. That is, the utility we consider in this chapter is the discounted sum, with discount factor $\gamma \in (0,1)$.

Denote

$$U_{min} = \sum_{t=0}^{\infty} \gamma^t \left(\min_{(z,a) \in Z \times A} \mathcal{R}(z,a) \right) = \frac{\min\limits_{(z,a) \in Z \times A} \mathcal{R}(z,a)}{1-\gamma} = \frac{\mathcal{R}_{min}}{1-\gamma} \tag{5.14a}$$

and

$$U_{max} = \sum_{t=0}^{\infty} \gamma^t \left(\max_{(z,a) \in Z \times A} \mathcal{R}(z,a) \right) = \frac{\max\limits_{(z,a) \in Z \times A} \mathcal{R}(z,a)}{1-\gamma} = \frac{\mathcal{R}_{max}}{1-\gamma} \tag{5.14b}$$

where $\mathcal{R}_{min} = \min\limits_{z,a_1,a_2} \mathcal{R}(z,a_1,a_2)$ and $\mathcal{R}_{max} = \max\limits_{z,a_1,a_2} \mathcal{R}(z,a_1,a_2)$. It is clear each reward is bounded between $\mathcal{R}_{min}$ and $\mathcal{R}_{max}$. Consequently, each utility function, and then the optimal value function, are bounded between U_{min} and U_{max}.

We establish the convexity and Lipschitz continuity of the optimal value function V^* defined in equations (3.8) and (3.9). Consider the following $||.||_1$-norm over the belief space as:

$$\|b\|_1 = \sum_{z \in Z} b(z). \tag{5.15}$$

Denote $\delta = \dfrac{|U_{max} - U_{min}|}{2} = \dfrac{U_{max} - U_{min}}{2}$. The following lemmas are proven in [32] and rewritten here for clarity.

Lemma 5.2.1 ([32]). *A function $f : \Delta(Z) \longrightarrow \mathbb{R}$ is continuous and convex if and only if f is a point-wise supremum over a set Γ of linear functions, i.e., $f(b) = \sup\limits_{\alpha \in \Gamma} \alpha(b)$ for every b in $\Delta(Z)$. Furthermore, if for a certain k all α in Γ are k-Lipschitz continuous, then f is k-Lipschitz continuous.*

Lemma 5.2.2 ([32]). *Any function bounded over $\Delta(Z)$ between U_{min} and U_{max} is δ-Lipschitz continuous.*

Therefore, based on previous lemma we have the following direct result.

Lemma 5.2.3. *The value function val_{σ_1} of any fixed strategy σ_1 of defender is δ-Lipschitz continuous.*

Proof. From lemma 3.5.1, val_{σ_1} is linear. As it is bounded between U_{min} and U_{max}, by lemma 5.2.2, it is δ-Lipschitz. □

Then, the following theorem gives the main result over the optimal value function V^*.

Theorem 5.2.1. *The optimal value function V^* of the game is convex and δ-Lipschitz continuous in the belief vector b.*

Proof. For any beliefs b and b' and for any strategy σ_1 of defender, the δ-Lipschiz property of val_{σ_1} (lemma (5.2.3)) gives: $\mathrm{val}_{\sigma_1}(b) - \mathrm{val}_{\sigma_1}(b') \leqslant \delta\|b - b'\|_1$, then $\mathrm{val}_{\sigma_1}(b) \leqslant \mathrm{val}_{\sigma_1}(b') + \delta\|b - b'\|_1$. So, for any beliefs b and b', $\max\limits_{\sigma_1\in\Sigma_1} \mathrm{val}_{\sigma_1}(b) \leqslant \max\limits_{\sigma_1\in\Sigma_1} (\mathrm{val}_{\sigma_1}(b') + \delta\|b - b'\|_1) = \max\limits_{\sigma_1\in\Sigma_1} (\mathrm{val}_{\sigma_1}(b')) + \delta\|b - b'\|_1$. In other notations, $V^*(b) \leqslant V^*(b') + \delta\|b - b'\|_1$. Hence, V^* is δ-Lipschitz.

The convexity comes from the fact that V^* is a point-wise maximum of the linear functions val_{σ_1}. Indeed, for any b, $b' \in \Delta(Z)$ and any $\lambda \in [0,1]$:

$$\begin{aligned}
\lambda V^*(b) + (1-\lambda) V^*(b') &= \lambda \max_{\sigma_1\in\Sigma_1} \mathrm{val}_{\sigma_1}(b) + (1-\lambda) \max_{\sigma_1\in\Sigma_1} \mathrm{val}_{\sigma_1}(b') \\
&= \max_{\sigma_1\in\Sigma_1} \lambda \mathrm{val}_{\sigma_1}(b) + \max_{\sigma_1\in\Sigma_1} (1-\lambda)\,\mathrm{val}_{\sigma_1}(b') \\
&\leqslant \max_{\sigma_1\in\Sigma_1} (\lambda \mathrm{val}_{\sigma_1}(b) + (1-\lambda)\,\mathrm{val}_{\sigma_1}(b')) \\
&= \max_{\sigma_1\in\Sigma_1} \mathrm{val}_{\sigma_1}(\lambda b + (1-\lambda) b') \qquad \text{(by linearity)} \\
&= V^*(\lambda b + (1-\lambda) b').
\end{aligned}$$

□

The result of the previous section (theorem 5.2.1) distinguishes the optimal value function among functions $\Delta(Z) \longrightarrow \mathbb{R}$. So, to compute this optimum, we present it as the limit of a converging sequence of such functions that we term value functions. The convergence is further guarantied by a Lipschitz-continuous operator, the backup operator.

Definition 5.2.1 (***value function***)**.** *A* value function *of our game is any real valued function $V : \Delta(Z) \longrightarrow \mathbb{R}$. The set of the value functions of the game is denoted by $\mathbb{V}$.*

5.3 Computing the optimal strategies

5.3.1 Definition of the Value Backup Operator

Starting from any value function, the value backup operator updates iteratively the value of the game. Assume that at a given period, the defender holds a belief vector b upon the system state, and the optimal value V of the next period onward is known to both players. That is, the overall POSSPG is summarized in a (static) zero-sum strategic-form game with action spaces A_1 and A_2. In this game, players initially receive an immediate reward that depends on their immediate actions. Then, since they know the optimal value of the subgame of the stochastic game that starts in the next period, they receive the optimal reward for subsequent actions, it is assumed that their behavior will be optimal. This second reward can be estimated from the players' stage strategies. This game is referred to as a stage

game. [3] The defender's utility in this stage game when strategies π_1 and π_2 are played is given by:

$$U^V_{\pi_1,\pi_2}(b) = \mathcal{R}^{\text{imm}}_{\pi_1,\pi_2}(b) + \gamma \mathcal{R}^{\text{subs}}_{\pi_1,\pi_2}(b, V), \tag{5.16a}$$

where

$$\begin{aligned} \mathcal{R}^{\text{imm}}_{\pi_1,\pi_2}(b) &= \sum_{(a,z)\in A\times Z} b(z)\,\pi_1(a_1)\,\pi_2(a_2 \mid z)\,\mathcal{R}(z,(a_1,a_2)) \\ &= \sum_{(a,z)\in A\times Z} b(z)\,\pi(a \mid z)\,\mathcal{R}(z,a) \end{aligned} \tag{5.16b}$$

is the reward in the stage game as defined in subsection 3.4.2, and

$$\mathcal{R}^{\text{subs}}_{\pi_1,\pi_2}(b, V) = \sum_{(a_1,o)\in A_1\times O} \pi_1(a_1)\,\mathbb{P}(o \mid b, a_1, \pi_2)\,V\Big(\tau(b|a_1,\pi_2,o)\Big) \tag{5.16c}$$

is the reward in the subsequent game. Let us now define a stage game.

Definition 5.3.1 (***stage game***)**.** *For all belief $b \in \Delta(Z)$ and value function $V \in \mathbb{V}$, the* stage game *associated with the belief b and the value function V is the zero-sum normal-form game* $\text{SG}(b, V)$ *with the two players, the defender as player* 1 *and the attacker as player* 2*, and action sets A_1 and A_2, respectively.*

For all strategy profile (π_1, π_2), the defender's expected reward $U^V_{\pi_1,\pi_2}(b)$ is the result of equations (5.16a). The optimal expected reward of the stage game is:

$$\begin{aligned} [HV](b) &= \max_{\pi_1\in\Delta(A_1)} \min_{\pi_2\in\Delta(A_2)} U^V_{\pi_1,\pi_2}(b) \\ &= \max_{\pi_1\in\Delta(A_1)} \min_{\pi_2\in\Delta(A_2)} \left[\mathcal{R}^{\text{imm}}_{\pi_1,\pi_2}(b) + \gamma \mathcal{R}^{\text{subs}}_{\pi_1,\pi_2}(b, V)\right], \end{aligned} \tag{5.16d}$$

which is the optimal value in belief b of the overall POSSPG, if it starts at the underlined period. That is, H is an operator that returns the optimal value function of the POSSPG at any period next to which the optimal function is known.

Definition 5.3.2 (***value backup operator***)**.** *The* (value) backup operator *is the application $H\colon \mathbb{V} \longrightarrow \mathbb{V}$, where the value HV is defined by the optimal value function given in equation (5.16d).*

In order to compare the properties of any value function and its backup value function, the stability of some subsets of value function is needed under the action of the operator H. One way to prove this main result is to show that the value backup operator H preserves the convexity and the Lipschitz continuity. In order to get these properties, we need the following set of technical lemmas. In the first one, lemma 5.3.1, we state the interchangeability of the sum and the supremum. The second one says that the value of the defender in the stage game would be linear and bounded if any linear function replaced the subsequent reward. Finally, the third one states the stability of some set of value functions under the backup operator.

[3] See subsection 3.4.2.

Lemma 5.3.1. *For every non-empty sets M and N and every function $a : M \times N \longrightarrow \mathbb{R}$, if M is finite, and $\{a(s,y) : y \in N\}$ and $\left\{\sum_{x\in M} a(x,y_x) : \overline{y} \in N^M\right\}$ are bounded from above for every $s \in M$ and every $\overline{y} \in N^M$, then it holds:*

$$\sum_{x\in M} \sup_{y\in N} a(x,y) = \sup_{\overline{y}\in N^M} \sum_{x\in M} a(x,y_x).$$

In this lemma, $\overline{y} = (y_s)_{s\in M}$ is a sequence of elements of N with indices in M.

Proof. Denote $A = \sum_{x\in M} \sup_{y\in N} a(x,y)$ and $B = \sup_{\overline{y}\in N^M} \sum_{x\in M} a(x,y_x)$. Take any sequence $\overline{y} \in N^M$. The inequality $a(x,y_x) \leqslant \sup_{y\in N} a(x,y)$ holds for any $x \in M$. So, $\sum_{x\in M} a(x,y_x) \leqslant \sum_{x\in M} \sup_{y\in N} a(x,y)$. Hence, $B \leqslant A$. To see the converse, consider the number $|M|$ of elements in M and take any $\varepsilon > 0$. For any $x \in M$, the inequality $a(x,y_x) \geqslant \sup_{y\in N} a(x,y) - \frac{\varepsilon}{|M|}$ holds for at least one y_x in N. Thus we have defined a sequence $\overline{y}$ such that $\sum_{x\in M} a(x,y_x) \geqslant \sum_{x\in M}\left(\sup_{y\in N} a(x,y) - \frac{\varepsilon}{|M|}\right) = \sum_{x\in M} \sup_{y\in N} a(x,y) - \varepsilon$. Applying to the supremum, we get $B \geqslant A - \varepsilon$ for every $\varepsilon > 0$. Then $B \geqslant A$ and, finally, $B = A$. □

In the remainder of this chapter, we use the notation $\overline{M}$ to represent any $(|A_1|, |O|)$-matrix $(M^{a_1,o})_{(a_1,o)\in A_1\times O}$.

Lemma 5.3.2. *For all set $\Gamma \subseteq \mathrm{lin}_{\Delta(Z)}$ of linear value functions over $\Delta(Z)$, bounded between U_{min} and U_{max}. For all sequence $\overline{\alpha} \in \Gamma^{A_1\times O}$ of these value functions and all defender stage strategy $\pi_1 \in \Pi_1$, consider the function $V^{SG}_{\pi_1,\overline{\alpha}} : \begin{array}{ccc} \Delta(Z) & \longrightarrow & \mathbb{R} \\ b & \longmapsto & \min_{\pi_2\in\Pi_2} u_{\pi_1,\pi_2}(\overline{\alpha}, b) \end{array}$ where, for all defender belief b and all attacker stage strategy π_2,*

$$\begin{aligned} u_{\pi_1,\pi_2}(\overline{\alpha}, b) = u_\pi(\overline{\alpha}, b) \;=\; & \sum_{(a,z)\in A\times Z} b(z)\,\pi(a \mid z)\,\mathcal{R}(z,a) + \\ & + \gamma \sum_{(a,z,z')\in A\times Z^2} T(z', o \mid z, a)\, b(z)\,\pi(a \mid z)\,\alpha^{a_1,o}(z'). \end{aligned} \quad (5.17)$$

Then $V^{SG}_{\pi 1,\overline{\alpha}}$ is linear and bounded between U_{min} and U_{max} (see lemma 3.5.1). It is therefore δ-Lipschitz continuous.

To better understand the function $V^{SG}_{\pi_1,\overline{\alpha}}$. Define a two-player (defender and attacker) zero-sum static one-sided partially observable $\overline{\alpha}$***-stage game*** that defers from the stage game defined in definition 5.3.1 uniquely in the reward. More precisely, for the $\overline{\alpha}$-stage

game, the defender accumulates two rewards: the first is that of defintion 5.16a, but the second reward is $\alpha^{a_1,o}(z')$ if the system transitions into state z' and generates observation o. The expected reward associated with strategy profile π is $u_\pi(\overline{\alpha}, b)$, and $V^{SG}_{\pi_1,\overline{\alpha}}(b)$ is the defender's value associated with strategy π_1 in the $\overline{\alpha}$-stage game.

Proof. We prove the Lipschitz property, then the linearity.

$\overline{\alpha} = (\alpha^{a_1,o})_{(a_1,o)\in A_1\times O}$ is a sequence of elements of Γ, i.e., for all $(a_1, o) \in A_1 \times O$, the function $\alpha^{a_1,o}\colon \Delta(Z) \longrightarrow \mathbb{R}$ is linear and bounded between U_{min} and U_{max}.

We have the following relationship:

$$u_\pi(\overline{\alpha}, b) = \sum_{(a,z,z')\in A\times Z} b(z)\,\pi(a\,|\,z)\,\mathcal{R}(z,a) + \gamma \sum_{(a_1,o)\in A_1\times O} \pi_1(a_1)\,\mathbb{P}^{\pi_2}_b(o\,|\,a_1)\,\mathcal{V}(a_1,o),$$

where $\mathcal{V}(a_1,o) = \dfrac{\displaystyle\sum_{(a_2,z,z')\in A_2\times Z^2} T(z', o\,|\,z, a_1, a_2)\,\alpha^{a_1,o}(z')}{\displaystyle\sum_{(a_2,z,z')\in A_2\times Z^2} T(z', o\,|\,z, a_1, a_2)}$ is the expected second reward (before the state transition) and $\mathbb{P}^{\pi_2}_b(o\,|\,a_1)$ is the probability that the system generates observation o whe the defender took action a_1. So, the boundedness of the $\alpha^{a_1,o}$'s between implies the boundedness of the $\mathcal{V}(a_1,o)$, then that of $(\overline{\alpha}, b)$ between the same limits. This is, for any $\overline{\alpha}$ and b:

$$U_{min} \leqslant u_\pi(\overline{\alpha}, b) \leqslant U_{max}.$$

So, $u_\pi(\overline{\alpha}, \cdot)$ and consequently $V^{SG}_{\pi_1,\overline{\alpha}}$ are bounded between U_{min} and U_{max}. Then from lemma 5.2.2 the linearity of $V^{SG}_{\pi_1,\overline{\alpha}}$ (if established) implies that $V^{SG}_{\pi_1,\overline{\alpha}}$ is also δ-Lipschitz continuous.

Now consider a defender's strategy π_1 and a sequence $\overline{\alpha} \in \Gamma^{A_1\times O}$ and consider the linear function $u'_{\pi_1}(\overline{\alpha}, .)$ defined on vertices of $\Delta(Z)$ by:

$$u'_{\pi_1}(\overline{\alpha}, z) = \min_{a_2} u_{\pi_1,a_2}(\overline{\alpha}, z),$$

with

$$\begin{aligned} u_{\pi_1,a_2}(\overline{\alpha}, z) &= \sum_{(a_1,z')\in A_1\times Z} \pi_1(a_1)\,\mathcal{R}(z,a_1,a_2) + \\ &\quad + \gamma \sum_{(a_1,o)\in A_1\times O} \pi_1(a_1)\,\mathbb{P}(o\,|\,z,a_1\,,a_2)\,\mathcal{V}(a_1,o) \\ &= \sum_{a_1\in A_1} \pi_1(a_1)\left[\sum_{z'\in Z} \mathcal{R}(z,a_1,a_2) + \gamma \sum_{o\in O} \mathbb{P}(o\,|\,z,a_1\,,a_2)\,\mathcal{V}(a_1,o)\right]. \end{aligned}$$

We can show that $\min_{\pi_2} u_{\pi_1,\pi_2}(\overline{\alpha}, b) = u'_{\pi_1}(\overline{\alpha}, b)$. In fact for any $\pi_2 \in \Delta(A_2)$ we have:

$$u_{\pi_1,\pi_2}(\overline{\alpha}, z) = \sum_{a_2} \pi_2(a_2\,|\,z) \sum_{a_1\in A_1} \pi_1(a_1)\left[\sum_{z'\in Z} \mathcal{R}(z,a_1,a_2) + \gamma \sum_{o\in O} \mathbb{P}(o\,|\,z,a_1\,,a_2)\,\mathcal{V}(a_1,o)\right]$$

$$\geqslant \sum_{a_2} \pi_2(a_2 \,|\, z)\, u'_{\pi_1}(\overline{\alpha}, z) = u'_{\pi_1}(\overline{\alpha}, z).$$

So $u'_{\pi_1}(\overline{\alpha}, b) \leqslant u_{\pi_1,\pi_2}(\overline{\alpha}, b)$ holds for any $b \in \Delta(Z)$. Now, we define a particular strategy π_2^* for the attacker as follows. Consider the non-empty set $\partial a_2 = \left\{a_2 \in A_2 : u_{\pi_1,a_2}(\overline{\alpha}, z) = u'_{\pi_1}(\overline{\alpha}, z)\right\}$ and say for any state $z \in Z$ and any attacker action $a_2 \in A_2$:

$$\pi_2^*(a_2 \,|\, z) = \begin{cases} \dfrac{1}{|\partial a_2|} & \text{if } a_2 \in A_2 \\ 0 & \text{otherwise} \end{cases}.$$

For any $z \in Z$, we get:

$$u_{\pi_1,\pi_2^*}(\overline{\alpha}, z) = \sum_{a_2 \in A_2} \pi_2^*(a_2 \,|\, z)\, u_{\pi_1,a_2}(\overline{\alpha}, z) = \sum_{a_2 \in \partial a_2} \frac{1}{|\partial a_2|} u'_{\pi_1}(\overline{\alpha}, z) = u'_{\pi_1}(\overline{\alpha}, z).$$

Then, $u_{\pi_1,\pi_2^*}(\overline{\alpha}, b) = \sum_{z \in Z} b(z)\, u_{\pi_1,\pi_2^*}(\overline{\alpha}, z) = \sum_{z \in Z} b(z)\, u'_{\pi_1}(\overline{\alpha}, z) = u'_{\pi_1}(\overline{\alpha}, b)$, which proves the linearity of $\min_{\pi_2} u_{\pi_1,\pi_2}(\overline{\alpha}, b)$ in $b \in \Delta(Z)$. □

Lemma 5.3.3. *The set of point-wise suprema of linear applications over $\Delta(Z)$ is stable under the value backup operator H.*

The lemma means that if a value function $V \in \mathbb{V}$ satisfies the following property:

$$\exists \Gamma \subseteq \mathrm{lin}_{\Delta(Z)} \quad \text{such that} \quad \forall b \in \Delta(Z), \quad V(b) = \sup_{\alpha \in \Gamma} \alpha(b), \tag{5.18}$$

then, the backup value function HV satisfies the same property.

Proof. Suppose the value V satisfies the property defined in equation (5.18) for some $\Gamma \subseteq \mathrm{lin}_{\Delta(Z)}$ and consider the stage game with value function V and belief b. For any action profile π, the reward of the subsequent game is:

$$\begin{aligned}
\mathcal{R}_\pi^{\mathrm{subs}}(b, V) &= \sum_{(a_1,o) \in A_1 \times O} \pi_1(a_1)\, \mathbb{P}_b^{\pi_2}(o \,|\, a_1)\, V\Big(\tau(b|a_1, \pi_2, o)\Big) \\
&= \sum_{(a_1,o) \in A_1 \times O} \pi_1(a_1)\, \mathbb{P}_b^{\pi_2}(o \,|\, a_1) \sup_{\alpha \in \Gamma} \alpha\Big(\tau(b|a_1, \pi_2, o)\Big) \\
&= \sum_{(a_1,o) \in A_1 \times O} \pi_1(a_1)\, \mathbb{P}_b^{\pi_2}(o \,|\, a_1) \times \\
&\qquad \times \sup_{\alpha \in \Gamma} \frac{\displaystyle\sum_{\substack{a_2 \in A_2 \\ (z,z') \in Z^2}} T(z', o \,|\, z, a)\, b(z)\, \pi_2(a_2 \,|\, z)\, \alpha(z')}{\mathbb{P}_b^{\pi_2}(o \,|\, a_1)}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{(a_1,o)\in A_1\times O} \sup_{\alpha\in\Gamma} \sum_{z'\in Z} \left(\sum_{(a_2,z)\in A_2\times Z} T\left(z', o \mid z, a\right) b\left(z\right) \pi\left(a \mid z\right) \right) \alpha\left(z'\right) \\
&= \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} \sum_{(a_1,o)\in A_1\times O} \sum_{z'\in Z} \left(\sum_{(a_2,z)\in A_2\times Z} T\left(z' \mid z, a\right) b\left(z\right) \pi\left(a \mid z\right) \right) \alpha^{a_1,o}\left(z'\right) \\
&= \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} \sum_{\substack{(a,z,z')\in A\times Z^2 \\ o\in O}} T\left(z', o \mid z, a\right) b\left(z\right) \pi\left(a \mid z\right) \alpha^{a_1,o}\left(z'\right),
\end{aligned}
$$

and the defender's utility of the stage game is:

$$
\begin{aligned}
U_{\pi}^{V}\left(b\right) &= \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} \Bigg[\sum_{(a,z,z')\in A\times Z} b\left(z\right) \pi\left(a \mid z\right) \mathcal{R}\left(z, a_1, a_2\right) + \\
&\qquad + \gamma \sum_{\substack{(a,z,z')\in A\times Z^2 \\ o\in O \\ o=o(a,z,z')}} T\left(z' \mid z, a\right) b\left(z\right) \pi\left(a \mid z\right) \alpha^{a_1,o}\left(z'\right) \Bigg] \\
&= u_{\pi}\left(\overline{\alpha}, b\right).
\end{aligned}
$$

Then the value backup operator of V is given by:

$$
\left[HV\right]\left(b\right) = \max_{\pi_1\in\Pi_1} \min_{\pi_2\in\Pi_2} \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} u_{\pi_1,\pi_2}\left(\overline{\alpha}, b\right).
$$

The utility $u_{\pi_1,\pi_2}\left(\overline{\alpha}, b\right)$ is linear in π_2. Without any loss of generality, we assume the convexity of Γ and consequently of $\Gamma^{A_1\times O}$. In fact, the point-wise supremum over a set of functions is also the point-wise supremum over its convex hull. Hence, from the convexity of Π_2, Sion's minmax theorem applies [32] and then:

$$
\begin{aligned}
\left[HV\right]\left(b\right) &= \max_{\pi_1\in\Pi_1} \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} \min_{\pi_2\in\Pi_2} u_{\pi_1,\pi_2}\left(\overline{\alpha}, b\right), \\
&= \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} \max_{\pi_1\in\Pi_1} \min_{\pi_2\in\Pi_2} u_{\pi_1,\pi_2}\left(\overline{\alpha}, b\right).
\end{aligned}
$$

As the set Π_1 is convex, $\left(\Gamma^{A_1\times O}\right) \times \Pi_1$ is also convex as well, and we get:

$$
\begin{aligned}
\left[HV\right]\left(b\right) &= \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} \max_{\pi_1\in\Pi_1} \min_{\pi_2\in\Pi_2} u_{\pi_1,\pi_2}\left(\overline{\alpha}, b\right) \\
&= \sup_{(\pi_1,\overline{\alpha})\in\Pi_1\times\Gamma^{A_1\times O}} \min_{\pi_2\in\Pi_2} u_{\pi_1,\pi_2}\left(\overline{\alpha}, b\right) \\
&= \sup_{(\pi_1,\overline{\alpha})\in\Pi_1\times\Gamma^{A_1\times O}} V_{\pi_1,\overline{\alpha}}^{SG}\left(b\right). \qquad (5.19)
\end{aligned}
$$

Then HV satisfies property (5.18). □

Based on these technical lemmas, the following theorem shows the convexity and the δ-Lipschitz continuity of the value backup operator H.

Theorem 5.3.1. *The set of convex continuous value functions and the set of convex δ-Lipschitz continuous value functions are stable under the operator H.*

Proof. For the stability of the set of convex continuous value functions, take any convex continuous value function V. Then, from lemmas 5.2.1 and 5.3.3, V satisfies the property (5.18). So, from lemma (5.3.3), HV satisfies this property as well, i.e., HV is convex continuous over the belief space $\Delta(Z)$.

For the stability of the set of convex δ-Lipschitz continuous value functions, assume that the value function V is δ-Lipschitz continuous over the belief space $\Delta(Z)$ and consider the set Γ verifying lemma (5.2.1). All linear value function α in Γ is δ-Lipschitz continuous. Hence, from lemma 5.2.1, the value backup function HV is also δ-Lipschitz continuous. □

5.3.2 Properties of the Value Backup Operator

We now look at specific properties of the value backup operator HV in our partially observable stochastic zero-sum game.

Lemma 5.3.4. *Let $V, W : \Delta(Z) \longrightarrow \mathbb{R}$ be two convex continuous functions, $b \in \Delta(Z)$ be a belief such that $[HV](b) \leqslant [HW](b)$. Denote by π^V and π^W Nash Equilibria of each stage game with belief b and values V and W respectively. If there exists a real non-negative number $C \geqslant 0$ such that $W\left(\tau\left(b|a_1, \pi_2^V, o\right)\right) - V\left(\tau\left(b|a_1, \pi_2^V, o\right)\right) \leqslant C$ holds whenever $\pi_1^W(a_1) \neq 0$, then*

$$[HW](b) - [HV](b) \leqslant \gamma C.$$

In other words, suppose that a period begins with defender's belief b over the network state, and it is not clear whether the value of the next period is V or W. However, the attacker will play her NE decision of the stage game corresponding to V. If, for any action supported by the defender's NE decision of the stage game corresponding to W, the two possible values of the subsequent games do not differ to more than C, then the current two possible values do not deffer to more than γC.

Proof. By unilaterally deviating from an NE of any stage game, a player decreases his reward by definition of a NE. Therefore, for the defender and the attacker, respectively:

$$U^V_{\pi_1^W, \pi_2^V}(b) \leqslant [HV](b) \quad \text{and} \quad [HW](b) \leqslant U^W_{\pi_1^W, \pi_2^V}(b).$$

Then we have:

$$\begin{aligned}[HW](b) - [HV](b) \;\leqslant\; & \gamma \sum_{a_1 \in A_1} \sum_{o \in O} \pi_1^W(a_1)\, \mathbb{P}^b_{\pi_2^V}(o \,|\, a_1) \times \\ & \times \Big(W\left(\tau\left(b|a_1, \pi_2^V, o\right)\right) - V\left(\tau\left(b|a_1, \pi_2^V, o\right)\right) \Big)\end{aligned}$$

$$\leqslant \gamma \sum_{a_1 \in A_1} \sum_{o \in O} \pi_1^W(a_1)\, \mathbb{P}^b_{\pi_2^V}(o \,|\, a_1)\, C = \gamma C.$$

□

Based on the previous lemma, the following theorem is the main result of our theoretical analysis of the value backup operator. This theorem shows that the value backup operator H is a γ-contracting mapping.

Theorem 5.3.2. *The value backup operator H is γ-Lipschitz continuous in the space of convex continuous value functions under the max-norm: $\|V\|_\infty = \max_{b \in \Delta(Z)} \|V(b)\|$. Therefore, if the discounted factor γ is strictly less than 1, the value backup operator H is a contraction mapping.*

Proof. Take two convex and continuous value functions $V, W \in \mathbb{V}$ and denote $C = \|V - W\|_\infty$. By definition, the inequality $|W(b) - V(b)| \leqslant C$ holds for all beliefs b, so for their updates $\tau\left(b \middle| a_1, \pi_2^V, o\right)$ as well, provided $\pi_1^W(a_1) \neq 0$. Then, from lemma 5.3.4, $[HW](b) - [HV](b) \leqslant \gamma C$ for all beliefs, i.e., $\|HV - HW\|_\infty \leqslant \gamma C = \gamma \|V - W\|_\infty$. □

Henceforth, if $\gamma < 1$, from the Banach fixed point theorem, the value backup operator H admits a fixed point V^*. Any sequence $(V_n)_{n \in \mathbb{N}^*}$ of convex continuous functions such that $V_{n+1} = HV_n$ for every n converges to V^*. We prove that the fixed point of H is the optimal value of the game. And, as stated in the following proposition, any such sequence of value functions converges to the optimum value function.

Proposition 5.3.1. *The optimal value function V^* is a stable point for the value backup operator H.*

Proof. From the definition (equation (3.9)), the optimal value function V^* is the point wise supremum over the set $\Gamma^* = \{\mathrm{val}_{\sigma_1} : \sigma_1 \in \Sigma_1\}$ of values associated with defender strategies.

Take any $\overline{\alpha} \in (\Gamma^*)^{A_1 \times O}$ and suppose that for all $(a_1, o) \in A_1 \times O$, the linear function $\alpha^{a_1,o}$ over $\Delta(Z)$ is the value associated with some defender strategy $\sigma_{\overline{\alpha}}^{a_1,o}$. Denote $\overline{\sigma_{\overline{\alpha}}} = \left(\sigma_{\overline{\alpha}}^{a_1,o}\right)_{(a_1,o) \in A_1 \times O}$ the sequence of these defender strategies, and for each $\pi_1 \in \Pi_1$ consider the defender strategy $\mathrm{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})$ that coincides with π_1 at the first period, and consists in the $\alpha^{a_1,o}$ strategy corresponding to his first action-observation, i.e., for all history $\hbar_1$ for the defender,

$$\mathrm{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})(\hbar_1) = \begin{cases} \pi_1 & \text{if } \hbar_1 = \emptyset \\ \sigma_{\overline{\alpha}}^{a_1,o}(\hbar_1') & \text{if } \hbar_1 = (a_1, o, \hbar_1') \end{cases}. \tag{5.20}$$

Assume that at period 1, the defender has the belief b and plays the strategy $\mathrm{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})$, and the attacker plays strategy π_2. The defender is rewarded $\sum_{(a,z,z') \in A \times Z} b(z)\, \pi(a|z)\, \mathcal{R}(z, a)$ immediately. For the subsequent game, if his effective move in the first period was a_1 and he made observation o, then the subsequent reward is $\alpha^{a_1,o}(z')$. Hence, the defender's

reward at this sub-game is $\sum_{(a,z,z')\in A\times Z^2} T(z',o|z,a)\, b(z)\, \pi(a|z)\, \alpha^{a_1,o}(z')$, and the total reward associated with strategy π_2 is $u_\pi(\overline{\alpha}, b)$. Since the attacker aims to minimize this reward, the value of the strategy $\text{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})$ in belief b is equal to $\min_{\pi_2\in\Pi_2} u_{\pi_1,\pi_2}(\overline{\alpha}, b)$, i.e., the value of the defender strategy $\text{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})$ is exactly $V^{SG}_{\pi_1,\overline{\alpha}}$. Clearly, each $\text{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})$ is some strategy σ_1. Conversely, by admitting constant $\overline{\alpha}$'s in (a_1, o), it appears that all defender strategy is a $\text{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})$. Thus, the set of $\text{comp}(\pi_1, \overline{\sigma_{\overline{\alpha}}})$'s is Γ^*, and V^*, which is the point-wise supremum over Γ^*, is the point-wise supremum of the $V^{SG}_{\pi_1,\overline{\alpha}}$'s. More precisely: $V^*(b) = \sup_{(\pi_1,\overline{\alpha})\in\Pi_1\times(\Gamma^*)^{A_1\times O}} \text{val}_{\text{comp}(\pi_1,\overline{\sigma_{\overline{\alpha}}})}(b) = \sup_{(\pi_1,\overline{\alpha})\in\Pi_1\times(\Gamma^*)^{A_1\times O}} V^{SG}_{\pi_1,\overline{\alpha}}(b) = [HV](b)$, where, from equation (5.19), $V(b) = \sup_{\alpha\in\Gamma^*} \alpha(b) = V^*(b)$. □

5.3.3 Computation of the Backup Value

For the sake of scalability, let us restrict the computation of the value backup on pointwise linear and convex (PWLC) functions over $\Delta(Z)$, i.e., point-wise suprema over finite sets of linear functions. After a linear program, we prove that the resulting backup values remain PWLC. In fact, for any PWLC function V over the belief space, consider the finite set $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of linear functions over which V is the point-wise supremum. Denote $\Gamma = \left\{ \sum_{i=1}^n \gamma_i \alpha_i \;\middle|\; (\gamma_1,\ldots,\gamma_n) \in \mathbb{R}^n_+ \text{ and } \sum_{i=1}^n \gamma_i = 1 \right\}$ its convex hull. The function V is also the point-wise maximum over Γ. The set $\Gamma^{A_1\times O}$ is convex and the backup value is given by:

$$
\begin{aligned}
[HV](b) &= \max_{\pi_1\in\Pi_1} \sup_{\overline{\alpha}\in\Gamma^{A_1\times O}} \min_{\pi_2\in\Pi_2} u_{\pi_1,\pi_2}(\overline{\alpha}, b) = \max_{\pi_1\in\Pi_1} \max_{\overline{\alpha}\in\Gamma^{A_1\times O}} \min_{a_2\in A_2} u_{\pi_1,a_2}(\overline{\alpha}, b) \\
&= \max_{\substack{\pi_1\in\Pi_1 \\ \overline{\alpha}\in\Gamma^{A_1\times O}}} \min_{a_2\in A_2} \sum_{z\in Z} b(z) \left(\sum_{a_1\in A_1} \pi_1(a_1)\, \mathcal{R}(z,a) + \right. \\
&\qquad \left. + \gamma \sum_{(a_1,o,z')\in A_1\times O\times Z} T(z', o \,|\, z, a)\, \pi_1(a_1)\, \alpha^{a_1,o}(z') \right).
\end{aligned}
$$

The problem of computing the backup value $[HV](b)$ is equivalent to maximizing the sum $\sum_{z\in Z} b(z)\, W(z)$ over $\pi_1 \in \Pi_1$, $\overline{\alpha} \in \Gamma^{A_1\times O}$ and $W \in \mathbb{V}$ under the following condition for all $(a_2, z) \in A_2 \times Z$:

$$
W(z) \leqslant \sum_{a_1\in A_1} \pi_1(a_1)\, \mathcal{R}(z,a) + \gamma \sum_{(a_1,o,z')\in A_1\times O\times Z} T(z' \,|\, z, a)\, \pi_1(a_1)\, \alpha^{a_1,o}(z').
$$

Then, our problem is equivalent to a two-player zero-sum game with strategy sets $\Pi_1 \times \Gamma^{A_1\times O}$ and A_2, and the defender's reward is defined by the following max-min problem:

$$
\max_{\substack{\pi_1\in\Pi_1 \\ \overline{\alpha}\in\Gamma^{A_1\times O}}} \min_{a_2\in A_2} \sum_{z\in Z} b(z)\, W(z).
$$

Henceforth, $[HV](b)$ is the solution of the following optimization problem:

$$\max_{\left(\pi_1,\left(\overline{\lambda_i}\right)_{i=1,\ldots,n},W\right)\in\Pi_1\times\Gamma^{A_1\times O\times\{1,\ldots,n\}}\times\mathbb{V}} \sum_{z\in Z} b(z)\,W(z) \tag{5.21a}$$

subject to

$$W(z) \leqslant \sum_{a_1\in A_1} \pi_1(a_1)\,\mathcal{R}(z,a_1,a_2) + \\ + \gamma \sum_{(a_1,o,z')\in A_1\times O\times Z} T(z',o\,|\,z,a_1,a_2)\,\pi_1(a_1)\,\alpha^{a_1,o}(z') \quad \forall\,(z,a_2)\in Z\times A_2; \tag{5.21b}$$

$$\pi_1(a_1) \geqslant 0 \qquad \forall a_1\in A_1; \tag{5.21c}$$

$$\sum_{a_1\in A_1} \pi_1(a_1) = 1; \tag{5.21d}$$

$$\alpha^{a_1,o}(z') = \sum_{i=1}^{n} \lambda_i^{a_1,o}\alpha_i(z') \qquad \forall\,(a_1,o,z')\in A_1\times O\times Z; \tag{5.21e}$$

$$\sum_{i=1}^{n} \lambda_i^{a_1,o} = 1 \qquad \forall\,(a_1,o)\in A_1\times O; \tag{5.21f}$$

$$\lambda_i^{a_1,o} \geqslant 0 \qquad \forall\,(a_1,o,i)\in A_1\times O\times\{1,\ldots,n\}. \tag{5.21g}$$

This problem is not linear since it includes the products $\pi_1(a_1)\,\alpha^{a_1,o}(z')$, which involves the products $\pi_1(a_1)\,\lambda_i^{a_1,o}$ of unknown variables. To make the problem linear, the following substitution $\pi_1(a_1)\,\alpha^{a_1,o}(z') = \widehat{\alpha}^{a_1,o}(z')$ can be used. The resulting linear optimization problem becomes:

$$\max_{\left(\pi_1,\left(\widehat{\lambda}_i\right)_{i=1,\ldots,n},W\right)\in\Pi_1\times\Gamma^{A_1\times O\times\{i=1,\ldots,n\}}\times\mathbb{V}} \sum_{z\in Z} b(z)\,W(z) \tag{5.22a}$$

subject to

$$W(z) \leqslant \sum_{a_1\in A_1} \pi_1(a_1)\,\mathcal{R}(z,a_1,a_2) + \\ + \gamma \sum_{(a_1,o,z')\in A_1\times O\times Z} T(z',o|z,a_1,a_2)\,\widehat{\alpha}^{a_1,o}(z') \quad \forall\,(z,a_2)\in Z\times A_2; \tag{5.22b}$$

$$\pi_1(a_1) \geqslant 0 \qquad \forall a_1\in A_1; \tag{5.22c}$$

$$\sum_{a_1\in A_1} \pi_1(a_1) = 1; \tag{5.22d}$$

$$\widehat{\alpha}^{a_1,o}(z') = \sum_{i=1}^{n} \widehat{\lambda}_i^{a_1,o}\alpha_i(z') \qquad \forall\,(a_1,o,z')\in A_1\times O\times Z; \tag{5.22e}$$

$$\sum_{i=1}^{n} \widehat{\lambda}_i^{a_1,o} = \pi_1(a_1) \qquad \forall (a_1, o) \in A_1 \times O; \tag{5.22f}$$

$$\widehat{\lambda}_i^{a_1,o} \geqslant 0 \qquad \forall (a_1, o, i) \in A_1 \times O \times \{1, \ldots, n\}. \tag{5.22g}$$

This is a linear optimization problem within the space $\mathbb{R}^{n|A_1|^2|O||Z|}$ of vectors of the form $\left(\pi_1, \left(\widehat{\lambda}_i^{a_1,o}\right)_{i=1,\ldots,n}^{(a_1,o)\in(A_1,O)}, W\right)$. Constraints (5.22b–5.22g) do not depend on b, so the constraint set and consequently the set Q of its vertices do not depend on b as well. Meanwhile, the maximum for each belief b is achieved at a vertex $q \in Q$ and corresponds to the linear function β^q defined at pure beliefs by $\beta^q(z) = W(z)$. Finally, we obtain the following expression for the value back operator:

$$[HV](b) = \max_{q\in Q} \min_{a_2} u_{\pi_1^q, a_2}(\overline{\alpha}^q, b) = \max_{q\in Q} V^{SG}_{\pi_1^q, \overline{\alpha}^q}(b), \tag{5.23}$$

i.e., HV is the point-wise supremum over the finite subset $\{\beta^q \mid q \in Q\}$ of lin_Δ. Then, the following proposition follows, as the problem can be described using a linear program.

Proposition 5.3.2. *The subset $\mathbb{V}_{\mathrm{PWLC}}$ of $\mathbb{V}$ of PWLC functions is stable under H.*

The above program is obtained by considering that the defender is maximizing the function that is minimized by the attacker. The dual problem can be considered and yields a simple optimization problem. Let us consider the following reward:

$$\mathcal{R}_\pi^{\mathrm{subs}}(b, V) = \sum_{(a_1,o)\in A_1\times O} \sup_{\alpha\in\Gamma} \sum_{z'\in Z} \pi_1(a_1) \left(\sum_{\substack{(a_2,z)\in A_2\times Z \\ o=o(a,z,z')}} T(z' \mid z, a)\, b(z)\, \pi_2(a_2 \mid z) \right) \alpha(z').$$

We get that the utility is:

$$\begin{aligned} U^V_{a_1,\pi_2}(b) &= \sum_{(a_2,z)\in A_2\times Z} b(z)\, \pi_2(a_2 \mid z)\, \mathcal{R}(z, a) + \\ &\quad + \gamma \sum_{o\in O} \sup_{\alpha\in\Gamma} \sum_{z'\in Z} \left(\sum_{\substack{(a_2,z)\in A_2\times Z \\ o=o(a,z,z')}} T(z' \mid z, a)\, b(z)\, \pi_2(a_2 \mid z) \right) \alpha(z') \\ &= \sum_{(a_2,z)\in A_2\times Z} \pi_2'(a_2 \wedge z)\, \mathcal{R}(z, a) + \\ &\quad + \gamma \sum_{o\in O} \sup_{\alpha\in\Gamma} \sum_{z'\in Z} \left(\sum_{\substack{(a_2,z)\in A_2\times Z \\ o=o(a,z,z')}} T(z' \mid z, a)\, \pi_2'(a_2 \wedge z) \right) \alpha(z'), \end{aligned}$$

where $\pi_2'(z \wedge a_2) = b(z)\,\pi_2(a_2 \mid z)$. Since $[HV](b) = \min_{\pi_2 \in \Pi_2} \max_{\pi_1 \in \Pi_1} U^V_{\pi_1,\pi_2} = \min_{\pi_2 \in \Pi_2} \max_{a_1 \in A_1} U^V_{a_1,\pi_2}$, the problem consists of minimizing a number $W \geqslant \max_{a_1 \in A_1} U^V_{a_1,\pi_2}$. Then, $[HV](b)$ is the solution of the following linear program:

$$\min_{(\pi_2', W) \in [0,1]^{Z \times A_2} \times \mathbb{R}} W \tag{5.24a}$$

subject to

$$W \geqslant \sum_{(a_2,z) \in A_2 \times Z} \pi_2'(z \wedge a_2)\,\mathcal{R}(z, a_1, a_2) + \gamma \sum_{o \in O} \widehat{W}(a_1, o) \qquad \forall a_1 \in A_1 \tag{5.24b}$$

$$\widehat{W}(a_1, o) \geqslant \sum_{z' \in Z} \widehat{k}(a_1, \pi_2, o)(z')\,\alpha_i(z') \qquad \forall (a_1, o, i) \in A_1 \times O \times \{1, \dots, n\} \tag{5.24c}$$

$$\widehat{k}(a_1, \pi_2', o)(z') = \sum_{\substack{(a_2,z) \in A_2 \times Z \\ o = o(a,z,z')}} T(z' \mid z, a)\,\pi_2'(a_2 \wedge z) \quad \forall (a_1, \pi_2', o) \in A_1 \times [0,1]^{Z \times A_2} \times O \tag{5.24d}$$

$$\sum_{a_2 \in A_2} \pi_2'(z \wedge a_2) = b(z) \qquad \forall z \in Z \tag{5.24e}$$

$$\pi_2'(z \wedge a_2) \geqslant 0 \qquad \forall (z, a_2) \in Z \times A_2 \tag{5.24f}$$

5.3.4 Value Backup Iteration

The authors in [35] proposes a point-wise iteration through algorithm (algorithm 1) which

Algorithm 1: HSVI algorithm for Discounted OS-POSGs [35]

while $\overline{V}(b^0) - \underline{V}(b^0) > \varepsilon$ **do**
  `Explore`(b^0, 1)
procedure `Explore`(b^{t-1}, t)
  if $\text{excess}_t(b^{t-1}) > 0$ **then**
    $\pi_1^t \leftarrow$ optimal strategy of player 1 in $[H\overline{V}](b^{t-1})$
    $\pi_2^t \leftarrow$ optimal strategy of player 2 in $[H\underline{V}](b^{t-1})$
    $b^t \leftarrow \underset{(a_1,o) \in A_1 \times O}{\arg\max}\ \mathbb{P}_{b^{t-1},\pi_1^t,\pi_2^t}(a_1, o) \times \text{excess}_{t+1}\left(\tau\left(b^{t-1}, a_1, \pi_2^t, o\right)\right)$
    `Explore`(b^t, $t+1$)
    Perform point-based update of $\overline{V}$ and $\underline{V}$ in b^{t-1}

converges the optimal value V^* is the fixed point of the backup function H defined for all

function $V\colon \Delta(Z) \longrightarrow \Delta(Z)$ by

$$[HV](b) = \max_{\pi_1 \in \pi_1} \min_{\pi_2 \in \pi_2} \left(\mathbb{E}^b_{\pi_1,\pi_2} [\mathcal{R}(z, a_1, a_2)] + \right.$$
$$\left. + \gamma \sum_{(a_1,o) \in A_1 \times O} \mathbb{P}^b_{\pi_1,\pi_2}(a_1, o)\, V(\tau(b, a_1, \pi_2, o)) \right).$$

The backup value, HV, is the result of the computation of the NE (lines 5 and 6) of the stage game $\mathrm{SG}(b, V)$. Suppose that currently the belief of player 1 is b and the optimal value function of the sub-game that begins at next period is V. If players have to run decision rules π of their choices, then player 1 is immediately rewarded $\mathbb{E}^b_{\pi_1,\pi_2} [\mathcal{R}(z, a_1, a_2)]$ in expectation. Also, he will play some action a_1 and make some observation o at probability

$$\mathbb{P}^b_{\pi_1,\pi_2}(a_1, o) = \sum_{(z',z,a_2) \in Z \times Z \times A_2} T(z', o|z, a_1, a_2)\, b(z)\, \pi_2(a_2|z)$$

and update his belief to $\tau(b, a_1, \pi_2, o)$ defined by

$$\tau(b, a_1, \pi_2, o)(z') = \frac{1}{\mathbb{P}^b_{\pi_1,\pi_2}(a_1, o)} \sum_{(z,a_2) \in Z \times A_2} T(z', o|z, a_1, a_2)\, b(z)\, \pi_2(a_2|z).$$

In their algorithm, authors assume players 1 and 2 respectively and simultaneously performing over time a point-based update of upper and lower bounds $\overline{V}$ and $\underline{V}$ (line 9) of the optimal value V^* by the application of the backup operator. To control the propagation of errors on the beliefs, the authors prove that the algorithm reaches the optimal value at some period t at which the belief is b with precision ε when the excess defined by

$$\mathrm{excess}_t(b) = \overline{V}(b) - \underline{V}(b) - \rho(t)$$

does not exceed ε. The function ρ is a monotonically increasing and unbounded defined on integer, positive numbers.

5.4 Simulations with Random Strategies

Some simple strategy simulations for both players are presented in this section, that assume a confrontation between an attacker and a defender acting randomly. We consider an Erdős-Réyni random graph with 50 nodes and a parameter 0.3 (probability to activate each edge), and assume the defender's resources limited at 3 IPSs. See figure 5.8. Both players' strategy is a fully random strategy without history. Meaning that the attacker chooses randomly a susceptible device from an infected device uniformly, and the defender chooses randomly the edges to allocate IPSs uniformly over the possible edges (edges that connect two not resistant nodes). A single node, chosen randomly, is assumed infected at period 1 and all

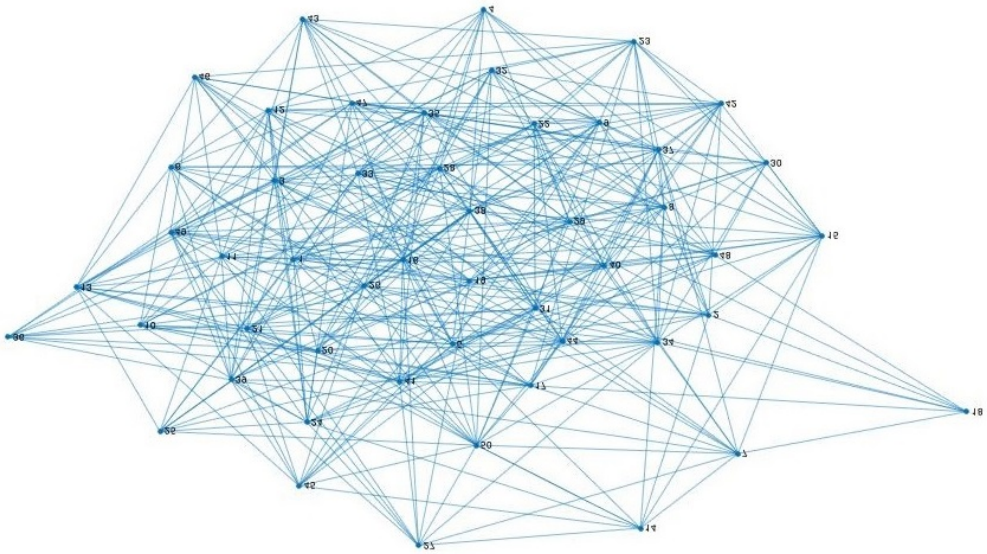


Figure 5.8: High connected graph

the other nodes are susceptible. We set the probability of transitions $I \to R$ to the value $\alpha = 0.5$ and the discount factor γ to $\rho = 0.99$, and we consider the following values of the parameters μ_i of the reward: $\begin{cases} \mu_3 = 1 \\ \mu_4 = 10 \\ \mu_0 = \mu_1 = \mu_2 = 0 \end{cases}$.

To begin, we observe the impact of the probability to change the default password on the defender's utility and also on the extinction time of the worm. Figure 5.9 shows that utility increases as the nodes are likely to become resistant.

Then, we set this probability of transitions to resistent state to $\rho = 0.1$. The simulations are depicted on figure 5.10. The number of IPSs h also has an important impact on these two performance measures. Other simulations, run 100 times with $\rho = 0.1$, show that the average utility goes from 83.03 with a 99% confidence interval [23.38, 142.68] to 385.40 with a 99% confidence interval [355.23, 415.58], when h goes from 3 to 10.

5.5 Conclusion

This chapter makes a census of all epidemic models with compartments S, I and R: 2 SI models and 14 SIR models. All epidemic models known so far are variants of these 16 models. For instance, many epidemics on networks, encompassing botnet spread, are SIR type, with transitions $S \to I$, $I \to S$ and $S \to R$. Nodes are not assumed intelligent and rational, so, to mitigate the propagation actively controlled by an attacker, we oppose a rational intelligent defender who will certainly play his optimal strategy. Both players are therefore involved in an end-impredictable scenario, and the transitions are influenced by the probabilistic decisions of network nodes. To solve the resulting POSSPG, we have proven

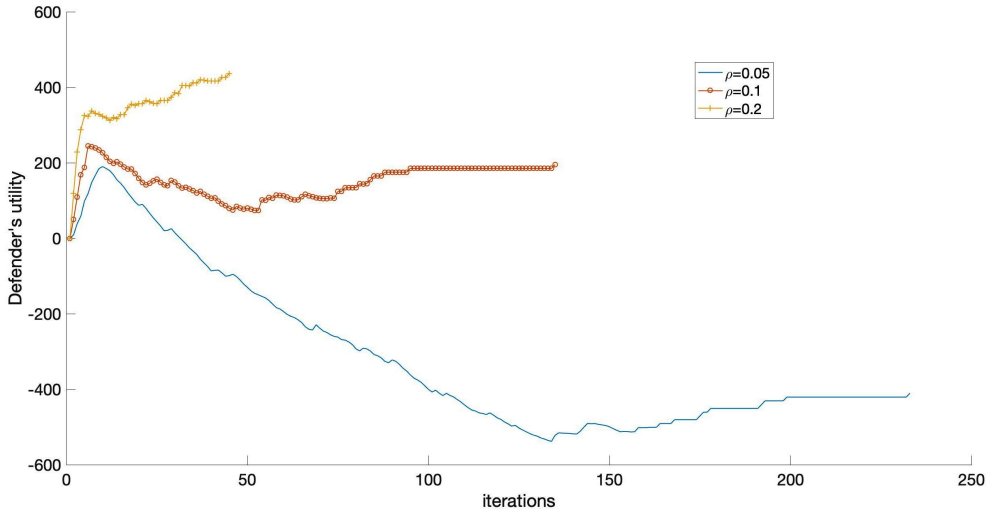


Figure 5.9: Influence of likelihood of becoming resistant
Output of simulations with $\gamma = 0.99$, $h = 3$, $\rho = 0.1$, $r_2 = 1$, $r_3 = 10$ *and* $\alpha = 0.5$.

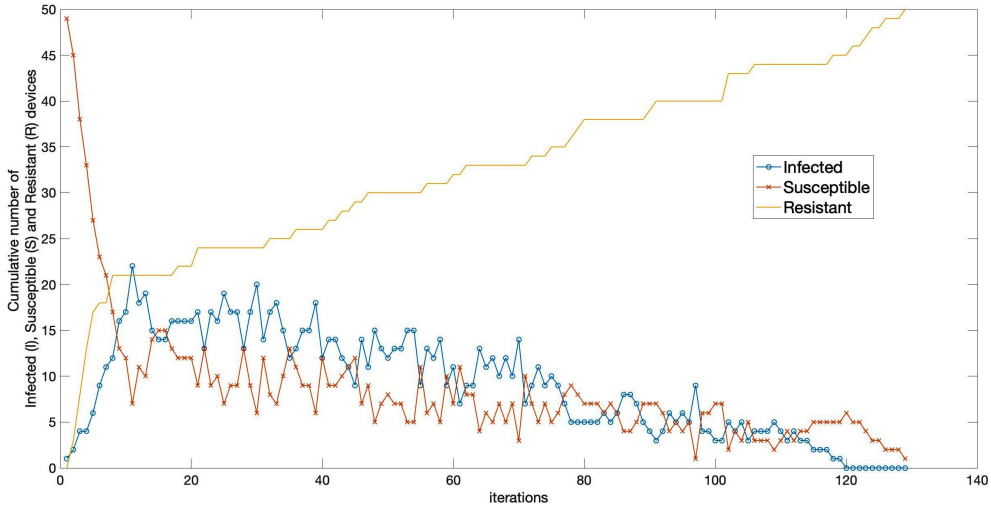


Figure 5.10: Evolution of each state categories
Output of simulations with $\gamma = 0.99$, $h = 3$, $\rho = 0.1$, $r_2 = 1$, $r_3 = 10$ *and* $\alpha = 0.5$.

that the VI algorithm holds even when the attacker cannot infer the actions of the defender. This algorithm, however, is not scalable even for less intricate lateral movement with two node states, S and I. To circumvent this issue, we propose (chapter 7) an equivalent game model that accounts the network topology. Another concern to rule out is the use of a discounted the sum instead of the sum. This is why we propose (chapter 6) another model

in which the utility is an extremum reward.

Chapter 6

Optimizing the IDPS Placement using a Partially Observable Stochastic Minimum-Threat Path Game

Contents

6.1 Introduction

In Markov decision processes as in stochastic games, the importance of payoff is captured by utility. Utility, viewed as a sum or discounted sum, has a computational advantage because of its interchangeability with mean. Indeed, the probabilistic nature of the transitions prevents calculations from being based on the exact rewards of the players, and thus requires that the exact values of the stage rewards be replaced by their mathematical expectations. In this way, the sum (be it discounted or not) of the mathematical expectation of the stage rewards is equal, by interchangeability, to the mathematical expectation of the sum of rewards, which is the utility. This is what guarantees the linearity of the utility function, then the meaning of the Bellman equation and, finally, the results established in the previous chapter. The author of [67] presents other utility functions in the framework of MDPs. However, since the epidemic is a threat to be minimized, this chapter does not consider it as an additive quantity, but rather expresses the threat as the attacker's reward. This being the case, the attacker's reward over the whole process is not the sum of his stage rewards, on the contrary, he will choose his strategy on the basis of the mathematical expectation of the maximum reward. Since the mathematical expectation of the maximum is not equal to the maximum of the mathematical expectations, this type of situation poses a still unsolved problem of MDPs. We solve this problem in this chapter.

In sum, the contribution of this threefold:

- We propose a new game model for adversarial epidemic control;
- We study extremum-utility function for two-player zero-sum POSGs, and we set an algorithm for optimal strategies;
- We prove the convergence of this algorithm in the context of adversarial SIR epidemic control. In fact, we prove this convergence for the more general case two-player zero-sum POSG where the utility for one player is the overall maximum of the rewards.

The rest of the chapter is organized as follows. In the next section, we prove that a POSMPG is equivalent to some POSSPG. That is, in Section 6.3, we derive the algorithm from the algorithm of POSSPG. Then, in Section 6.4, we present the optimal control of the active threat spread. We end with the conclusion.

6.2 Game Modification

Due to the non-interchangeability of the mean and the maximum, solving a POSG when the utility is the maximum remains a challenge. However, the case where utility is the sum of rewards is solved for the discounted sum and for POSSPGs. In order to take advantage of the resolution of this second type of POSGs, we extend the state of the system by adding a field that keeps in memory the player's maximum global reward 2, thus transforming the initial game into a POSSPG whose NE is that of the initial POSMPG $\mathcal{G} = \left(Z, A, T, r, b^0\right)$.

6.2.1 State-Extended Game

In the remainder of this thesis, we denote $\mathcal{M} = \{\mu(z'|z,a) : (z',z,a) \in Z \times Z \times A\}$ the set of possible rewards in game $\mathcal{G}$. For all application $\varphi\colon Z \longrightarrow \mathcal{M}$, [1] we call *state φ-extended game* or simply *extended game*, the POSSG $\widetilde{\mathcal{G}} = \left(\widetilde{Z}, O, A, \widetilde{T}, \widetilde{r}, \widetilde{b}^0, \widetilde{Z}_{\text{goals}}\right)$ defined as follows:

- The extended state space and set of goal states

$$\widetilde{Z} = Z \times \mathcal{M} \qquad \text{and} \qquad \widetilde{Z}_{\text{goals}} = Z_{\text{goals}} \times \mathcal{M}; \tag{6.1}$$

- The reward function due to the increase of memory and defined by $\widetilde{r} = -\widetilde{\mu}$, where $\widetilde{\mu}$ is defined for all (z,m), $(z',m') \in Z \times \mathcal{M}$ and $a \in A$ by:

$$\widetilde{\mu}(z',m'|z,m,a) = \begin{cases} 0 & \text{if } z \in Z_{\text{goals}} \\ \max(m'-m,0) & \text{otherwise} \end{cases}, \tag{6.2}$$

  where $\mu = -r$;

- The initial belief defined by:

$$\widetilde{b}^0(z,m) = \begin{cases} b^0(z) & \text{if } m = \varphi(z) \\ 0 & \text{otherwise} \end{cases}; \tag{6.3}$$

- The extended transition function defined for all $\widetilde{z} = (z,m)$ and $\widetilde{z}' = (z',m')$ with $z, z' \in S$, $m, m' \in M$, and for all $a \in A$ and $o \in O$ by:

$$\widetilde{T}(z',m',o|z,m,a) = \begin{cases} T(z',o|z,a) & \text{if } \begin{cases} \mu(z'|z,a) \leqslant m \\ m' = m \end{cases} \text{ or } \begin{cases} \mu(z'|z,a) > m \\ m' = \mu(z'|z,a) \end{cases} \\ 0 & \text{otherwise} \end{cases}. \tag{6.4}$$

A state of the system in the POSSPG corresponds to a state of the system in the POSMPG with a number that represents a memory of the last maximum reward attained by player 2. That is, at the end of each period, the memory is incremented with the eventual additional reward as the current reward is compared to the last maximum reward attained and this incremented reward is the reward for the POSSPG.

The following propositions respectively witness the straightforwardness and the importance of the above definition of the extended POSG.

Proposition 6.2.1. *(a) $\widetilde{Z}$ is a finite set.*

(b) For all $(\widetilde{z}, a) \in \widetilde{Z} \times A$, $\widetilde{T}(\cdot|\widetilde{z}, a)$ is a probability distribution on $\widetilde{Z} \times A$.

(c) All state in $\widetilde{Z}_{\text{goals}}$ is a goal state.

[1] In the case of epidemic for example, $\varphi(z)$ is the number of infected nodes in state z.

Proof. From $\mathcal{M}$ is the image set of the function μ defined on the finite set $S \times S \times A$, it comes that the sets $\mathcal{M}$ then $\widetilde{Z}$ are finite. The proof of (b) is the following: (1) the obvious non-negativity of T and (2): for all $\widetilde{z} = (z, m) \in \widetilde{Z}$ and all $a \in A$, we get:

$$\begin{aligned}
\sum_{(\widetilde{z}', o) \in \widetilde{Z} \times O} \widetilde{T}\left(\widetilde{z}', o | \widetilde{z}, a\right) &= \sum_{(z', m', o) \in Z \times \mathcal{M} \times O} \widetilde{T}\left(z', m', o | z, m, a\right) \\
&= \sum_{\substack{(z', o) \in Z \times O \\ \mu(z'|z,a) \leqslant m}} \widetilde{T}\left(z', m, o | z, m, a\right) + \sum_{\substack{(z', o) \in Z \times O \\ \mu(z'|z,a) > m}} \widetilde{T}\left(z', \mu\left(z' | z, a\right), o | z, m, a\right) \\
&= \sum_{\substack{(z', o) \in Z \times O \\ \mu(z'|z,a) \leqslant m}} T\left(z', o | z, a\right) + \sum_{\substack{(z', o) \in Z \times O \\ \mu(z'|z,a) > m}} T\left(z', o | z, a\right) = 1.
\end{aligned}$$

For the proof of (c), if $(z, m) \in \widetilde{Z}_{\text{goals}}$ and $m' \neq m$, from (6.4), $\widetilde{T}\left(z, m', o_{\text{reach}} | z, m, a\right) \neq 0$ only if $\mu\left(z | z, a\right) > m$. Since $\mu\left(z | z, a\right) = 0$, this last statement would contradict the non-negativeness of the values in $\mathcal{M}$. □

Furthermore, the state extension makes the POSMPG a particular POSSPG as it is shown in the following proposition.

Proposition 6.2.2. *If the path $\hbar = \left(z^t, a^t, o^t\right)_{t=1}^{\infty}$ is realized in game $\mathcal{G}$ and generates the output $(\mu_t)_{t=1}^{\infty}$ for player 2, then, by taking $\mu_0 = m_1 = \widetilde{\mu}_0$, the associated path $\widetilde{\hbar} = \left(\widetilde{z}^t, a^t, o^t\right)_{t=1}^{\infty}$ in game $\widetilde{\mathcal{G}}$ (with $\widetilde{z}_t = (z_t, m_t)$) and the resulting output $(\widetilde{\mu}_t)_{t=1}^{\infty}$ satisfy for all period $t \geqslant 1$:*

$$m_t = \sum_{n=0}^{t-1} \widetilde{\mu}_n = \max_{n=0,\ldots,t-1} \mu_n. \tag{6.5}$$

Proof. By induction. The equations clearly hold for $t = 1$.

From proposition 6.2.1.(c), if the equations hold until a goal state is reached, then it hold onward.

Suppose it holds for some $t \geqslant 1$ and no goal state is reached. Note that from (6.4), it comes:

$$m_{t+1} = \begin{cases} m_t & \text{if } \mu\left(z_{t+1} | z_t, a_t\right) \leqslant m_t \\ \mu\left(z_{t+1} | z_t, a_t\right) & \text{if } \mu\left(z_{t+1} | z_t, a_t\right) > m_t \end{cases}, \tag{6.6a}$$

then

$$m_{t+1} = \max\left(m_t, \mu_t\right). \tag{6.6b}$$

So, $m_{t+1} = \max_{n=0,\ldots,t} \mu_n$.

From 6.2, if $z_t \notin Z_{\text{goals}}$, then

$$\widetilde{\mu}_t = m_{t+1} - m_t. \tag{6.7}$$

In case $z_t \in Z_{\text{goals}}$, for all state $z' \in Z$ and action profile $a \in A$, we get $\mu\left(z' | z_t, a\right) = 0 \leqslant m$, which from (6.4) implies $m_{t+1} = m_t$. We also get $\mu_{t+1} = 0$. So, equation (6.7) holds in any case, and $m_{t+1} = m_t + \widetilde{\mu}_t = \sum_{n=0}^{t} \widetilde{\mu}_n$. □

Then, based on the previous result, we have that the supremum of rewards earned in the POSMPG is the sum of the rewards earned in the POSSPG and, at the same time, the value in memory.

6.2.2 Relation between the POSMPG and the POSSPG

We closely examine the relation between the two games, and we prove that their resolutions are equivalent.

Proposition 6.2.3. *Different POSMPGs extend to different POSSPGs.*

Proof. Suppose games $\mathcal{G} = \left(Z, O, A, T, \mu, b^0\right)$ and $\mathcal{G}' = \left(Z', O', A', T', \mu', b'^0\right)$ are equally state-extended to $\overline{\mathcal{G}} = \left(\overline{Z}, \overline{O}, \overline{A}, \overline{T}, \overline{\mu}, \overline{b}^0\right)$. Immediately, we get $O = \overline{O} = O'$ and $A = \overline{A} = A'$, and from equation (6.3) it comes $b^0 = \overline{b}^0 = b'^0$. So, from equation (6.1), we get $Z = Z'$ and the equality of image sets $\mathcal{M} = \{\mu(z'|z,a) : (z,z',a) \in Z \times Z \times A\}$ and $\mathcal{M}' = \{\mu'(z'|z,a) : (z,z',a) \in Z \times Z \times A\}$, which, coupled to equation (6.6b), imply the equality of the reward functions of both POSMPGs. From this equality and equation (6.4), we get the equality of the transition functions T and T'. □

Respectively denote $\widetilde{\mathsf{out}}_1$, $\widetilde{\mathsf{out}}_2$ and $\widetilde{\theta}$ the players 1 and 2 outputs and the history in game $\widetilde{\mathcal{G}}$. Consider the natural projection $s\colon (z,m) \longmapsto z$ of the state space $\widetilde{Z}$ of the POSSPG onto the state space Z of the POSMPG, that consists in relaxing the memory. This projection induces other projections, that we also denote s:

(i) A projection of the history set $\widetilde{H}$ in the POSSPG onto the corresponding history sets in the POSMPG, defined by $s\left(\left(z^t, m^t, a^t, o^t\right)_{t=1}^{n-1}, z^n, m^n\right) = \left(\left(z^t, a^t, o^t\right)_{t=1}^{n-1}, z^n\right)$;

(ii) A projection of the player 2 history set $\widetilde{H}_2$ in the POSSPG onto the corresponding history sets in the POSMPG, defined by $s\left(\left(z^t, m^t, a_2^t\right)_{t=1}^{n-1}, z^n, m^n\right) = \left(\left(z^t, a_2^t\right)_{t=1}^{n-1}, z^n\right)$.

Note that player 1 has the same history in both games. That is, each strategy σ_2 for player 2 in the POSMPG induces the strategy $\sigma_2 \circ s$. The converse is established in the following proposition:

Proposition 6.2.4. *For any period $n \geqslant 2$ and any two histories $\widetilde{h}_1 = \left(\left(z^t, m_1^t, a^t, o^t\right)_{t=1}^{n-1}, z^n, m_1^n\right)$ and $\widetilde{h}_2 = \left(\left(z^t, m_2^t, a^t, o^t\right)_{t=1}^{n-1}, z^n, m_2^n\right)$ in the POSSPG, if the equality $m_1^t = m_2^t$ holds for $t = 1$, then it holds for all period $t \in \{1, \ldots, n\}$.*

In other words, a history $\hbar$ in the POSSPG is uniquely determined by its projection $s(\hbar)$ onto the POSMPG and its first value in memory $m^1(\hbar)$. This is, the application $\overline{s}\colon \widetilde{\hbar} \longmapsto \left(s\left(\widetilde{\hbar}\right), m^1\left(\widetilde{\hbar}\right)\right)$ is an injection from $\widetilde{H}$ to H. It is a bijection by (6.6).

Proof. By induction. Equation (6.6a) implies that at time t where the system is in state z_t with memory m_t, if action a_t is taken and makes the system state transition to z_{t+1}, the new value m_{t+1} in memory is known. □

In a mater of consequence, since a strategy profile $\widetilde{\sigma}$ in the POSSPG maps a decision profile to any extended history, which is nothing but a couple $(\hbar, m)$ composed of a history $\hbar$, and its first value in memory $m = m^1(\hbar)$, for any m, the partial application $\widetilde{\sigma}(\cdot, m)$ is a strategy profile in the POSMPG. Note that $m^1 = \varphi\left(z^1\right)$. That is, one strategy profile in the POSMPG, note it $\varphi(\widetilde{\sigma}) = \widetilde{\sigma}(\cdot, m)$, is associated with the strategy profile $\widetilde{\sigma}$ of the POSSPG.

Theorem 6.2.1. *The two strategy profiles $\widetilde{\sigma}$ and $\varphi(\widetilde{\sigma})$ bring on the same utility value.*

Proof. Suppose the initial state in the POSMPG is z. The initial state is the POSSPG is $\widetilde{z} = (z, \varphi(z))$. The utility value of player 2 is:

$$\begin{aligned}
\widetilde{U}_{2,\widetilde{\sigma}}(z, \varphi(z)) &= \mathbb{E}_{\widetilde{\sigma}}\left(\sum_{t=1}^{\infty} \widetilde{\mu}_t \middle| z^{[1]} = z, m^{[1]} = \varphi(z)\right) \\
&= \mathbb{E}_{\widetilde{\sigma}}\left(\max_{t \geqslant 1} \mu_t \middle| z^{[1]} = z, m^{[1]} = \varphi(z)\right) \\
&= U_{2,\sigma}(z).
\end{aligned}$$

□

The operator φ is a bijection from the strategies of the POSSPG to the strategies of the POSMPG. That is, the two games are played concurrently with equivalent outputs for equivalent strategies and therefore same actions.

An NE $\widetilde{\sigma}^*$ of the POSSPG exists and is associated with the optimal value function $V^* = U_{\widetilde{\sigma}}$ defined by $V^*(b) = \max_{\widetilde{\sigma}_1} \min_{\widetilde{\sigma}_2} U_{(\widetilde{\sigma}_1, \widetilde{\sigma}_2)}(b)$. From the above theorem, V^* is also the optimal value function of the POSMPG, and $V^* = U_{\varphi(\widetilde{\sigma})}$. Therefore, the resolution of the state POSSPG is mathematically equivalent to the resolution of the POSMPG. We choose the computational resolution of the first one. Each player will play in the POSMPG the NE strategy corresponding to the POSSPG NE.

6.3 Computing the NE of a POSMPG

6.3.1 Solving POSSPGs

An algorithm for POSSPGs is provided in [85] with the assumption that the reward of player 2 is strictly positive until a goal state is reached. We prove that this assumption can be relaxed when the POSSPG is the state-extension of game in which the utility is the maximum. In the remaining of this section, we note

$$\widetilde{R}(\widetilde{z}, a) = \sum_{(\widetilde{z}', o) \in \widetilde{Z} \times O} \widetilde{T}(\widetilde{z}', o | \widetilde{z}, a) \times \widetilde{r}(\widetilde{z}' | \widetilde{z}, a). \tag{6.8}$$

In [85] the authors propose an algorithm that converges only when the instantaneous reward admits some negative bound before a goal state is reached, i.e., for some $\widetilde{R}_{\max}$, the inequality $\widetilde{R}(\widetilde{z}, a) \leqslant \widetilde{R}_{\max}$ holds for all action profile $a \in A$ taken in all non-goal state $z \in Z \setminus Z_{\text{goals}}$. For all period $k \geqslant 1$, they define the *k-cutoff game* as the k-period prefix of the POSSPG with the recommendation that player 1 is forced to play uniform strategy $\widetilde{\sigma}_{\text{unif}}$ from period $k+1$ onward. The utility of a strategy $\widetilde{\sigma}$ is defined by

$$U_{\widetilde{\sigma}}^{k-}\left(\widetilde{b}\right) = \mathbb{E}_{\widetilde{b}}^{\widetilde{\sigma}}\left(\sum_{t=1}^{k} \widetilde{R}\left(\widetilde{z}^{[t]}, a^{[t]}\right) + \text{val}_{\widetilde{\sigma}_{\text{unif}}}\left(\widetilde{b}^{[k]}\right)\right), \tag{6.9}$$

where $\widetilde{b}^{[k]}$ is the result of the successive updates of player 1's belief at the end of each of the k first periods. Authors prove the following results in the general case and use the strict negativity of the reward of player 1 while no goal state is reached to bound $V^* - V^{*,k-}$.

Theorem 6.3.1 ([85]). *For arbitrary period $k \geqslant 1$, it holds:* $\text{val}_{\widetilde{\sigma}_{\text{unif}}} \leqslant V^{*,k-} \leqslant V^*$.

Keeping in mind that V^* is bounded, consider the *k-horizon limited game* of the POSSPG, for which the utility of any strategy $\widetilde{\sigma}$ is defined by

$$U_{\widetilde{\sigma}}^{k+}\left(\widetilde{b}\right) = \mathbb{E}_{\widetilde{\sigma}}^{\widetilde{b}}\left(\sum_{t=1}^{k} \widetilde{R}\left(\widetilde{z}^{[t]}, a^{[t]}\right)\right). \tag{6.10}$$

We then get the following result:

Theorem 6.3.2. *For an arbitrary period $k \geqslant 1$, it holds: $V^* \leqslant V^{*,k+}$.*

Proof. Take any belief $\widetilde{b}$ and consider any strategy profile $\widetilde{\sigma}$ of the infinite horizon POSSPG and its restriction, also noted $\widetilde{\sigma}$ on the set of histories before period $k+1$. The negativity of the function $\widetilde{R}$ implies $U_{\widetilde{\sigma}}\left(\widetilde{b}\right) \leqslant U^{\widetilde{\sigma},k+}\left(\widetilde{b}\right)$, then $U_{\widetilde{\sigma}}\left(\widetilde{b}\right) \leqslant V^{*,k+}\left(\widetilde{b}\right)$. So, for the NE $\widetilde{\sigma}^*$, it comes $V^*\left(\widetilde{b}\right) = U_{\widetilde{\sigma}^*}\left(\widetilde{b}\right) \leqslant V^{*,k+}\left(\widetilde{b}\right)$. □

This result yields the boundedness $V^{*,k-} \leqslant V^* \leqslant V^{*,k+}$.

6.3.2 Algorithm for the Extended Game

Algorithm for the POSSPG with Discounted Sum

The algorithm for the resolution of the extremum-utility POSSPGs is adapted from the algorithm for the resolution of sum-utility POSSPGs proposed in [35]. This algorithm is adapted from the algorithm designed for discounted sum-utility POSGs (algorithm 2), with discount factor any $\gamma \in (0,1)$, which approaches the optimal value V^* as the fixed point of the backup function H defined for all function $V\colon \Delta\left(\widetilde{Z}\right) \longrightarrow \Delta\left(\widetilde{Z}\right)$ by

$$[HV]\left(\widetilde{b}\right) = \max_{\widetilde{\pi}_1 \in \widetilde{\Pi}_1} \min_{\widetilde{\pi}_2 \in \widetilde{\Pi}_2} \left(\mathbb{E}_{\widetilde{b},\widetilde{\pi}_1,\widetilde{\pi}_2}\left[\widetilde{R}(z, a_1, a_2)\right] +\right.$$

$$+\gamma\times\sum_{(a_1,o)\in A_1\times O}\mathbb{P}_{\widetilde{b},\widetilde{\pi}_1,\widetilde{\pi}_2}\left[a_1,o\right]\times V\left(\tau\left(\widetilde{b},a_1,\widetilde{\pi}_2,o\right)\right)\Bigg).$$

The backup value, HV, is the result of the computation of the NE (lines 5 and 6) of the

Algorithm 2: HSVI algorithm for Discounted OS-POSGs [85]

Data: b^0, φ, $\overline{V}=\overline{V}$, $\underline{V}=\underline{V}$
Result: $\left(\widetilde{\pi}^t\right)_{t\geqslant 1}$

1 **while** $\overline{V}\left(\widetilde{b}^0\right)-\underline{V}\left(\widetilde{b}^0\right)>\varepsilon$ **do**
2 `Explore(`$\widetilde{b}^0$`, 1)`
3 **procedure** `Explore(`$\widetilde{b}^{t-1}$`, `t`)`
4 **if** $\text{excess}_t\left(\widetilde{b}^{t-1}\right)>0$ **then**
5 $\widetilde{\pi}_1^t\leftarrow$ optimal strategy of player 1 in $\left[H\overline{V}\right]\left(\widetilde{b}^{t-1}\right)$
6 $\widetilde{\pi}_2^t\leftarrow$ optimal strategy of player 2 in $\left[H\underline{V}\right]\left(\widetilde{b}^{t-1}\right)$
7 $\widetilde{b}^t\leftarrow\underset{(a_1,o)\in A_1\times O}{\arg\max}\ \mathbb{P}_{\widetilde{b}^{t-1},\widetilde{\pi}_1^t,\widetilde{\pi}_2^t}\left(a_1,o\right)\times\text{excess}_{t+1}\left(\tau\left(\widetilde{b}^{t-1},a_1,\widetilde{\pi}_2^t,o\right)\right)$
8 `Explore(`$\widetilde{b}^t$`, `$t+1$`)`
9 Perform point-based update of $\overline{V}$ and $\underline{V}$ in $\widetilde{b}^{t-1}$

game termed *stage game* $\left[HV\right]\left(\widetilde{b}\right)$. Suppose that currently the belief of player 1 is $\widetilde{b}$ and the optimal value function of the sub-game that begins at next period is V. If players have to run decision rules $\widetilde{\pi}$ of their choices, then player 1 is immediately rewarded $\mathbb{E}_{\widetilde{b},\widetilde{\pi}_1,\widetilde{\pi}_2}\left[\widetilde{R}\left(z,a_1,a_2\right)\right]$ in expectation. Also, he will play some action a_1 and make some observation o with probability

$$\mathbb{P}_{\widetilde{b},\widetilde{\pi}_1,\widetilde{\pi}_2}\left[a_1,o\right]=\sum_{(\widetilde{z}',\widetilde{z},a_2)\in\widetilde{Z}\times\widetilde{Z}\times A_2}\widetilde{T}\left(\widetilde{z}',o|\widetilde{z},a_1,a_2\right)\times\widetilde{b}\left(\widetilde{z}\right)\times\widetilde{\pi}_2\left(a_2|\widetilde{z}\right)$$

and update his belief to $\tau\left(\widetilde{b},a_1,\widetilde{\pi}_2,o\right)$ defined by

$$\tau\left(\widetilde{b},a_1,\widetilde{\pi}_2,o\right)\left(\widetilde{z}'\right)=\frac{1}{\mathbb{P}_{\widetilde{b},\widetilde{\pi}_1,\widetilde{\pi}_2}\left[a_1,o\right]}\sum_{(\widetilde{z},a_2)\in\widetilde{Z}\times A_2}\widetilde{T}\left(\widetilde{z}',o|\widetilde{z},a_1,a_2\right)\times\widetilde{b}\left(\widetilde{z}\right)\times\widetilde{\pi}_2\left(a_2|\widetilde{z}\right).$$

Authors of [85] assume optimistic players 1 and 2 respectively and simultaneously performing over time a point-based update of upper and lower bounds $\overline{V}$ and $\underline{V}$ (line 9) of the optimal value V^* by the application of the backup operator. Player 2 starts with a point-wise maximum $\underline{V}\colon\widetilde{b}\longmapsto\max_{\alpha\in\Gamma}\left\langle\alpha,\widetilde{b}\right\rangle=\max_{\alpha\in\Gamma}\sum_{\widetilde{z}\in\widetilde{Z}}\alpha\left(\widetilde{z}\right)\times\widetilde{b}\left(\widetilde{z}\right)$ over some set Γ of linear functions α

termed α-vectors. The update of $\underline{V}$ adds an α-vector $\mathcal{L}\Gamma\left(\widetilde{b}\right)$ corresponding to an NE strategy of player 1 in the game $[H\underline{V}]\left(\widetilde{b}\right)$ and expands the set Γ of α-vectors to $\Gamma \cup \left\{[H\underline{V}]\left(\widetilde{b}\right)\right\}$. Player 1 starts from an upper bound $\overline{V}$ of a restriction of the optimal value V^* on a finite set Υ of beliefs. The update of $\overline{V}$ at some belief $\widetilde{b}''$ consists in (1) updating the definition of $\overline{V}$ by $\overline{V}\left(\widetilde{b}\right) \leftarrow \inf_{\widetilde{b}' \in \Upsilon}\left\{\overline{V}\left(\widetilde{b}'\right) + (U-L) \times \|\widetilde{b} - \widetilde{b}'\|_2\right\}$, where $U = \max_{(\widetilde{z},a)\in\widetilde{Z}\times A} \widetilde{R}(\widetilde{z},a)$, $L = \min_{(\widetilde{z},a)\in\widetilde{Z}\times A} \widetilde{R}(\widetilde{z},a)$ and $\|\cdot\|_2$ is the euclidean norm, and in (2) expanding Υ to $\Upsilon \cup \left\{\widetilde{b}''\right\}$.

The gap $\overline{V} - \underline{V}$ converges to zero. However, to control the propagation of errors on the beliefs, define the excess, by

$$\text{excess}_t\left(\widetilde{b}\right) = \overline{V}\left(\widetilde{b}\right) - \underline{V}\left(\widetilde{b}\right) - \rho(t)$$

where ρ is a function defined on integer, positive numbers by

$$\rho(1) = \varepsilon \qquad \rho(t+1) = \frac{\rho(t) - 2 \times \delta \times D}{\gamma},$$

$\delta = \dfrac{\max \widetilde{R}(\cdot) - \min \widetilde{R}(\cdot)}{2 \times (1-\gamma)}$ and D is any parameter that satisfy $0 < D < \dfrac{(1-\gamma)\times\varepsilon}{2\times\delta}$, i.e., ρ is monotonically increasing and unbounded. The algorithm reaches the optimal value at some period t when the excess does not exceed ε.

Algorithm for the POSSPG

The boundedness of the value of the POSSPG established in theorems 6.3.1 and 6.3.2 suggests that one can iteratively apply the value iteration algorithm on the upper and lower bounds of $V^{*,k+}$ and $V^{*,k-}$ to bound V^*. However, the algoritm 2 applies only for discounted infinite horizon POSGs while the k-cutoff and the k-horizon limited games are finite, and the reward is not discounted. To adapt this algorithm to a N-horizon POSG with not discounted sum $\mathcal{G}^{[N]} = \left(S, A, O, T, R, b^0\right)$, authors of [85] make the problem an infinite-horizon POSG with γ-discounted sum $\mathcal{G}^{\gamma} = \left(S^{\gamma}, A, O, T^{\gamma}, R^{\gamma}, b^0\right)$. For $\gamma \in (0,1)$, the game $\mathcal{G}^{\gamma}$ are defined as follows:

- Games $\mathcal{G}^{[N]}$ and $\mathcal{G}^{\gamma}$ have the action and observation sets;
- The set S_n^{γ} of possible states of $\mathcal{G}^{\gamma}$ when n periods remain to be played in $\mathcal{G}^{[N]}$, $n \in \{0, \dots N\}$ can be iteratively obtained as follows: [2]

$$\begin{cases} S_N^{\gamma} = \text{supp}\left(b^0\right) \times \{N\} = \left\{(s,N)\middle| s \in S \text{ and } b^0(s) > 0\right\} \\ S_n^{\gamma} = \left\{(s',n)\middle| \text{ for some } (s,n+1) \in S_{n+1}^{\gamma} \text{ the transition from } s \text{ to } s' \text{ is possible in } \mathcal{G}^{[N]}\right\} \end{cases}.$$

[2] One can take $S_N^{\gamma} = S \times \mathbb{N}$.

- The transition function in $\mathcal{G}^\gamma$ respects the transition probabilities in $\mathcal{G}^{[N]}$, i.e.:
  - $T^\gamma(o, s', n|s, n+1, a) = T(o, s'|s, a)$
  - $T^\gamma(\hat{o}, s, 0|s, 0, a) = 1$ for an arbitrary fixed observation $\hat{o}$, i.e., there is no effective transition in $\mathcal{G}^\gamma$ when it remains no period to play in $\mathcal{G}^{[N]}$;
- About the reward function: $\begin{cases} R^\gamma(s, n, a) = \dfrac{R(s, a)}{\gamma^{N-n}} & \text{if } n \in \{1, \dots, N\} \\ R^\gamma(s, 0, a) = 0 \end{cases}$.

Both games have the same value function. So, the value iteration algorithm applies for the k-cutoffs and the k-horizon limitations of the POSSPG.

Algoritm 3 leverages the boundedness of the value function of the POSSPG between the finite horizon games to iteratively approach the value of this first value function. Therein,

Algorithm 3: HSVI algorithm for the POSSPG

Data: b^0, φ, $\mathrm{val}_{\widetilde{\sigma}_{\mathrm{unif}}}$, V^{*0_+}
Result: $\left(\widetilde{\pi}^t\right)_{t \geqslant 1}$

1 $k \leftarrow 1$
2 $\underline{V}^{k_-} \leftarrow \mathrm{val}_{\widetilde{\sigma}_{\mathrm{unif}}}$
3 $\overline{V}^{k_+} \leftarrow V^{*0_+}$ ▷ the zero function
4 **while** $\overline{V}^{k_+}\left(\widetilde{b}^0\right) - \underline{V}^{k_-}\left(\widetilde{b}^0\right) > \varepsilon$ and $k \leqslant K$ **do**
5 `Explore(`$\widetilde{b}^0$`, 1)`
6 $k \leftarrow k + 1$
7 **procedure** `Explore(`$\widetilde{b}^{t-1}$`, t)`
8 **if** $\mathrm{excess}_t\left(\widetilde{b}^{t-1}\right) > 0$ **then**
9 $\widetilde{\pi}_1^t \leftarrow$ optimal strategy of player 1 in $\left[H\overline{V}^{(k-t)_+}\right]\left(\widetilde{b}^{t-1}\right)$
10 $\widetilde{\pi}_2^t \leftarrow$ optimal strategy of player 2 in $\left[H\underline{V}^{(k-t)_-}\right]\left(\widetilde{b}^{t-1}\right)$
11 $\widetilde{b}^t \leftarrow \underset{(a_1, o) \in A_1 \times O}{\arg\max}\ \mathbb{P}_{\widetilde{b}^{t-1}, \widetilde{\pi}_1^t, \widetilde{\pi}_2^t}(a_1, o) \times \mathrm{excess}_{t+1}\left(\tau\left(\widetilde{b}^{t-1}, a_1, \widetilde{\pi}_2^t, o\right)\right)$
12 `Explore(`$\widetilde{b}^t$`, t + 1)`
13 Perform point-based update of $\overline{V}^{(k-t)_+}$ and $\underline{V}^{(k-t)_-}$ in $\widetilde{b}^{t-1}$

players 1 and 2 respectively play the successive cutoff and limited games (lines 9, 10 and 13). When a precision ε and a number K of periods are given, the algorithm terminates at the K-th period, unless the precision ε is reached before (line 4). The lower and upper bounds are respectively initialized in lines 2 and 3 at the value of uniform strategy of player 1 and the zero function.

We cannot guarantee the convergence of algoritm 3 but at least we can state the monotonicity of the value of the uniform strategy and deduce that the algorithm does not take us away from the solution.

Theorem 6.3.3. *For all belief $\widetilde{b}$ in the POSSPG, the sequence $\left(\mathrm{val}_{\widetilde{\sigma}_{\mathrm{unif}}}\left(\widetilde{b}^{[n]}\right)\right)_{n=0}^{\infty}$ increases in n.*

So, either the sequence is stationary or it converges.

Proof. Recall $b^{[0]} = b$. Take $\mathrm{lowermem}_n = \min\left\{m \,\middle|\, (z,m) \in \mathrm{supp}\left(\widetilde{b}^{[n]}\right) \text{ for some } z \in Z\right\}$, i.e., $\mathrm{lowermem}_n$ is the minimum possible value in memory at period n. Player 1 knows that the memory cannot decrease, so, $\mathrm{lowermem}_{n+1} \geqslant \mathrm{lowermem}_n$ for all n. □

In case $\left(\mathrm{val}_{\widetilde{\sigma}_{\mathrm{unif}}}\left(\widetilde{b}^{[n]}\right)\right)_{n=0}^{\infty}$ is constant from some period N onward, the uniform strategy achieves the maximum expected value of $\lim_{n\to\infty} m^n$ and is therefore the optimal strategy for player 1 from period N onward.

6.4 Algorithm for Defender optimal Strategy

We extend the POSMG with the above function φ. So, the number of infected nodes at the beginning of the game is the first value in memory. The game needs not reach the goal state for the algorithm to terminate. Indeed, it is possible to check out if the uniform strategy is stationary and from the underlined period onward, the defender should run this strategy. Note that the defender might be not aware that his strategy $\widetilde{\pi}_1$ is already stationary. So, the algorithm terminates anyway. This verification is made in line 4 of algorithm 3 through function *ShouldContinue* (see algorithm 4). To know if the value cannot be improved, the algorithm checks if the POSMPG has returned to visited state while the maximum number of infected nodes has not been improved (lines 9 to 18). From the finitude of the state space, it comes that the algorithm converges. In case needed to force the algorithm to terminate before the convergence, we keep the control $k \leqslant K$.

6.5 Smart Defense Strategies against Random Attack

Some simple strategy simulations for both players are presented in this section. The simplest ones assume a confrontation between an attacker and a defender acting randomly, while in the others at least one player thinks intelligently. We report here results of simulations we have conducted in order to compare some offensive strategies.

6.5.1 Description

Our stochastic game is a static game, in the sense that both players, attacker and defender, determine respectively at the beginning of the game which attack and which deception

Algorithm 4: HSVI algorithm for Adversarial Epidemic Control

Data: b^0, $\mathrm{val}_{\widetilde{\sigma}_{\mathrm{unif}}}$, V^{*0_+}
Result: $\left(\widetilde{\pi}^t\right)_{t\geqslant 1}$

$k \leftarrow 1$
$\underline{V}^{k_-} \leftarrow \mathrm{val}^{\widetilde{\sigma}_{\mathrm{unif}}}$
$\overline{V}^{k_+} \leftarrow V^{*0_+}$ ▷ the zero function
StuckStates $\leftarrow \emptyset$ ▷ in the POSMG
CheckMemory $\leftarrow -1$
while $\overline{V}^{k_+}\left(\widetilde{b}^0\right) - \underline{V}^{k_-}\left(\widetilde{b}^0\right) > \varepsilon$ *and* $k \leqslant K$ *and* `ShouldContinue()` **do**
  `Explore`$(\widetilde{b}^0, 1)$
  $k \leftarrow k+1$
Function `ShouldContinue()`
  if new maximum number of infected nodes = CheckMemory **then**
    if state ∈ StuckStates **then**
      return false
    else
      StuckStates ← StuckStates ∪ {state}
      return true
  else
    CheckMemory ← ∅
    return true
procedure `Explore`$(\widetilde{b}^{t-1}, t)$
  if $\mathrm{excess}_t\left(\widetilde{b}^{t-1}\right) > 0$ **then**
    $\widetilde{\pi}_1^t \leftarrow$ optimal strategy of player 1 in $\left[H\overline{V}^{(k-t)_+}\right]\left(\widetilde{b}^{t-1}\right)$
    $\widetilde{\pi}_2^t \leftarrow$ optimal strategy of player 2 in $\left[H\underline{V}^{(k-t)_-}\right]\left(\widetilde{b}^{t-1}\right)$
    $\widetilde{b}^t \leftarrow \underset{(a_1,o)\in A_1\times O}{\arg\max}\ \mathbb{P}_{\widetilde{b}^{t-1},\widetilde{\pi}_1^t,\widetilde{\pi}_2^t}(a_1,o) \times \mathrm{excess}_{t+1}\left(\tau\left(\widetilde{b}^{t-1}, a_1, \widetilde{\pi}_2^t, o\right)\right)$
    `Explore`$(\widetilde{b}^t, t+1)$
    Perform point-based update of $\overline{V}^{(k-t)_+}$ and $\underline{V}^{(k-t)_-}$ in $\widetilde{b}^{t-1}$

strategy to play. Then, the botnet propagation process is simulated considering these two strategies, and two performance metrics are evaluated in order to illustrate the impact of these strategies onto the system. These metrics are : the maximum proportion of infected devices and the botnet time to extinction. The latter is defined as the smallest period such that the number of infected devices is equal to 0, i.e. $\min_{t\geqslant 0}\{|I_t| = 0\}$. Several IoT network topologies are considered for simulations and several types of strategies for both

players. In order to illustrate performances of simple cyber deception techniques against epidemic botnet, our case studies are focused on random and smart deception strategies. Moreover, the attacker's strategy is restricted to random choice of targeted devices. Based on these strategies of the attacker (targeted neighbor devices) and defender (IPSs localization), epidemic propagation is simulated using a discrete events framework that follows:

- given a current state z of the system, for each infected device a targeted number of susceptible neighbors is determined,
- IPSs edges are selected following the deception strategy used,
- intermediate state $a(z)$ is computed based on previous actions,
- based on user's probability reaction ρ and α the system reaches a new state z'.

Every simulation is performed 500 times, and we plot at each time the 90% confidence interval. In particular, based on [57], two extreme cases topologies with 50 devices and different connectivity level are used. Network topologies have an important impact on the botnet propagation process, as it is shown in the simulations. We consider the following probabilistic reaction : $\rho = 0.2$ and $\alpha = 0.1$. It means that 20% of the devices change their password and 10% of the infected devices accepts the patch. This number seems to be realistic in botnet propagation, as in order to accept the patch, an IoT's user has to be aware that his device is corrupted, which is not usually the case with botnets.

6.5.2 Defense Deception Strategies

We consider several deception strategies for the defender and show their performances in terms of maximum proportion of infected devices and time to extinction (TTE) of the epidemic. First, defender's strategy is a fully random strategy without history. The deception strategy is called the ***Randomized Deception Strategy (RDS)***. First, the defender determines totally randomly and uniformly the edges on which to allocate IPSs on the overall network topology. Second, the defender implements a more sophisticated strategy based on his observation. In particular, once the defender observes an attack trough an IPS edge, its strategy for the next period is to put an IPS on top of the infected device observed. Specifically, if an attack from an infected device i and a susceptible device j has been captured by an IPS at period t, then, at period $t + 1$, the device i is in the susceptible state and the defender puts an IPS on an edge between device i and one of his neighbor except device j. This strategy is inspired by the one-by-one neighboring contamination process employed by the botnet and noting that device j cannot be infected anymore from device i. A generalization of this deception defense strategy is to consider that the defender can place several IPSs on top of the previous infected device i. We denote by ***k-Smart Deception Strategy (k-SDS)*** the strategy which consists in placing k IPSs over the maximum h of available IPSs, on top of any trapped infected device (i.e. an edge that linked any previous infected device and a non-resistant one). The remaining $h-k$ IPSs are randomly positioning

on other edges of the network. Note that if the degree d of the trapped infected device is strictly smaller than h, the $h-d-1$ remaining IPSs are used to discover new infected devices randomly in the network.

6.5.3 Low Connected Network Topology

A sparsely connected topology has been generated as an Erdős-Réyni random graph with 50 devices and parameter 0.06 (probability to activate each edge). See figure 6.1. The average

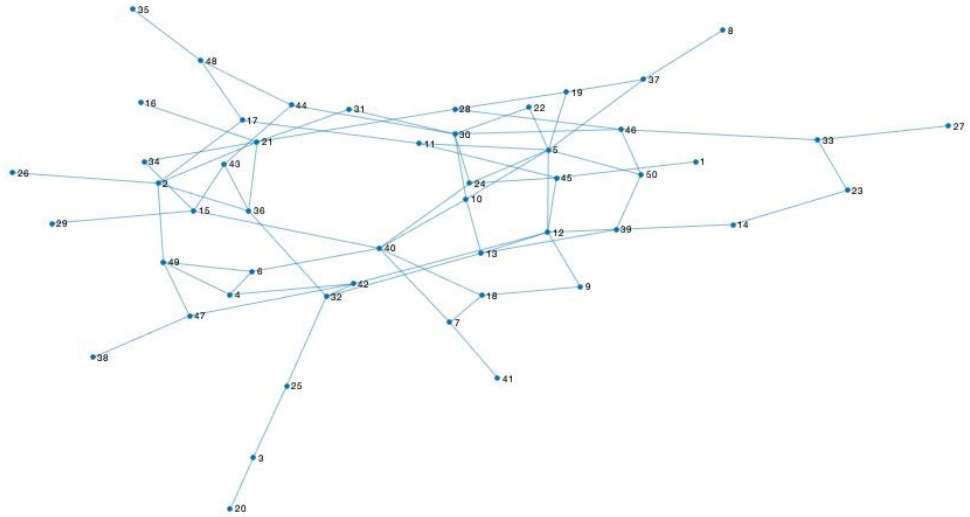


Figure 6.1: Sparsely connected graph

degree is 3 and the total number of edges is 75.

Cyber Attack Strategies

Figure 6.2 illustrates the performances in terms of maximum proportion of infected devices of the k-SDS deception mechanisms against three different cyber attack strategies: unicast, half and broadcast. These spreading strategies are defined by the proportion of susceptible neighbors infected by each infected node to which each node transmits the malware. The unicast strategy consists in making each infected node infect a single susceptible neighbor; the broadcast strategy consists in making each infected node infect all the susceptible neighbors; the intermediate half cast strategy consists in making each infected node infect half of the susceptible neighbors. As expected, the broadcast cyber attack yields the highest maximum proportion of infected devices in average.

Initial infected devices set

The initial set I_0 of infected devices has a strong influence on the performances of the k-SDS deception mechanisms related to the maximum proportion of infected devices. In fact, on

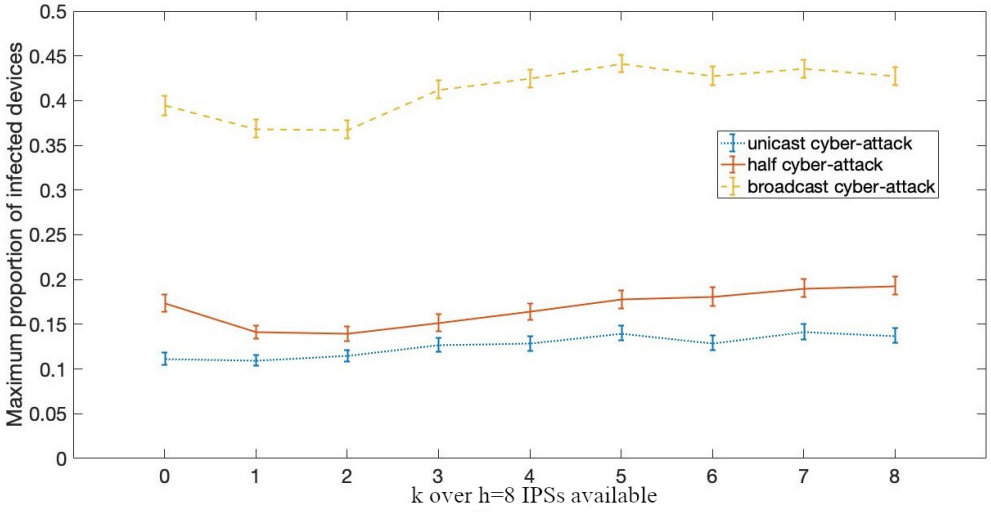


Figure 6.2: Maximum proportion of infected devices in the sparsely connected IoT network considering the k-SDS strategies with $k = 0, \ldots, 8$ and $h = 8$ available IPSs at each period and single initially infected device $I_0 = \{19\}$. *The cyber attack strategy has a big impact on the maximum proportion of infected devices.*

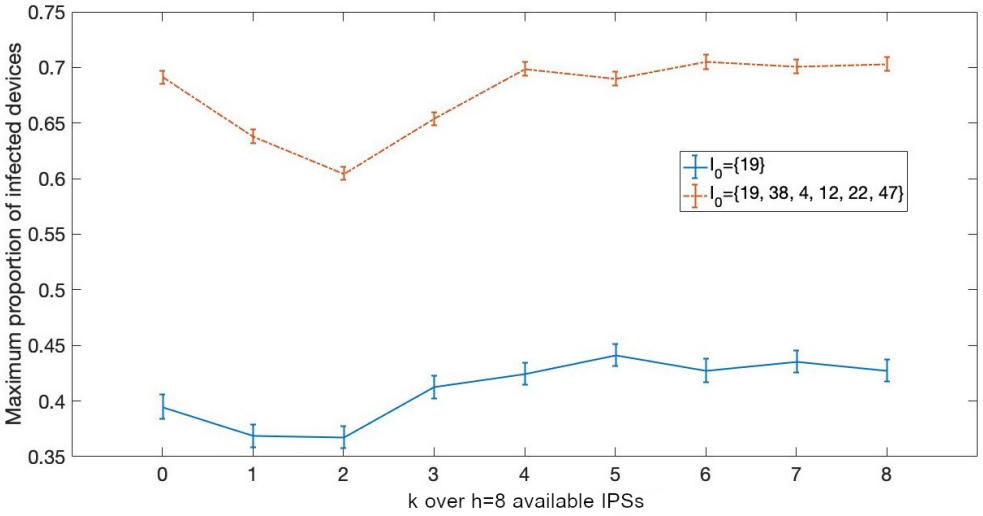


Figure 6.3: Maximum proportion of infected devices in the sparsely connected IoT network considering the k-SDS strategies with $k = 0, \ldots, 8$ and $h = 8$ available IPSs at each period and the broadcast cyber attack strategy. *The deception strategy used has no big impact compared to the initial infected devices set I_0.*

figure 6.3 we observe that the curves are almost flat depending on h but at different levels. In particular, it is between 35% and 45% when I_0 is a singleton, and between 60% and 70% for the other larger initial infected devices set. Note that the best deception strategy is the 2-SDS mechanism in this case.

However, the k-SDS mechanisms do not have the same performance in terms of botnet extinction time as illustrated on figure 6.4. Particularly, when the initial set of infected

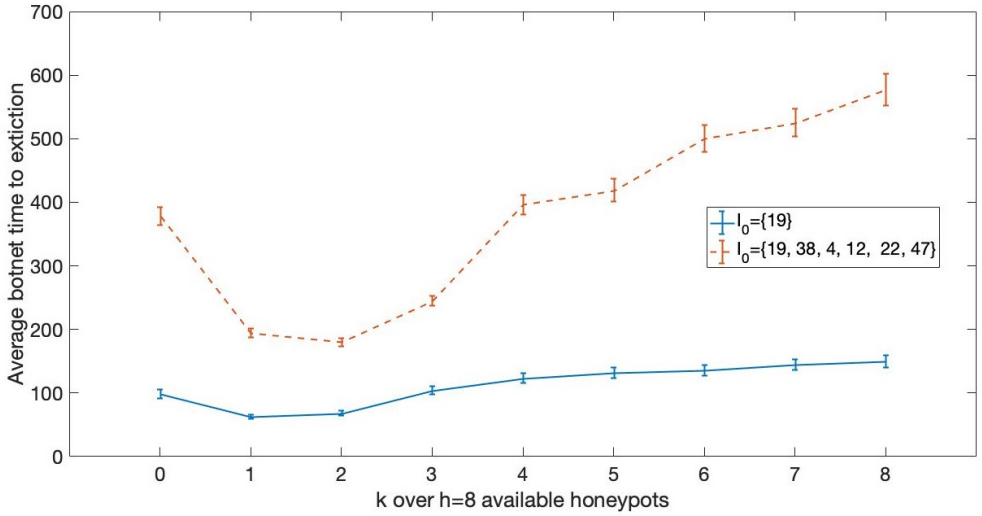


Figure 6.4: Time measured in periods to botnet extinction in the sparsely connected IoT network varying k over the $h = 8$ IPSs available at each period with the broadcast cyber attack strategy. *For this metric, the deception strategy used has an important effect when more than one device is initially infected:* 1*-SDS makes three times less periods to extinct the botnet compared to the* 7*-SDS mechanism, when the initial set of infected devices is larger.*

devices is larger, i.e. $I_0 = \{19, 38, 4, 12, 22, 47\}$, the performance ratio between the 2-SDS (best one) and 8-SDS (worst one) is around 3, meaning that it takes 3 times more iterations in order to extinct the botnet. This observation means that the exploration usage of IPSs is important in this case. In fact, IPSs are used to cure infected devices but also to discover new infected devices in the network. Then, it is not optimal to put all IPSs around infected devices.

6.5.4 High Connected Network Topology

Another Erdős-Réyni random graph with 50 devices and parameter 0.3 (probability to activate each edge) has been generated. The average degree is 15.36 and the network is composed of 384 edges.

cyber attack strategies

A very interesting observation is that cyber attack strategies do not have the same impact on the maximum proportion of infected devices in average, and specifically the broadcast cyber attack does not lead to the higher value. Precisely (figure 6.6), the half cyber attack strategy, which consists in choosing to propagate the botnet to half of the susceptible device neighbors from each infected ones, yields to 60% of maximum proportion of infected devices

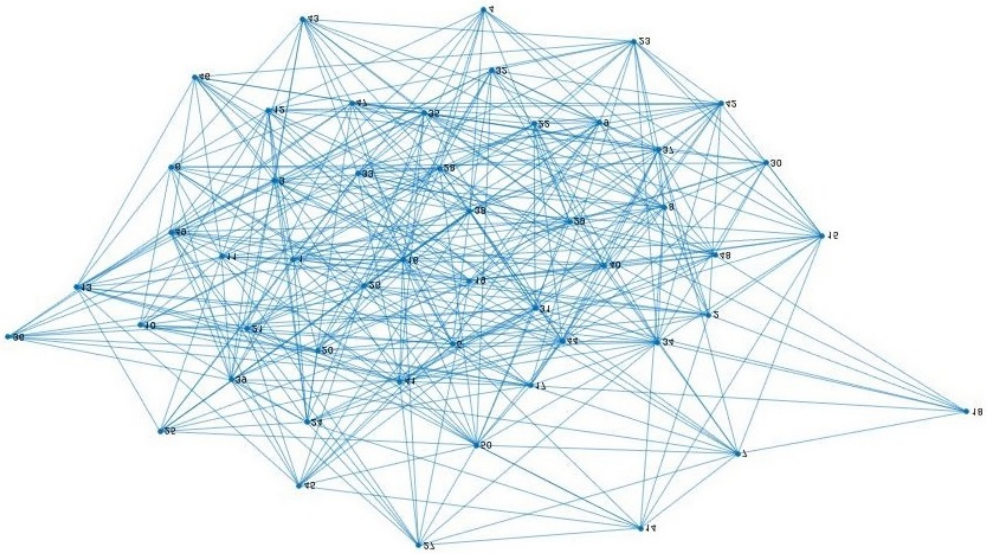


Figure 6.5: High connected graph

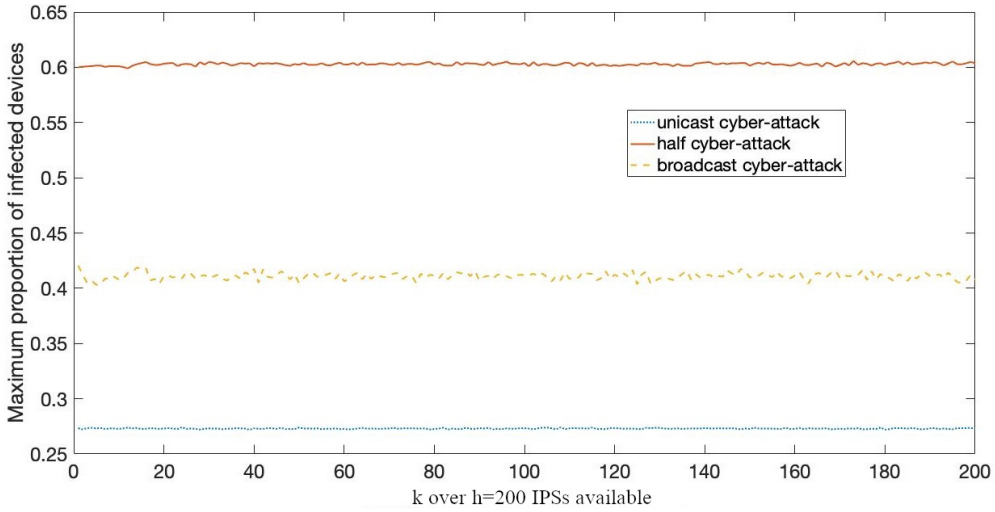


Figure 6.6: Maximum proportion of infected devices considering several cyber attack strategies. *Broadcast cyber attack does not lead to the worst scenario in terms of the maximum proportion of infected devices. In fact, by contaminating too many devices, botnet can be more easily detected by IPSs.*

whereas the broadcast cyber attack only 40%. This can be explained as contaminating too many devices, botnet can be more easily detected by IPSs.

Initial infected devices set

The set I_0 of initial infected devices has also an important impact on the performances of the k-SDS. As illustrated on figure 6.7, the smallest maximum number of infected devices is obtained with the 3-SDS mechanism when $I_0 = \{19\}$ is a singleton whereas it is the 7-SDS mechanism when the initial set of infected devices is bigger, i.e. $I_0 = \{19, 38, 4, 12, 22, 47\}$.

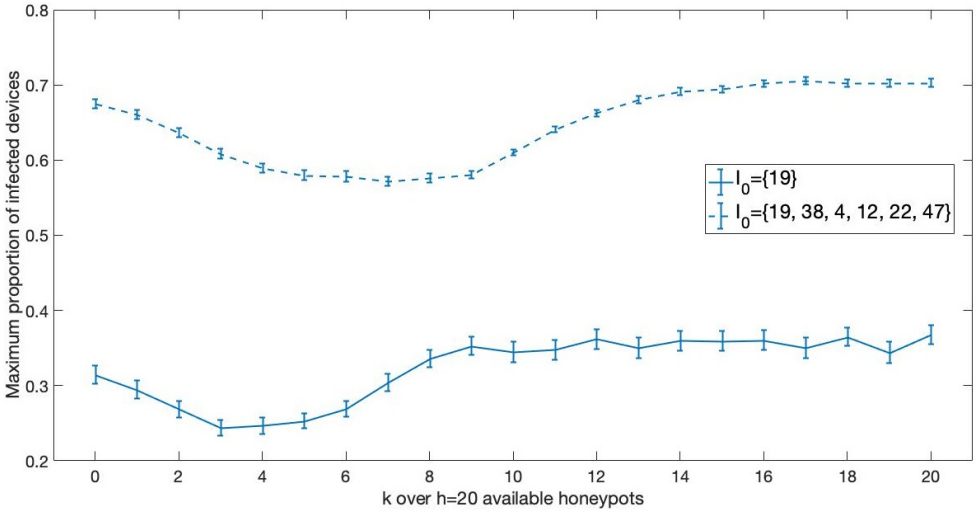


Figure 6.7: Maximum proportion of infected devices in the high connected network considering the k-SDS strategies with $k = 0, \ldots, 20$ and $h = 20$ available IPSs at each period. *An exploration-exploitation type of phenomenon is observed and the optimal deception strategy seems to be the* $3 - SDS$ *one.*

Note also that, as expected as the epidemic has more sources and therefore has a more powerful contamination strength, the smallest maximum number of infected devices obtained in both cases has a ratio of around 50%, i.e. values are less than 30% for I_0 singleton and 60% for the other case. Finally, a last interesting observation is that there exists some kind of plateau such that increasing k does not lead significant improvement of the metric. This phenomenon can be explained as when so much IPSs are placed around infected devices, less are used to discover other infected devices in other places in the network. Do not forget that IPSs have such a double goal, to cure for an infection and to discover infected devices.

Similar plateau is observed on the botnet time to extinction, on figure 6.8. In fact, as we have seen in the model description, botnet epidemic is a stochastic process such that when no device is infected, the botnet does no more exist. A very interesting observation about this metric over the simulations is that the 6-SDS mechanism leads almost the same performance in terms of botnet time to extinction for both initial infected devices sets. The values obtained through the simulations are 87.85 and 102.89, which are pretty close, whereas for the 20-SDS the gap is enormous (245.28 vs 1378) and also for the 0-SDS (104.5 vs 585.22) with both a ratio of 5. It shows that choosing a good deception mechanism has an important impact on this metric and also that a medium value of k such that enough

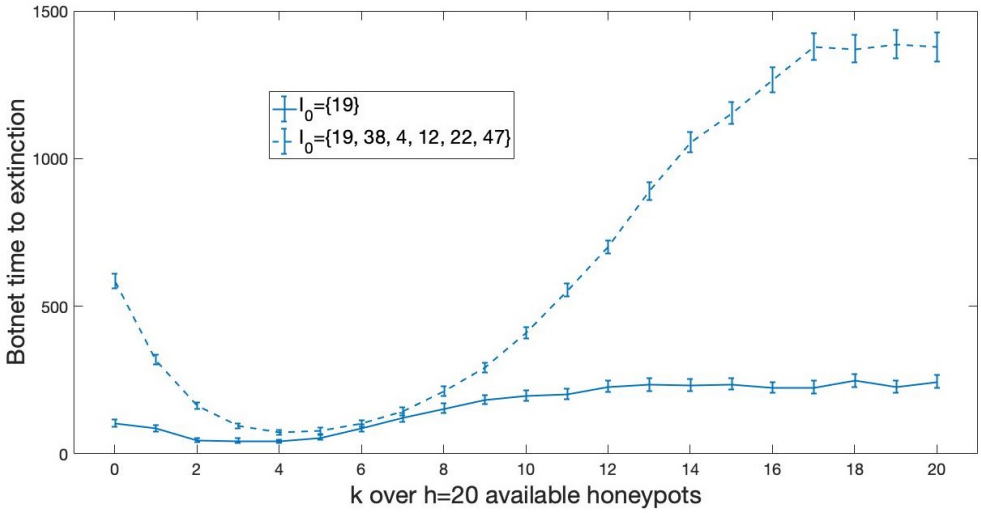


Figure 6.8: Time measured in periods to botnet extinction in the high connected network varying k over the $h = 20$ IPSs available at each period. *It takes 5 times more periods in order to eradicate the botnet using the* 18*-SDS compared to using the* 4*-SDS.*

IPSs are available to discover new infected devices, leads the best performances for this type of deception strategies.

6.5.5 Comparison of RDS and k^*-SDS for High Connected IoT Network

We have observed that for each value of h available IPSs at each period, there is an optimal value k^* that minimizes our performance metric. Here the strategy of the attacker is fixed to the unicast cyber attack which is one of the simplest cyber attack to implement because the first neighbor device in the susceptible state can be targeted. Then, this section illustrates the best-response strategy of the defender against such cyber attack over a particular set of smart deception strategies. We denote by k^*-SDS such optimal deception strategy in the set of all k-SDS for a given value of h. The value k^* means that k^* IPSs are used to control local infected devices on top of an infected one (exploitation), and $h - k^*$ IPSs are randomly positioned in other edges in order to capture other cyber attacks of the botnet (exploration).

For both metric, performances of k^*-SDS and RDS are pretty close when h is small, i.e. $h = 1$ (see figures 6.9 and 6.10). But, as we observe on figure 6.10 for the botnet time to extinction, the performance of k^*-SDS outperforms a lot the RDS mechanism when h becomes larger.

Increasing the number of IPSs improves the number of periods in order to eradicate the botnet with both k^*-SDS and RDS strategies as observed in figure 6.10. But we can see that even for small number of IPSs $h = 4$, the k^*-SDS strategy is very efficient compared to RDS. When $I_0 = \{19, 38, 4, 12, 22, 47\}$ it takes only few hundreds of periods to eradicate the botnet using 4-SDS, whereas it takes almost 6000 periods using RDS.

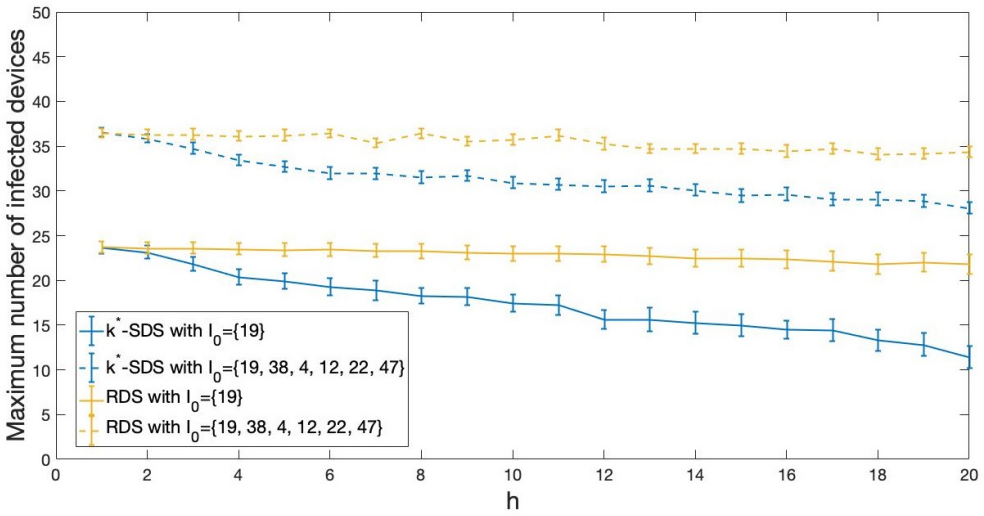


Figure 6.9: Comparison of the maximum number of infected devices with the k^*-SDS and RDS deception mechanisms in the high connected network. *When h is small, both k^*-SDS and RDS have similar performances for both initial infected devices scenarios. Whereas, as h is increasing, the gain in terms of reduction of the maximum number of infected devices is getting better and better for the k^*-SDS compared to RDS.*

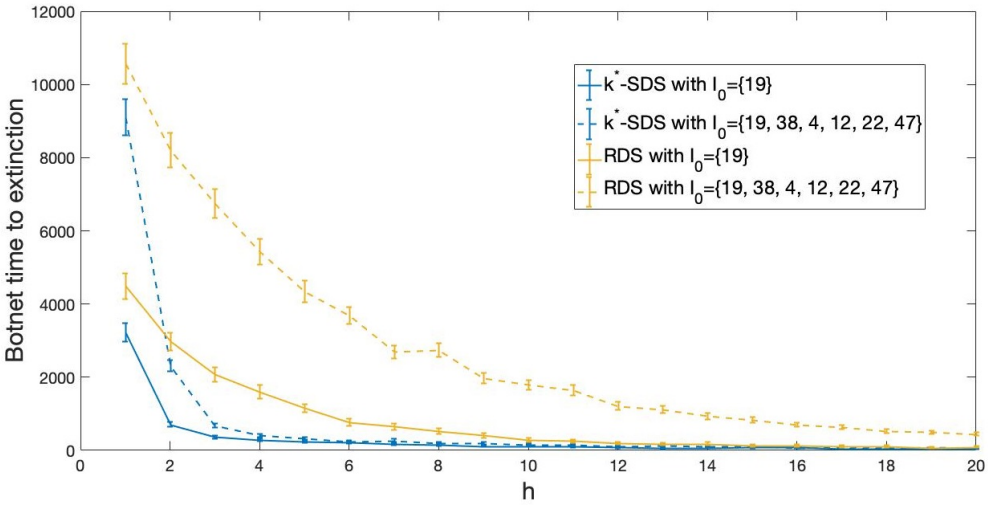


Figure 6.10: Comparison of the time to epidemic extinction of the k^*-SDS and RDS deception mechanisms. *k^*-SDS outperforms significantly RDS when h becomes larger enough.*

6.5.6 Comparison with Defense Techniques in the Literature

Finally, figure 6.11 illustrates the performances of our 6-SDS defense mechanism compared to traditional defense techniques which are the user-based and the network-based approaches [43], on the average number of infected devices for the highly connected network topology. The user-based technique assumes that each infected device recovers itself, for example

by installing the patch. The network-based technique assumes that a network authority disconnect the device in order to recover it. For the simulations, each infected device is disconnected of the network during one period.
Our scheme outperforms largely this two traditional defense mechanisms. After only 50 periods, the number of infected devices is less than 5% with our technique compared to more than 25% for the network-based approach and 65% for the user-based. The user-based technique has very bad performances because only individual decisions are performed and moreover, not all infected devices will be restored because it depends on the user's probability α to install the patch. For all the simulations, we have considered the same probability $\alpha = 0.1$ in order to be consistent for the comparison between the different defense mechanisms. Then, we conclude that IPSs based cyber-deception defense technique outperforms both traditional user-based and network-based approaches.

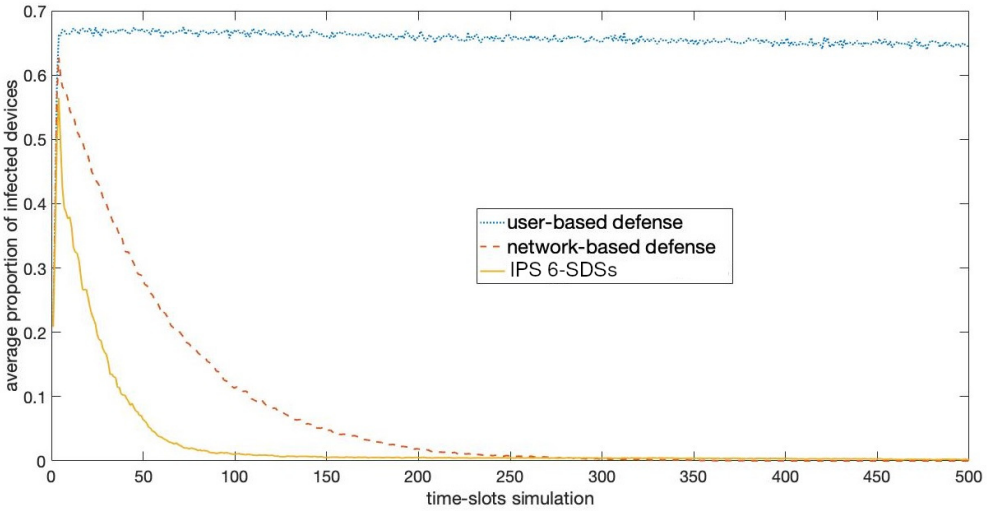


Figure 6.11: Comparison between user-based, network-based defense mechanisms and our IPS 6-SDS mechanism with 20 IPSs. *Our smart deception strategy outperforms largely both traditional defense techniques.*

6.6 Importance of the Degree

We supervised other simulations showing this time the impact of using an offensive strategy taking into account the network topology. For this experiment, a graph of 50 nodes was generated by the Erdős-Réyni method, each edge being activated with a probability of 0.3. We assume that the defender plays a 2-SDS with 8 IPSs available. Here, we present the result of simulations performed with two strategies for the attacker, which we refer to as k-smart attack strategies (k-SAS). As presented on figures 6.12 and 6.13, the k-SAS consists of propagating the worm from each infected node to a proportion of k of its likely higher degree neighbors, the degree of a node being defined as the number of its neighbors. In

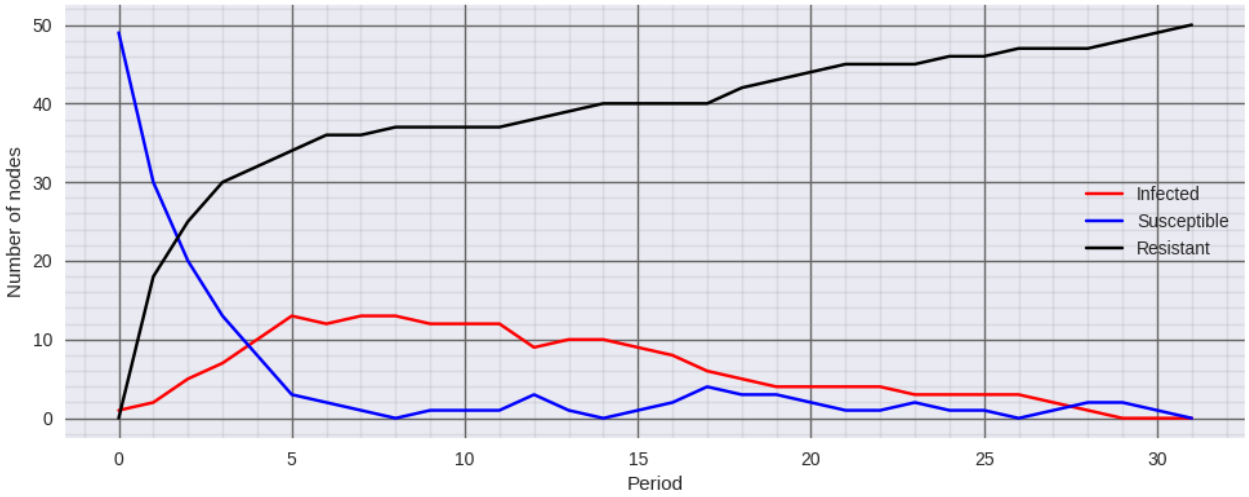


Figure 6.12: 30-SAS against 2-SDS *At each period, the attacker transmits the worm to* 30% *of the higher degree susceptible neighbors of each infected nodes, while the defender allocate* 2 *IPSs for exploitation and* 6 *IPSs for exploration. The epidemic reaches a peak of* 13 *nodes.*

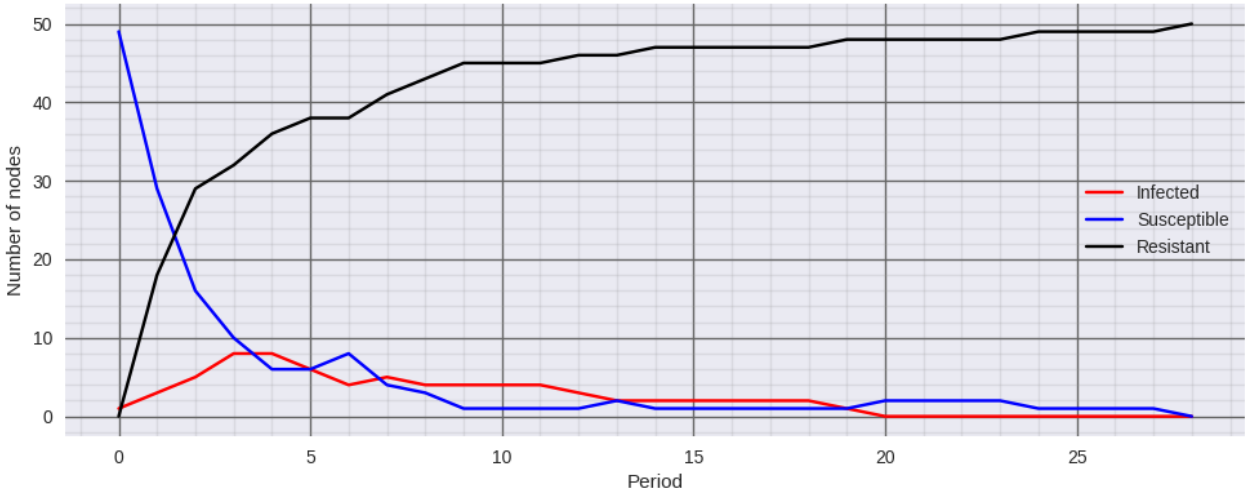


Figure 6.13: 60-SAS against 2-SDS *At each period, the attacker transmits the worm to* 60% *of the higher degree susceptible neighbors of each infected nodes, while the defender allocate* 2 *IPSs for exploitation and* 6 *IPSs for exploration. The epidemic reaches a peak of* 8 *nodes.*

the trial conducted, the 0.3-SAS produced a peak of 13 infected nodes, higher than the peak of 8 infected nodes produced by the more aggressive 0.6-SAS. This suggests that the degree of the node the attacker chooses to infect plays an important role in the spread of the epidemic, and that it is therefore counterproductive to attack any node without first assessing its importance.

6.7 Conclusions

The control of strategic SIR epidemic spread with opposite decision makers can be studied as POSMPGs with infinite horizon two-player zero-sum POSGs with goal states. The outcome of any player is captured by the overall extremum value in the instantaneous utility. The overall maximum value in the outcome of the attacker is the peak of the threat, and the defender aim to minimize it. To circumvent the non-interchangeability of the mean and the maximum, we derive a POSSPG from the POSMPG that admits the same optimal strategy profiles. Henceforth, the resolution of the first game may be done through the resolution of the latest one. However, unlike in [85], our POSSPG may generate zero reward while a goal state is not yet reached and, consequently, the boundedness of the optimal solution determined by the previous authors is not proved in our context. Nevertheless, we iteratively bound the optimal value of our POSMPG between the optimal values of two fine horizon POSGs. In order to obtain the algorithm for our inferred POSSPG, we first transform the finite horizon POSGs into infinite horizon discounted sum POSG, then the optimal value of the infinite horizon game is iteratively approximated by the point-based update of the finite bounds. We finally propose an algorithm that converges in the particular case of epidemic control.

Chapter 7

Using Graph Centrality for Smart Defense

Contents

7.1 Introduction

Chapters 5 and 6 propose two one-sided partially observable stochastic game framework to determine an optimal strategy for the defender. The proposed value iteration (VI) algorithm presents a major problem related to scalability (number of nodes for which the solution is

applicable). To overcome this problem, some authors have developed in [31] a so-called double oracle algorithm coupled with a compact representation of the defender belief. This approach allows to increase the scalability of the solution up to 40 nodes in the context of lateral movement problems with lower dimensional state and belief spaces. Meanwhile, the epidemic control problem generally applies to networks with numerous devices and, henceforth, requires more efficient tools.

To overcome this problem, we still model our epidemic process using well-known compartmental framework like the SIR framework. However, rather than proposing history-based strategies, and its high memory-consumption value iteration algorithm, we seek for a smart defensive strategy based on the possibility of threat transmission. To this end, we take into account the network topology, which consists in integrating the fact that the conflicting agents take their actions regarding the importance of the nodes in the network. Several metrics capture the importance of a node in a network, of which the degree, the eigenvalue centrality, the betweenness centrality, the local centrality with coefficient and the closeness centrality are among them [13, 50].

The centrality of a node is interpreted here as the ease with which its position allows the attacker to reach the greatest number of targets. This leads us to model the conflict as a two-player game on a graph between the attacker and the defender. The attacker sequentially chooses which hosts to attack, with the goal of increasing its reachability to as many other nodes as possible, possibly in several hops. Knowing this, the defender opts for an IPS deployment strategy that is supposed to minimize the potency that the attacker will realize. It is important to remember that neither player can guess the other's strategy unless they know the equilibrium strategy profile. The purpose of this chapter is to determine the Nash equilibrium of the game thus defined. The assumption of partial observability on the attacker side, according to which only the attacker knows the state of the network, remains valid.

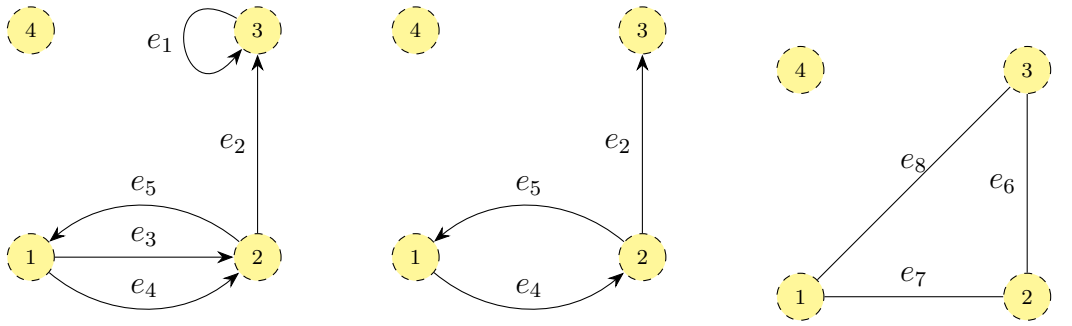
Before the formal definition and clear resolution of this game, section 7.3, we introduce the notion of centrality in a brief reminder on graph theory in the next section.

7.2 A Short Overview on Graph Theory

7.2.1 Definition of a Graph

A graph can be seen as a set of relations between elements, where the term relation can refer to a social link, a communication channel or something else. In [9], authors formalize this understanding of a ***graph*** as tuple $G = (V, E, \phi)$, where V and E are disjoint sets and ϕ is an application from E to V^2. However, this most general definition may be assigned to the ***multigraph***. More restricted notions of graph are given later in this subsection. Multiple relationships may exist between two elements. $\phi(e) = (u, v)$ means that e is a relationship from u to v. If applicable, u and v are called the ***ends*** of e, then u and v are ***neighbors*** of each other or, equivalently, v and is said to be ***adjacent*** to u. e is called a ***self-loop*** or a ***link***, depending on whether $u = v$ or $u \neq v$. A graph without self-loop is said to be

simple. Elements of E are referred to as ***edges***, elements of V are referred to as ***vertices*** or ***nodes***, and ϕ is called the ***incidence function***. For example, figure 7.1 represents three multigraphs with the same set of vertices, $V = \{1, 2, 3, 4\}$. On figure 7.1a, the set



(a) A non-directed graph (b) A directed graph (c) An undirected graph

Figure 7.1: Example of multigraphs

of edges is $E = \{e_1, e_2, e_3, e_4, e_5\}$ and the incidence function ϕ is defined by $\phi(e_1) = (3, 3)$, $\phi(e_2) = (2, 3)$, $\phi(e_3) = (1, 2)$, $\phi(e_4) = (1, 2)$, $\phi(e_5) = (2, 1)$. Edge e_1 is a self-loop and the other edges ares links.

When there is at most one relationship from any element a to any other element b, i.e., ϕ is an injection of E in V^2, if in addition there is no self-loop, the graph is said to be directed:

Definition 7.2.1 (***directed graph***)**.** *A* directed graph *is any tuple $G = (V, E)$ such that V is any set and $E \subseteq V^2 \setminus \delta(V)$, where $\delta(V) = \{(a, a) \mid a \in V\}$ is the* diagonal *of V.*

There is a one-to-one corespondance between finite directed graphs on V and $|V| \times |V|$ matrices of coeficients in $\{0, 1\}$ and 0 in the diagonal. Such a matrix is called ***adjacency matrix***. The coeficient at i-th line and j-th takes the value 1 iff j is adjacent to i. In figure 7.1b for example, the edges are $e_2 = (2, 3)$, $e_4 = (1, 2)$ and $e_5 = (2, 1)$, and the adjacency matrix is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. A symmetric adjacency matrix means that (i, j) is an edge iff (j, i) is an edge, and the graph is said to be directed.

Definition 7.2.2 (***undirected graph***)**.** *An* undirected graph *is any tuple $G = (V, E)$ such that V is any set and $E \in \mathcal{P}_2(V)$ is any subset of E of two elements.*

Except for this subsection, throughout this document, a ***graph*** is always undirected. In figure 7.1c, the edges are $e_6 = \{2, 3\}$, $e_7 = \{1, 2\}$ and $e_8 = \{1, 3\}$, and the symmetric adjacency matrix is $\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

The number of neighbors of a node j, i.e., the sum of coefficients in the j-th line of the adjacency matrix, is the ***degre*** of the node. When all nodes have the same degree, the graph is said ***regular***.

7.2.2 Paths in a Graph

For all $N \in \mathbb{N}^*$, any sequence $w = (e_n)_{n=1}^{N}$ of edges of a multigraph, for which there exists a sequence $(v_n)_{n=1}^{N+1}$ of vertices such that $\phi(e_n) = (v_n, v_{n+1})$ for all $n \in \{1, \dots, N\}$ is called a finite ***walk***. The walk is said ***closed*** if the two ends are equal, i.e., if $v_{n+1} = v_1$; it is called ***trail*** if no edge is repeated, i.e., if $e_m \neq e_n$ for all $m \neq n$ in $\{1, \dots, N\}$. A trail for which all vertices are different is called ***path***. A closed trail is called ***cycle***. In figure 7.1a for example, (e_5, e_4, e_2, e_1) is an open trail with sequence of edges $(2, 1, 2, 3, 3)$, (e_5, e_4) is a cycle and (e_3, e_4, e_2, e_1) is not a walk. Any link is a walk.

If the set of indices is $\mathbb{N}^*$, the walk is said semi-infinite, and is called a ***ray***. If the set of indices is $\mathbb{Z}$, the walk is said infinite. For a sequence w taken in a graph, first check if any edge is repeated $2k+1$ times ($k \in \mathbb{N}^*$) consecutively and remove it k times. The sequence w called walk if the remaining sequence w' meets the property $e'_n \cap e'_{n+1} \neq \emptyset$ at all positions, i.e., if the remaining consecutive edges are adjacent. A walk in a graph can be more simply defined as any sequence of alternatively vertices and edges, i.e., $\left((v_n, e_n)_{n=1}^{N}, e_{N+1}\right)$ [60] for a finite walk, $(v_n, e_n)_{n\in\mathbb{N}^*}$ for a semi-infinite walk, and $(v_n, e_n)_{n\in\mathbb{Z}}$ for a infinite walk, such that it always holds $e_n = \{v_n, v_{n+1}\}$. The ***length*** of a finite path is the number of its edges. Two nodes are ***connected*** if they are the ends of some path. When all nodes are connected to each other, the graph is said ***connected***. If all nodes are neighbors, the graph is said ***complete***.

7.2.3 Graph Random Generation

One important thing while performing simulations in network study is the random generation of a graph. The random generation involves an algorithm that will generate the graph, called ***graph generator***, and the rule that the generator will follow, called ***graph model***. [22] proposes a survey of these models. It defines a ***random graph*** for some set of vertices as "any model wherein is specified a probability distribution over a set of graphs". The model used in this thesis is the Erdős-Réyni one. In this model, the generator assumes that all nodes are "non-activated", then it activates nodes independently to each other with the same probability. The model may be expressed with the probability p of each edge activation, and noted $G(v, p)$, or with the number N of edges to activate of each node, and noted $G(v, N)$, where v is the number of vertices.

7.2.4 Graph Nodes Influence

General Definition of the Centrality, Degree Centrality

Degree can be used to measure the popularity of nodes in a graph [93], insofar as popularity can be seen as the number of nodes with which the node is connected. The degree effectively captures influence when nodes are in communication with each other, with each node communicating only with its immediate neighbors. This is the case, for example, in an attack that consists of injecting a worm into a device that will be propagated to its immediate neighbors to damage them. In this case, the attacker will give priority to the nodes with the highest degree. If, on the other hand, the attacker has a specific target in the network, she will target the nodes closest to it. In this case, importance is more a function of distance. If the attacker has to scan the network to identify her target, she will look for the node that is most influential in terms of its ability to reach any target. To propose a formal definition of the influence, also called centrality, without loss of generality, we assume that all vertices of the graph are positive integers.

Definition 7.2.3 (***centrality***)**.** *A* centrality measure *is any application* $c\colon \mathcal{P}_2(\mathbb{N}^*) \longrightarrow (\mathbb{R}_+)^{\mathbb{N}^*}$*, i.e., for all set* E *of pairs of positive integers,* $c(\cdot \mid E)$ *is an application from* $\mathbb{N}^*$ *to* $\mathbb{R}_+$*. For all positive integer* i*, the number* $c(i \mid E)$ *is the* centrality value *of* i*.*

Intuitively, if E is the set of edges, then the set of vertices can be restricted to $V = \bigcup_{e \in E} e$ and only these nodes may have a positive centrality. When the set E of edges is given, the centrality measure will be simply noted c, and the centrality value of a node i will be noted $c(i)$ instead of $c(i \mid E)$.

The ***degree centrality*** for example is the centrality defined by $c(i) = |N_j|$, where $N_j = \{j \mid j \in E\}$ is the set of node i neighbors.

Some Other Centrality Measures

The ***closeness centrality*** of a node is the average distance between the node and all other nodes in the graph. In the example above, which considers an attack on an unknown target, the attacker will prioritize the nodes with the largest closeness centrality. Regarding this measure, the more central a node is, the closer it is to all other nodes:

$$c_{\text{closeness}}(i) = \sum_{j \in E} \frac{1}{d(i,j)}, \tag{7.1}$$

where $d(i,j)$ is the distance between nodes i and j, i.e., the length of the shortest path between the two nodes.

The ***eigenvector centrality*** assigns relative scores to all nodes in the network, based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. The centrality of a node is determined from its direct connections with other nodes, and its power is derived from the centrality

values of these neighbors directly and other nodes in the network indirectly. From the Perron-Frobenius theorem, the adjacency matrix admits at least one positive eigenvalue [80]. Take the largest one, λ. If $\mathbf{x}$ is the vector whose components are the eigenvector centrality, then $\mathbf{x}$ is the eigenvector associated with the eigenvalue λ. Concretely, if $\mathbf{x}$ is the influence vector of the nodes of the graph, Ax is the vector returning for each node the sum of the influences of its neighbors. Thus, the equality

$$A = \lambda \mathbf{x} \tag{7.2}$$

means that by multiplying the influence of each node by the same constant λ, we obtain the sum of the influences of its neighbors. This means that the influence of each node is proportional to the sum of the influences of its neighbors. This measure is therefore applicable if the influence of each node is only related to that of its neighbors, for example for feature selection some affinity relationship is defined in the set of features [71].

Since information circulates in a network, if we admit that to go from a source to its destination it will always go through the shortest path, then the importance of a node i relative to two vertices s and t is the proportion $\dfrac{\sigma_{st}(i)}{\sigma_{st}}$ of the paths passing through i among the minimal paths connecting s and t. The influence of a node seen as the sum of these values on all pairs of the graph is called betweenness centrality. More precisely, the ***betweenness centrality*** is defined by

$$c_{\text{betweenness}}(i) = \sum_{\substack{s,t \in E \\ s \neq i \neq t}} \frac{\sigma_{st}(i)}{\sigma_{st}}, \tag{7.3}$$

where $\sigma_{st}(i)$ is the number of minimal paths connecting s and t through i, and σ_{st} is the total number of minimal paths connecting s and t.

7.3 Centrality Game

7.3.1 Reward Associated with an Action Profile

The ***centrality game*** is the (static) two-player zero-sum partially observable game defined as follows:

- The strategy makers and the actions are the same as in the stochastic game defined in chapter 5, i.e.:
  - The two players are the defender (player 1) and the attacker (player 2), and the system is the network (with the nodes),
  - the players actions are defined in equations (5.2), namely, defender actions are of type $W \in \mathcal{P}_{\leqslant h}(\mathbb{S}_b)$ and an attacker actions in state z is any sequence $\text{Tar} = (\text{Tar}_i)_{i \in I} \in (\mathbb{S}_z)^I$ of edges such that any two edges with distinct infected ends are disjoint, i.e., $i \neq j$ should imply $u \cap v = \emptyset$ for all edges u and v members of Tar_i and Tar_j respectively (see equations (5.1) for the stake),

– After the players' actions, each infected node transitions to susceptible state at probability α, and each susceptible node transitions to resistant state at probability ρ;

- The rewards are determined by the nodes and are additive. Namely, each node rewards with its centrality: the defender if it becomes susceptible, or the attacker if it becomes infected.

When an action profile $(W, \mathrm{Tar}) \in A_1 \times A_2$ is taken, the centrality value of any node is rewarded to the defender, if the node transits from infected to susceptible, or to the attacker, if the node transits from susceptible to infected. The expected partial reward of the defender is depicted in table 7.1, where c_k stands for the centrality measure of node k. The reward

		ATTACKER: Propagate $i \to j$?	
		Propagate ($\{i,j\} \in \partial\mathrm{Tar}$)	No propagate ($\{i,j\} \notin \partial\mathrm{Tar}$)
DEFENDER: Watch $\{i,j\}$?	Watch ($\{i,j\} \in W$)	$(1-\rho)\, c_i$	0
	No watch ($\{i,j\} \notin W$)	$-(1-\alpha)\, c_j$	0

Table 7.1: The defender's expected reward resulting from a joint action (W, Tar) on one edge

associated with an action profile $(W, \mathrm{Tar}) \in A_1 \times A_2$ is

$$\mathcal{R}(W, \mathrm{Tar} \,|\, z) = \sum_{\substack{\{i,j\} \in \mathbb{S}_z \cap \mathrm{Tar} \setminus W: \\ i \in I}} (1-\rho)\, c_i - \sum_{\substack{\{i,j\} \in \mathbb{S}_z \cap \mathrm{Tar} \setminus W: \\ i \in I}} (1-\alpha)\, c_j.$$

Remark 7.3.1. *In case there are exactly or less than h edges in the stake, the centrality game is no worth termed so since the defender will actually have no more than one option. So, in the remaining of this section, we assume that the stake contains more than h edges.*

7.3.2 Reward Associated with a Strategy Profile

Players play strategies $\pi_1 \in \Delta(A_1)$ and $\pi_2 : Z \longrightarrow \Delta(A_2)$. Denote Π_i the strategy space for player i. The expected reward of the defender with belief $b \in \Delta(Z)$ when the strategy profile $\pi = (\pi_1, \pi_2)$ is $\mathcal{R}(\pi|b) = \sum_{z \in Z} b(z)\, \mathcal{R}(\pi|z)$, where

$$\mathcal{R}(\pi|z) = \sum_{(W,\mathrm{Tar}) \in A_1 \times A_2} \pi_1(W)\, \pi_2(\mathrm{Tar}|z)\, \mathcal{R}(W, \mathrm{Tar} \,|\, z)$$

$$
\begin{aligned}
&= \sum_{\substack{(W,\mathrm{Tar})\in A_1\times A_2\\ \{i,j\}\in \mathbb{S}_z\cap \mathrm{Tar}_i\cap W:\\ i\in I}} \pi_1(W)\,\pi_2(\mathrm{Tar}|z)\,(1-\rho)\,c_i + \\
&\quad + \sum_{\substack{(W,\mathrm{Tar})\in A_1\times A_2\\ \{i,j\}\in \mathbb{S}_z\cap \mathrm{Tar}_i\setminus W:\\ i\in I}} \pi_1(W)\,\pi_2(\mathrm{Tar}|z)\,(\alpha-1)\,c_j \\
&= \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} \sum_{\substack{(W,\mathrm{Tar})\in A_1\times A_2:\\ \{i,j\}\in \mathrm{Tar}_i\cap W}} \pi_1(W)\,\pi_2(\mathrm{Tar}|z)\,(1-\rho)\,c_i + \\
&\quad + \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} \sum_{\substack{(W,\mathrm{Tar})\in A_1\times A_2:\\ \{i,j\}\in \mathrm{Tar}_i\setminus W}} \pi_1(W)\,\pi_2(\mathrm{Tar}|z)\,(\alpha-1)\,c_j \\
&= \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} \pi_1(\{i,j\})\,\pi_2(\{i,j\}|z)\,(1-\rho)\,c_i + \\
&\quad + \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} (1-\pi_1(\{i,j\}))\,\pi_2(\{i,j\}|z)\,(\alpha-1)\,c_j \\
&= \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} \pi_1(\{i,j\})\,\pi_2(\{i,j\}|z)\left((1-\rho)\,c_i+(1-\alpha)\,c_j\right) + \\
&\quad - \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} (1-\alpha)\,\pi_2(\{i,j\}|z)\,c_j \\
&= \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} \pi_2(\{i,j\}|z)\left(\pi_1(\{i,j\})\left((1-\rho)\,c_i+(1-\alpha)\,c_j\right)-(1-\alpha)\,c_j\right),
\end{aligned}
$$

$\pi_1(u) = \sum_{\substack{W\in A_1:\\ u\in W}} \pi_1(W)$ and $\pi_2(u|z) = \sum_{\substack{\mathrm{Tar}\in A_2:\\ u\in \mathrm{Tar}}} \pi_2(\mathrm{Tar}|z)$ being respectively the probabilities that the defender watches edge u and the attacker targets edge u.

So, the expected reward associated with the strategy profile π under the defender belief b is

$$
\begin{aligned}
\mathcal{R}(\pi|b) = &\sum_{z\in Z} \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} \pi_1(\{i,j\})\,\pi_2(\{i,j\}|z)\,b(z)\left((1-\rho)\,c_i+(1-\alpha)\,c_j\right) \\
&- \sum_{z\in Z} \sum_{\substack{\{i,j\}\in \mathbb{S}_z:\\ i\in I}} (1-\alpha)\,\pi_2(\{i,j\}|z)\,b(z)\,c_j
\end{aligned}
$$

$$= \sum_{z \in Z} \sum_{\substack{\{i,j\} \in \mathbb{S}_z : \\ i \in I}} \pi_2 \left(\{i,j\} | z \right) b \left(z \right) \left(\pi_1 \left(\{i,j\} \right) \left((1-\rho) c_i + (1-\alpha) c_j \right) - (1-\alpha) c_j \right),$$

or

$$\begin{aligned} \mathcal{R} \left(\pi | b \right) = & \sum_{\substack{\{i,j\} \in \mathbb{S}_z : \\ i \in I}} \pi_1 \left(\{i,j\} \right) \pi_2 \left(\{i,j\} | b \right) \left((1-\rho) c_i + (1-\alpha) c_j \right) \\ & - \sum_{\{i,j\} \in \mathbb{S}_z} (1-\alpha) \pi_2 \left(\{i,j\} | b \right) c_j \\ = & \sum_{\substack{\{i,j\} \in \mathbb{S}_z : \\ i \in I}} \pi_2 \left(\{i,j\} | b \right) \left(\pi_1 \left(\{i,j\} \right) \left((1-\rho) c_i + (1-\alpha) c_j \right) - (1-\alpha) c_j \right). \end{aligned}$$

One can also write

$$\begin{aligned} \mathcal{R} \left(\pi | b \right) &= \sum_{u \in \mathbb{S}_z} \pi_1 \left(u \right) \varphi_1 \left(u | b, \pi_2 \right) - \sum_{u \in \mathbb{S}_z} \psi_1 \left(u | b, \pi_2 \right) && (7.4) \\ &= \sum_{u \in \mathbb{S}_z} \pi_2 \left(u | b \right) \varphi_2 \left(u | \pi_1 \right), && (7.5) \end{aligned}$$

where, for all edge $u = \{i,j\}$ with $i \in I$,

$$\psi_1 \left(u | b, \pi_2 \right) = (1-\alpha) \pi_2 \left(u | b \right) c_j$$

is the expected marginal loss of the defender for not protecting the edge u, $(1-\rho) \pi_2 \left(u | b \right) c_i$ is a sort of attacker marginal "preserved advantage" if the defender does not protect the edge u,

$$\varphi_1 \left(u | b, \pi_2 \right) = \pi_2 \left(u | b \right) \left((1-\rho) c_i + (1-\alpha) c_j \right)$$

is the defender's marginal resulting unsatisfaction for not protecting the edge u,

$$\begin{aligned} \varphi_2 \left(u | \pi_1 \right) &= \pi_1 \left(u \right) \left((1-\rho) c_i + (1-\alpha) c_j \right) - (1-\alpha) c_j \\ &= \pi_1 \left(u \right) (1-\rho) c_i - \left(1 - \pi_1 \left(u \right) \right) (1-\alpha) c_j \end{aligned}$$

is the expected reward of the defender in case the attacker targets edge u, and

$$\pi_2 \left(u | b \right) = \sum_{u \in \mathbb{S}_z} \pi_2 \left(u | z \right) b \left(z \right)$$

is the marginal probability that the attacker spreads the virus through edge u, given that the defender's belief b. Note that:

1. For all $z \in Z$ and all $u \in E \setminus \mathbb{S}_z$, we have $\pi_2(u|z) = 0$.

2. For all $u \in E \setminus \mathbb{S}_b$, with $\mathbb{S}_b = \bigcup_{z \in \text{supp}(b)} \mathbb{S}_z$, we have $\pi_1(u) = 0$, where

$$\text{supp}(b) = \{z \in Z : b(z) \neq 0\}$$

is the ***support*** of the probability distribution b.

7.3.3 Best Responses to Players' Strategies

Remember that the defender wishes to maximize the payoff, while the attacker wishes to minimize it. This subsection advices each player when the opponent's strategy is known.

Best Response of the Defender to the Attacker's Strategy

A strategy π_1 for the defender is a best response to some strategy π_2 of the attacker when π_1 maximizes the reward $\mathcal{R}(\pi|b) = \sum_{u \in \mathbb{S}_z} \pi_1(u)\,\varphi_1(u|b,\pi_2) - \sum_{u \in \mathbb{S}_z} \psi_1(u|b,\pi_2)$. And the maximum payoff for a fixed attacker strategy π_2 and consequently fixed coefficients $\varphi_1(u|b,\pi_2)$ and $\psi_1(u|b,\pi_2)$ corresponds to the maximum of the $\sum_{u \in \mathbb{S}_z} \pi_1(u)\,\varphi_1(u|b,\pi_2)$'s and is obtained by taking $\pi_1(u) = 0$ whenever u is not top ranked according to $\varphi_1(\cdot|b,\pi_2)$, i.e., the defender does not have the greatest marginal unsatisfaction in expectation.

Note that the marginal probability distribution π_1 over $\mathbb{S}_b$ should be consistent with some probability distribution over A_1, and this is true only if the number of edges to watch is important enough to receive all the IPSs. In other words, the defender should focus on the h-top ranked $u \in \mathbb{S}_b$. To express this mathematically, define over the stake the rank $r(u|b,\pi_2) = 1 + \left|\left\{v \in \mathbb{S}_b : \varphi_1(v|b,\pi_2) > \varphi_1(u|b,\pi_2)\right\}\right|$ that should guide the choice for the defender.

This done, π_1 best responds to π_2 iff $\pi_1(u) = 0$ whenever $r(u|b,\pi_2) > h$, i.e., iff $\pi_1(u) = 0$ for all u not in the set $\text{SL}_1(\pi_2) = \left\{v \in \mathbb{S}_b : r(v|b,\pi_2) \leqslant h\right\}$ of the h-top ranked in the stake according to $\varphi_1(\cdot|b,\pi_2)$. However, for all $u \in \mathbb{S}_b$, $\pi_1(u) = 0$ iff

$$\forall W \in A_1, \quad u \in W \Longrightarrow \pi_1(W) = 0.$$

Practically, this means that the defender focuses on edges u with maximal values of $\varphi_1(u|b,\pi_2)$ and ignores the other edges, which is always possible. $\text{SL}_1(\pi_2)$ is the ***short list*** for the defender best responding to π_2 attacker strategy. The effective probability distribution over this short list should meet some probability distribution over A and it is not excluded that $\pi_1(u) = 0$ for some $u \in \text{SL}_1(\pi_2)$.

Best Response of the Attacker to the Defender's Strategy

Similarly, a strategy π_2 for the attacker best responds to some strategy π_1 of the defender when π_2 minimizes the reward $\mathcal{R}(\pi|b) = \sum_{u\in\mathbb{S}_z}\sum_{z\in Z}\pi_2(u|z)\,b(z)\,\varphi_2(u|\pi_1)$. And the minimum reward, for a fixed defender strategy π_1 and consequently fixed coefficients $\varphi_2(u|\pi_1)$ and $b(z)$, is realized if $\pi_2(u|z) = 0$ in any possible state z ($b(z) \neq 0$) in which $\varphi_2(u|\pi_1)$ is not minimal, i.e., if, for all possible state $z \in Z$, $\pi_2(u|z) = 0$ whenever u is not in the set $\mathrm{SL}_2(\pi_1) = \{v \in \mathbb{S}_b : \forall x \in \mathbb{S}_b, \varphi_2(x|\pi_1) \geqslant \varphi_2(v|\pi_1)\}$ of $\varphi_2(\cdot|\pi_1)$-minimally valued possibilities in the stake. However, for all $u \in \mathbb{S}_b$, $\pi_2(u|b) = 0$ iff

$$\forall z \in \mathrm{supp}(b), \pi_2(u|z) = 0,$$

and $\pi_2(u|z) = 0$ iff

$$\forall \mathrm{Tar} \in A_2, \quad u \in \mathrm{Tar} \Longrightarrow \pi_2(\mathrm{Tar}|z) = 0.$$

Practically, this means that when the network state is z, the attacker may transmit the virus through some edge u only if $\varphi_2(u|\pi_1)$ is minimal over the stake. In other words, the attacker transmits the virus through the edges of lower expected reward for the defender. $\mathrm{SL}_2(\pi_1)$ is the ***short list*** for the attacker best responding to the defender's strategy π_1.

7.3.4 Nash Equilibria

At Nash equilibrium (NE) $\pi^* = (\pi_1^*, \pi_2^*)$, each player best responds to his/her opponent's strategy and therefore, for all $u \notin \mathrm{SL}_n(\pi_{-n})$, player n assigns the probability 0 to u.

Lemma 7.3.1. *If* $\mathrm{SL}_1(\pi_2^*) \neq \mathbb{S}_b$, *then* $\mathrm{SL}_1(\pi_2^*) \subseteq \mathrm{SL}_2(\pi_1^*)$.

At NE, unless the short list for the defender extends to the hole stake, all member of this short list is in the short list for the attacker. In other words, unless the defender wishes to watch all edges, he does not care attacker's a priori irrelevant targets.

Proof. Suppose that $\mathrm{SL}_1(\pi_2^*) \neq \mathbb{S}_b$ and take any $u = \{i,j\} \in \mathbb{S}_b \setminus \mathrm{SL}_2(\pi_1^*)$, with $i \in I$. For all $z \in Z$, if $b(z) \neq 0$, then $\pi_2^*(u|z) = 0$. So, for all $z \in Z$, it holds $\varphi_1(u|b,\pi_2^*) = \pi_2^*(u|z)\,b(z)\left((1-\rho)\,c_i + (1-\alpha)\,c_j\right) = 0$. That is, u is minimally ranked according to $\varphi_1(\cdot|b,\pi_2^*)$ because $\varphi_1(v|b,\pi_2^*) \geqslant 0$, $\forall v \in \mathbb{S}_b$.

Since $\mathrm{SL}_1(\pi_2^*) \neq \mathbb{S}_b$, at least one $v \in \mathbb{S}_b$ is not h-top ranked according to $\varphi_1(\cdot|b,\pi_2^*)$. As $\varphi_1(v|b,\pi_2^*) \geqslant 0$, we conclude that, $u \notin \mathrm{SL}_1(\pi_2^*)$. □

Lemma 7.3.2. *1. For all* $\{i,j\} \in \mathrm{SL}_1(\pi_2^*)$, *all* $\{k,l\} \in \mathrm{SL}_2(\pi_1^*)\backslash\mathrm{SL}_1(\pi_2^*)$ *and all* $\{x,y\} \in \mathbb{S}_b \setminus \mathrm{SL}_2(\pi_1^*)$, *with* $i,k,x \in I$, *it holds:* $\begin{cases} c_j \geqslant c_l > c_y \\ c_j > c_l \Longleftrightarrow \pi_1^*(\{i,j\}) > 0 \end{cases}$.

2. For all pairs $\{k,l\}, \{k',l'\}$ *in* $\mathrm{SL}_2(\pi_1^*) \setminus \mathrm{SL}_1(\pi_2^*)$ *with* $\{k,k'\} \in I$, *it holds:* $c_l = c_{l'}$.

At NE: (1) the defender protects centrality values down to a certain threshold θ_1 while the attacker targets centrality values down to a not more important threshold θ_2; (2) all possible attacker target nodes that the defender should not protect have the same centrality value. Note that this lemma does not state that $\mathrm{SL}_2(\pi_1^*) \setminus \mathrm{SL}_1(\pi_2^*)$ is a non-empty set.

Proof. From $\{k,l\} \notin \mathrm{SL}_1(\pi_2^*)$, it comes $\pi_1^*(\{k,l\}) = 0$. From $\{k,l\} \in \mathrm{SL}_2(\pi_1^*)$, it comes that $\varphi_2(\{k,l\}|\pi_1^*) = -(1-\alpha)c_l$ is minimal. This point witnesses the second statement of the lemma. Note that $\{x,y\} \notin \mathrm{SL}_2(\pi_1^*)$, and from lemma 7.3.1, $\{x,y\} \notin \mathrm{SL}_1(\pi_2^*)$. So, $\pi_1^*(\{x,y\}) = 0$. Then $\varphi_2(\{x,y\}|\pi_1^*) = -(1-\alpha)c_y$ and, since $\varphi_2(\{k,l\}|\pi_1^*) = -(1-\alpha)c_l$ is minimal, it comes $-(1-\alpha)c_l < -(1-\alpha)c_y$ and, consequently, $c_l > c_y$.

In addition, the minimality of $\varphi_2(\{k,l\}|\pi_1^*)$ also applies to $(\{i,j\})$ and, therefrom, the equality $\pi_1^*(\{i,j\})\Big((1-\rho)c_i + (1-\alpha)c_j\Big) - (1-\alpha)c_j = -(1-\alpha)c_l$. Then, we get $\pi_1^*(\{i,j\})\Big((1-\rho)c_i + (1-\alpha)c_j\Big) = (1-\alpha)(c_j - c_l)$. The positivity of $c_j - c_l$ relies on that of $(1-\rho)c_i + (1-\alpha)c_j$. □

Theorem 7.3.1. *For some centrality values θ_1 and θ_2, it holds:*

1. $\mathrm{SL}_p\left(\pi_{-p}^*\right) = \left\{\{i,j\} \in \mathbb{S}_b : \quad j \in S \iff c_j \geqslant \theta_p\right\}$, *for* $p = 1,2$;

2. $\theta_2 \leqslant \theta_1$;

3. *If* $\theta_2 < \theta_1$, *then no centrality value can lie in the interval* (θ_2, θ_1).

Proof. Consider $\theta_p = \min\limits_{\substack{\{\text{source,target}\} \in \mathrm{SL}_p\left(\pi_{-p}^*\right): \\ \text{target} \in S}} c_{\text{target}}$, for $p = 1,2$. By this definition, $c_j \geqslant \theta_p$ for any $\{i,j\} \in \mathrm{SL}_p\left(\pi_{-p}^*\right)$. Conversely, on the one hand, take any $\{k,l\} \in \mathbb{S}_b$ such that $c_l \geqslant \theta_2$. The minimum value θ_2 is attained for some $\{k',l'\} \in \mathrm{SL}_2$. Then, from the inequality $c_l \geqslant c_{l'}$ and lemma 7.3.2 (point 1), it comes $\{k,l\} \in \mathrm{SL}_p\left(\pi_{-p}^*\right)$. Indeed, if $\{k,l\} \in \mathbb{S}_b \setminus \mathrm{SL}_p\left(\pi_{-p}^*\right)$, then $c_l < c_{l'}$. On the other hand, take any $(i,j) \in \mathbb{S}$ such that $c_j \geqslant \theta_1$. Point 1 is proven.

Since $\mathrm{SL}_1(\pi_2^*) \subseteq \mathrm{SL}_2(\pi_1^*)$ and from the definition of θ_p, $p = 1,2$, we have $\theta_2 \leqslant \theta_1$ (and more specifically $\theta_2 < \theta_1$ iff $\mathrm{SL}_1(\pi_2^*) \subsetneq \mathrm{SL}_2(\pi_1^*)$). Point 2, is proven.

For the proof of point 3, let's assume that there is $\{x,y\} \in \mathrm{SL}_2(\pi_1^*)$ and $\{i,j\} \in \mathbb{S}_b$ with $y,j \in S$ such that $\theta_2 = c_v$ and c_j is in the space (θ_2, θ_1). In this case, $\{i,j\} \in \mathrm{SL}_2(\pi_1^*) \setminus \mathrm{SL}_1(\pi_2^*)$. Therefore, $\{i,j\}, \{x,y\} \in \mathrm{SL}_2(\pi_1^*) \setminus \mathrm{SL}_1(\pi_2^*)$ and, by lemma 7.3.2 (point 2), $c_j = \theta_2$. Which is absurd. Point 3 is proven. □

The above results establish a connection between the Nash equilibrium and the centralities to be protected or defended. It is important to be able to leverage them for the final setting of the NE.

7.4 Computation of the Nash Equilibria

We assume that the players are playing an NE strategy profile π. That is, the supports $\operatorname{supp}(\pi_i)$, $i = 1, 2$ of their strategies π_i are included in their respective shortlists $\mathrm{SL_i}(\pi_{-\mathrm{i}})$ that depends upon respective thresholds θ_i.

7.4.1 Shortlists Exploitation

In this subsection, we note s the minimum value of φ_2 under the NE, i.e., $s = \min_{u \in \mathbb{S}} \varphi_2(u|\pi_1)$. Also, when the pair $\{i, j\}$ represents an edge of the stake, by abuse, we admit that j is the infected node.

Proposition 7.4.1. *The shortlists and the NE strategies are subject to the following properties:*

1. $$s = -(1-\alpha)\frac{\left(\sum\limits_{\substack{\{i,j\}\in\mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{c_j}{(1-\rho)c_i + (1-\alpha)c_j}\right) - \frac{h}{1-\alpha}}{\sum\limits_{\substack{\{i,j\}\in\mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{1}{(1-\rho)c_i + (1-\alpha)c_j}}.$$

2. *For all* $\{i, j\} \in \mathbb{S}$,
$$\begin{cases} \{i,j\} \in \mathrm{SL}_2(\pi_1) \Longrightarrow \varphi_2(\{i,j\}|\pi_1) = s \text{ and } \pi_1(\{i,j\}) = \dfrac{s + (1-\alpha)c_j}{(1-\rho)c_i + (1-\alpha)c_j} \\ \{i,j\} \notin \mathrm{SL}_2(\pi_1) \Longrightarrow \varphi_2(\{i,j\}|\pi_1) > s \text{ and } \pi_1(\{i,j\}) = 0 \end{cases}.$$

3. $$\sum_{\substack{\{i,j\}\in\mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{c_j - \theta_1}{(1-\rho)c_i + (1-\alpha)c_j} \leqslant \frac{h}{1-\alpha}.$$
If $\theta_1 > \theta_2$, *then* $$\sum_{\substack{\{i,j\}\in\mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{c_j - \theta_2}{(1-\rho)c_i + (1-\alpha)c_j} = \sum_{\substack{\{i,j\}\in\mathbb{S}_b \\ c_j \geqslant \theta_2}} \frac{c_j - \theta_2}{(1-\rho)c_i + (1-\alpha)c_j} = \frac{h}{1-\alpha}.$$

4. *If* Last_h *is any subset of* $\left\{ \{i,j\} \in \mathbb{S}_b \colon c_j \geqslant \theta_1 \right\}$ *consisting in* h *last ranked elements of the plausible stake according to* $\pi_1(\{i,j\})$, *then:*
$$\sum_{\{i,j\}\in\mathrm{Last}_h} \pi_1(\{i,j\}) \geqslant \frac{(h-1)h}{|\mathbb{S}_b| - 1}.$$

5. *If* $s \geqslant 0$ *then* $\theta_1 = \min\limits_{\{i,j\}\in\mathbb{S}_b} c_j$.
If $s < 0$ *then the attacker infects a susceptible node* j *if and only if that for some infected node* i, *it holds* $\varphi_2(i, j|\pi_1) = s$.

Proof. The comparison of $\varphi_2(\{i,j\}|\pi_1)$ and s comes from the definition of the attacker's shortlist; in case $\{i,j\} \in \mathrm{SL}_2(\pi_1)$, the value of $\pi_1(\{i,j\})$ comes from this comparison. The point 2 is proven. Point 1 comes from the facts that the probabilities $\pi_1(\{i,j\})$ sum to the number h of edges the defender chooses, and $\pi_1 = 0$ out of the defender's shortlist. The non-negativeness of π_1 implies $(1-\alpha)\,c_j \geqslant s$ then $(1-\alpha)\,\theta_1 \geqslant s$ on the defender's shortlist. In case $\theta_1 > \theta_2$, for some $(i,j) \in \mathrm{SL}_2 \setminus \mathrm{SL}_1$, it holds $c_j = \theta_2$ and $\pi_1(\{i,j\}) = \dfrac{s + (1-\alpha)\,c_j}{(1-\rho)\,c_i + (1-\alpha)\,c_j} = 0$. So, we get $s = (1-\alpha)\,\theta_2$. Thus, $(1-\alpha)\,c_j \geqslant s$ in the general case, and $s = (1-\alpha)\,\theta_2$ in case $\theta_1 > \theta_2$. This witnesses point 3. Point 4 is a condition that comes from [101]. Note that it is equivalent to

$$\sum_{\{u,v\} \in \mathrm{Last}_h} \frac{-(1-\alpha)\dfrac{\left(\displaystyle\sum_{\substack{\{i,j\} \in \mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{c_j}{(1-\rho)\,c_i + (1-\alpha)\,c_j}\right) - \dfrac{h}{1-\alpha}}{\displaystyle\sum_{\substack{\{i,j\} \in \mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{1}{(1-\rho)\,c_i + (1-\alpha)\,c_j}} + (1-\alpha)\,c_y}{(1-\rho)\,c_x + (1-\alpha)\,c_y} \geqslant \frac{(h-1)\,h}{|\mathbb{S}_b| - 1}. \tag{7.6}$$

For the proof of 5, suppose $s \geqslant 0$. That is, for any $\{i,j\} \in \mathbb{S}_z$, we get successively:

$$\begin{aligned}
&\varphi_2(\{i,j\}|\pi_1) \geqslant 0,\\
&\pi_1\{i,j\}\,((1-\rho)\,c_i + (1-\alpha)\,c_j) - (1-\alpha)\,c_j \geqslant 0,\\
&\pi_1\{i,j\} \geqslant \frac{(1-\alpha)\,c_j}{(1-\rho)\,c_i + (1-\alpha)\,c_j} > 0,\\
&\{i,j\} \in \mathbb{S},\\
&c_j \geqslant \theta_1.
\end{aligned}$$

Suppose on the other hand that $s < 0$. From the definition of the attacker'r shortlist, it comes: $\mathcal{R}(\pi|b) = \displaystyle\sum_{\substack{\{i,j\} \in \mathbb{S}_z \\ \varphi_2(i,j|\pi_1) \text{ is minimal}}} \pi_2(\{i,j\}|b)\,\varphi_2(i,j|\pi_1) = s \sum_{\substack{\{i,j\} \in \mathbb{S}_z \\ \varphi_2(i,j|\pi_1) = s}} \pi_2(\{i,j\}|b)$. The minimization of this result imposes maximization of the $\pi_2(\{i,j\}|b)$'s whenever $\varphi_2(\{i,j\}|\pi_1)$ is minimal. □

By taking $p = \displaystyle\sum_{\substack{\{i,j\} \in \mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{c_j}{(1-\rho)\,c_i + (1-\alpha)\,c_j}$ and $q = \displaystyle\sum_{\substack{\{i,j\} \in \mathbb{S}_b \\ c_j \geqslant \theta_1}} \frac{1}{(1-\rho)\,c_i + (1-\alpha)\,c_j}$, Point 3 of the above proposition can be rewritten: $\begin{cases} p - \theta_1 q \leqslant \dfrac{h}{1-\alpha} \\ \theta_1 > \theta_2 \Longrightarrow p - \theta_2 q = \dfrac{h}{1-\alpha} \end{cases}$ or, equivalently,

$\begin{cases} \theta_1 \geqslant \theta \\ \theta_1 > \theta_2 \Longrightarrow \theta_2 = \theta \end{cases}$, where $\theta = \dfrac{p - \dfrac{h}{1-\alpha}}{q}$.

Suppose that $\theta_1 = \theta$. Then, $\theta_1 = \theta_2$ because either, we get the inequality $\theta_1 > \theta_2$, and from it, we derive its contradiction $\theta_1 = \theta_2$.

Conversely, suppose that $\theta_1 > \theta$. If $\theta_1 > \theta_2$, then $\theta_2 = \theta$, and θ is a centrality value. Suppose now that $\theta_1 = \theta_2$ and θ is the centrality value of some susceptible node in the stake, i.e., for some $\{i, j\} \in \mathrm{S}_b$, it holds $c_j = \theta$. Note that $s = -(1-\alpha)\theta$. For $\{i, j\} \notin \mathrm{SL}_1(\pi_2)$, we get $\pi_1(\{i, j\}) = 0$ and $\varphi_2(\{i, j\}|\pi_1) = -(1-\alpha)c_j = s$, which contradict the above proposition since $\{i, j\} \notin \mathrm{SL}_2(\pi_1)$.

Proposition 7.4.2. *The attacker targets susceptible nodes of centrality values above or equal to θ.*

7.4.2 Simulations

We performed simulations of the centrality set with the dual purpose of verifying whether the NE of the centrality set is unique and whether this solution is more scalable than the POSG solution. For this experiment, we assumed that the defender knows the state of the network. Our method consists in considering beforehand that all the centrality values of the nodes of the graph are possible values of θ_1. Then, to each possible value of θ_1, we associate the possible values of s, p, q and θ, and we eliminate the entries that do not meet proposition 7.4.1. We find that a single entry satisfies all of this proposition, which implies the relevance of Nash equilibria, in that a player's outcome does not depend on the NE strategy profile in which he is participating.

We observed the complete course of the game considering that centrality is the degree. We also assumed that the graph contains 2000 nodes with 120 initially infected ones and that the defender has 100 IPSs. We then performed 350 trials, each trial starting with a new randomly generated graph following an Erdős-Réyni graph model, aiming at an average degree of 8 per node. We observe that the 350 constructed graphs have an average degree of about 8.00745, with a small relative standard deviation, of 6.36%.

The calculations involved a total of 8977.59 seconds, or about 25.6503 seconds per trial. $\alpha = 0.1$, and $\rho = 0.5$. Figure 7.2 shows an overview of the average result. The epidemic reaches its peak, of 148, in the interval $[146.735, 149.322]$ with a degree of confidence of 99%, which corresponds to the 2nd period which, with the same degree of confidence, is in the interval $[2.01077, 2.05781]$. After the 4th period ($[3.89515, 3.97914]$), i.e., at the 5 period, the number of IPSs is no longer lower than the number of edges to protect, and the epidemic is under control. All nodes are cleaned after the first 11 periods ($[10.6675, 11.224]$) and all nodes are resistant after 15 periods ($[15.0343, 15.6172]$).

Similar simulations were performed using other centrality measures, including the betweenness, the closeness, the eigenvector centralities, and the semi-local centrality with coefficient. It was found that the results obtained did not depend on the centrality measure. This is a predictable result given the nature of the graph used. Indeed, the Erdős-Réyni

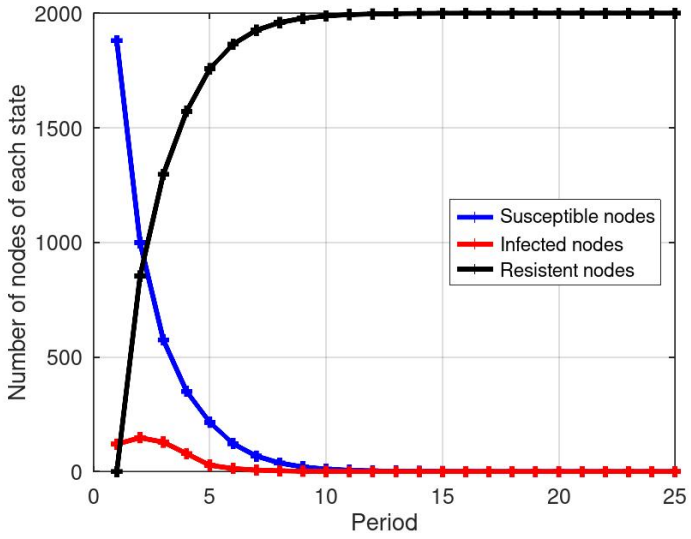


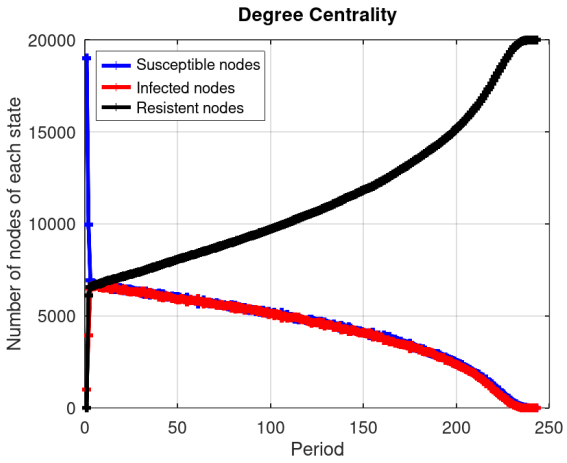
Figure 7.2: Centrality game, simulations with 2000 nodes
Number of nodes: 2000*, of which* 120 *initially infected nodes. Peak:* 148.029*, after* 2.03429 *periods. Under control after* 3.93714 *periods. Extinction after* 10.9457 *periods. Network safe after* 15.3257 *periods.*

graph model leads to an almost regular graph and, if all the nodes have the same degree, it is logical to think that all the above centrality measures are identical.

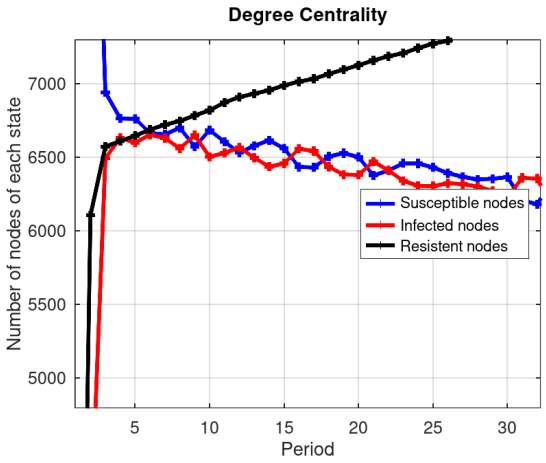
Finally, we performed a simulation on a graph generated by the same method, with 20000 nodes and an expected average degree of 20 per node. The generated graph has this average degree. Keeping the same transition probabilities, we simulated a game where the defender has 30 IPSs to protect the 20000 nodes network already containing 1000 infected nodes. Performed with a computer with $8Gb$ RAM, the simulation lasted 23271.3 seconds and rendered the figure 7.3. The peak, of 651 infected nodes, is reached in the 6th period and almost reached again in the 9th period (649 nodes). The defender takes control of the epidemic in period 235 and the epidemic dies out in period 238, 5 periods before all nodes are resistant.

7.5 Conclusion

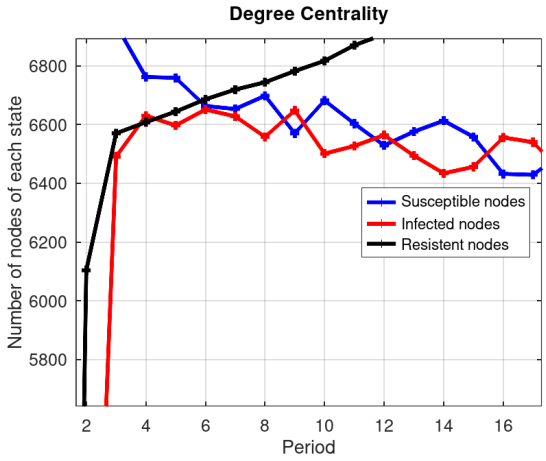
In order to circumvent the hardness of the value iteration algorithm, we study two smart deception defense strategies against a random attack strategy and proved that the optimal smart defense strategy outperforms the random defense strategy. Then, we propose a game theoretic framework based on nodes' centrality. At each period, each player optimizes the centrality values of nodes under his/her control. The attacker and the defender are therefore involved in a Bayesian game in which the type of attacker corresponds to the state of the network. The NE corresponds to a situation where each player has fixed a threshold and pro-



(a) Complete plot



(b) Zoom on the 30 first periods



(c) Zoom on periods 2 through 16

Figure 7.3: Simulation with 20000 nodes

tects/attacks only nodes of centrality bounded from below by this threshold, the defender's being at least equal to that of the attacker. These properties yield equations that allow a computational resolution determination of the NE. Experiments prove that the game is solved in a very short time, even for an important number of nodes, and the defender takes control of the epidemic at the earlier stages of the confrontation. Our framework applies to any centrality measure, and we used the degree of the nodes. We have solved the problem in the case where the information is complete for both the defender and the attacker and assuming that the agents play with the same centrality measures.

Chapter 8

Conclusion

Contents

8.1 Defensive Cyber Deception and Network Epidemics

8.1.1 Cyber Security and Network Epidemics

Taking into account the possible compartmentalization of a population and the possible transitions of an individual between classes, a simple combinatorial analysis reveals 16 general epidemic models, 14 of which are SIR. One of these models describes the active and stealthy diffusion of a threat in a network. After giving an understanding of cyber attacks and cyber security rules, we present a defense technique in this post-penetration context, which consists in trapping worm transmissions between devices, and we propose smart behaviors to stop such a threat. To do so, we assume limited resources, and we examine the impact of network exploration and exploitation of enemy vulnerabilities on the quality of the outcome. These strategies are proposed taking into account different forms that attacker smartness could take.

8.1.2 Cyber Deception against Cyber Deception

To give an optimal response to the attack, we proceed to the formal description of the context and the perpetrators' action. The context gives the attacker the advantage of knowing the state of the network. Secondly, they use obfuscation cyber deception to prevent collaboration between the protectors and the users of the network. This results in a doubly asymmetric battle. To reduce asymmetry, we propose that the defense performs mixing cyber deception. The confrontation is then an SG with incomplete information on one side and imperfect information on both sides. In line with our contribution to solving this type of game, we argue that the proposed algorithm for classical POSGs suits it. We consider the non-additivity of the threat and make the realistic assumption that protectors seek only to optimally minimize the maximum threat.

8.2 Game Theoretical Solutions

Game theory is the of the essence point of view for an effective and efficient analysis of conflicts between intelligent and rational individuals. Its optimal exploitation requires that it be studied in its principles and that, eventually, the tools it proposes be adapted to the particular context in which it is to be implemented. Its criterion par excellence for option selection is the Nash equilibrium, which is defined as a strategy profile such that, if adopted, no player has an advantage in deviating from it.

8.2.1 Review of Stochastic Game Resolution

In the case of a two-player stochastic zero-sum game, which involves an unmonitored transition of the system after each period, one player minimizes what the other player maximizes, and the Nash equilibrium is the behavioral strategy profile that corresponds to the optimal value. Some convergent value iteration algorithm describes the process of reaching the optimal value of a POSG when both incomplete and imperfect information is assumed to be one-way. Returning to the Bellman equation, we show that the VI algorithm remains convergent when the imperfect information is two-sided. This is demonstrated when utility is the sum of the gains or losses generated throughout the process, each gain discounted beforehand, the discount factor being a number $\lambda \in (0, 1)$. This assumption is inappropriate when the maximizer is only motivated by the maximum payoff over the overall process. In this case, thanks to an apt modification of the set of states of the system, we show that by playing on the maximum-utility, the players play a parallel game of discounted sum-utility. We deduce an algorithm converging to the solution of maximum-utility POSGs. We thus exploit existing results of game theory to develop new ones, which improve the fight against epidemics in cyber space.

8.2.2 Graph Theory in support of Game Theory

To circumvent the non-scalability of the VI algorithm, we propose to study the confrontation from the perspective of targeting devices with better potential to continue the propagation. This implies taking into account the network topology. Graph theory is the discipline par excellence for studying the importance of a node in a network. We briefly discuss it to give a general definition of centrality, which is already known as a measure of influence. We formulate a static game model for dynamic epidemic control. The game thus formulated is Bayesian, i.e., a game with incomplete information. The information of at least one player is thus a probability distribution over the different possibilities of the information. This requires that the Nash equilibrium be sought among mixed strategies, i.e. probability distributions over pure strategies. Solving this game allows to improve considerably the scalability.

8.3 Future Work

Through this diverse contribution, which harmoniously and in an original way combines game theory, graph theory, the mathematical study of epidemic models and cyber deception to safe the cyber space, our contribution opens the way to promising perspectives in cyber security. Here are a few of them:

- We propose to deceive the attacker by letting him record a victory made transient without her knowledge, which has the effect of further informing the defender about the system state. Perhaps a deception should be applied that goes beyond a time period and causes the attacker to launch his DDoS from a honey-net and/or even on a fake machine.

- We solve the POSMPG by pairing it with the POSSPG. In this POSSPG, a number of periods, an unlimited number in probability, can pass without a reward being recorded. The sequence of non-zero rewards is consequently very sparse. An exploitation of this sparsity will probably improve the scalability of the game.

- Game theory provides the right answer to an intelligent and rational opponent. This brings a complication in its implementation in a context where the defender cannot count on the rationality of the users who, by the way, are deceived by the attacker. We have represented in terms of transition probabilities the result of his solicitation of the users, and we have assumed these probabilities to be constant. By using machine learning, we could know more about the users and give an even better response. Moreover, rather than getting bogged down in a search for an optimal strategy taking into account all the history, one could make use of machine learning this time to improve or circumvent the scalability of the VI algorithm.

- We replace a SG with a Bayesian game involving the centrality of nodes in a graph. This option goes beyond the epidemic setting and applies to all two-player, zero-sum SGs. The search for theoretical equivalence between the respective solutions of Bayesian and

stochastic games will not only solve the scalability of SGs and MDPs on graphs, but also answer the thorny question of the significance of a centrality measure. It will also allow to improve learning algorithms, which rely on MDPs.

Bibliography

[1] Charalambos D Aliprantis, Dionysius Glycopantis, Allan Muir, et al. On mixed and behavioural strategies. *Economics Bulletin*, 29(3):1783–1795, 2009.

[2] Manos Antonakakis, Tim April, Michael Bailey, Matthew Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. Understanding the mirai botnet. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 1093–1110, Vancouver, BC, Canada, 08 2017. USENIX Association.

[3] Manos Antonakakis, Tim April, Michael Bailey, Matthew Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. Understanding the mirai botnet. In *26th USENIX Security Symposium*, pages 1093–1110, 2017.

[4] Neetu Bansla, Swati Kunwar, and Khushboo Gupta. Social engineering: A technique for managing human behavior. *Journal of Information Technology and Sciences*, 5(1):18–22, 2019.

[5] Arnab Basu and L. Stettner. Finite- and infinite-horizon shapley games with nonsymmetric partial observation. *SIAM Journal on Control and Optimization*, 53:3584–3619, 01 2015.

[6] D. Bernstein, R. Givan, N. Immerman, , and S. Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of Operations Research*, 27(4):819–840, 2002.

[7] Michael Bloem, Tansu Alpcan, and Tamer Basar. Intrusion response as a resource allocation problem. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 6283–6288. IEEE, 2006.

[8] Sally M Blower, Angela R Mclean, Travis C Porco, Peter M Small, Philip C Hopewell, Melissa A Sanchez, and Andrew R Moss. The intrinsic transmission dynamics of tuberculosis epidemics. *Nature medicine*, 1(8):815–821, 1995.

[9] John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al. *Graph theory with applications*, volume 290. Macmillan London, 1976.

[10] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jure Leskovec, and Christos Faloutsos. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security*, 10(4), 2008.

[11] PR Chamberlain. Twitter as a vector for disinformation. *Journal of Information Warfare*, 9(1):11–17, 2010.

[12] Krishnendu Chatterjee and Laurent Doyen. Partial-observation stochastic games: How to win when belief fails. *ACM Transactions on Computational Logic*, 15, 07 2011.

[13] Duanbing Chen, Linyuan Lü, Ming-Sheng Shang, Yi-Cheng Zhang, and Tao Zhou. Identifying influential nodes in complex networks. *Physica a: Statistical mechanics and its applications*, 391(4):1777–1787, 2012.

[14] Lili Chen, Zhen Wang, Fenghua Li, Yunchuan Guo, and Kui Geng. A stackelberg security game for adversarial outbreak detection in the internet of things. *Sensors*, 20:804, 02 2020.

[15] Taolue Chen, Vojtěch Forejt, Marta Kwiatkowska, Aistis Simaitis, Ashutosh Trivedi, and Michael Ummels. Playing stochastic games precisely. In *International Conference on Concurrency Theory*, pages 348–363. Springer, 2012.

[16] Zesheng Chen. Modeling and defending against internet worm attacks. *PhD thesis*, 2007.

[17] Zesheng Chen, Lixin Gao, and Kevin Kwiat. Modeling the spread of active worms. In *IEEE INFOCOM*, volume 3, pages 1890–1900. IEEE, 2003.

[18] Nicholas Cifranic, Roger A Hallman, Jose Romero-Mariona, Brian Souza, Trevor Calton, and Giancarlo Coca. Decepti-scada: A cyber deception framework for active defense of networked critical infrastructures. *Internet of Things*, 12:100320, 2020.

[19] Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. Efficient immunization strategies for computer networks and populations [j]. *Physical Review Letters*, 91:247901, 09 2013.

[20] Richard de Beer, Adrie Stander, and J Van Belle. Anti-forensic tool use and their impact on digital forensic investigations: A south african perspective. In *Proceedings of the International Conference on Information Security and Digital Forensics*, pages 7–20, 2014.

[21] Yuri Diogenes and Erdal Ozkaya. *Cybersecurity–Attack and defense strategies: Counter modern threats and employ state-of-the-art tools and techniques to protect your organization against cybercriminals*. Packt Publishing Ltd, 2019.

[22] Mikhail Drobyshevskiy and Denis Turdakov. Random graph modeling: A survey of the concepts. *ACM Computing Surveys (CSUR)*, 52(6):1–36, 2019.

[23] Ling Feng, Yanqing Hu, Baowen Li, H Eugene Stanley, Shlomo Havlin, and Lidia A Braunstein. Competing for attention in social media under information overload conditions. *PloS one*, 10(7):e0126090, 2015.

[24] Jerzy Filar and Koos Vrieze. *Competitive Markov decision processes*. Springer Science & Business Media, 2012.

[25] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[26] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.

[27] Andreas Fuchsberger. Intrusion detection systems and intrusion prevention systems. *Information Security Technical Report*, 10(3):134–139, 2005.

[28] Neeraj Garg and Daniel Grosu. Deception in honeynets: A game-theoretic analysis. *2007 IEEE SMC Information Assurance and Security Workshop*, pages 107–113, 2007.

[29] "IT" Global. Security risks survey.(2014). distributed denial of service (ddos) attacks. Technical report, Technical report, Kaspersky. Retrieved from `https://media.kaspersky.com/en/B2B-International-2014-Survey-DDoS-Summary-Report.pdf`, 2014.

[30] Eric Hansen. Solving pomdps by searching in policy space. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 01 2013.

[31] Karel Horák, Branislav Bošansky, Christopher Kiekintveld, and Charles Kamhoua. Compact representation of value function in partially observable stochastic games. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 350–356, 2019.

[32] Karel Horák. Scalable algorithms for solving stochastic games with limited partial observability. *PhD thesis, Czech Technical University in Prague*, 2019.

[33] Karel Horák and Branislav Bosansky. Solving partially observable stochastic games with public observations. *AAAI Conference on Artificial Intelligence*, 33:2029–2036, 2019.

[34] Karel Horák, Branislav Bosansky, Petr Tomášek, Christopher Kiekintveld, and Charles Kamhoua. Optimizing honeypot strategies against dynamic lateral movement using partially observable stochastic games. *Computers and Security*, 87:101579, 07 2019.

[35] Karel Horák, Branislav Bošanský, and Michal Pěchouček. Heuristic search value iteration for one-sided partially observable stochastic games. *International Joint Conference on Artificial Intelligence*, 31:558–564, 2017.

[36] Diana Mertz Hsieh. False excuses: Honesty, wrongdoing, and moral growth. *Journal of value inquiry*, 38(2):171, 2004.

[37] Keman Huang, Michael Siegel, and Stuart Madnick. Systematically understanding the cyber attack business: A survey. *ACM Computing Surveys (CSUR)*, 51(4):1–36, 2018.

[38] Nick Ierace, Cesar Urrutia, and Richard Bassett. Intrusion prevention systems. *Ubiquity*, 6(19):2–2, 2005.

[39] Masaaki Ishikawa. Stochastic optimal control of an sir epidemic model with vaccination. In *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications*, volume 2012, pages 57–62. The ISCIE Symposium on Stochastic Systems Theory and Its Applications, 2012.

[40] Robert Kaiser. The birth of cyberwar. *Political Geography*, 46:11–20, 2015.

[41] Dhruva Kartik and Ashutosh Nayyar. Zero-sum stochastic games with asymmetric information. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4061–4066, 2019.

[42] Jagpreet Kaur and KR Ramkumar. The recent trends in cyber security: A review. *Journal of King Saud University-Computer and Information Sciences*, 2021.

[43] S. Khattak, N. Ramay, K. Khan, A. Syed, and S. Khayam. A taxonomy of botnet behavior, detection, and defense. *IEEE Communications Surveys and Tutorials*, 16(2), 2014.

[44] István Z Kiss, Joel C Miller, Péter L Simon, et al. *Mathematics of Epidemics on Networks*, volume 598. Springer, 2017.

[45] Richard Kissel. *Glossary of key information security terms*. Diane Publishing, 2011.

[46] Jacob C. Koella. On the use of mathematical models of malaria transmission. *Acta Tropica*, 49(1):1–25, 1991.

[47] Constantinos Kolias, Georgios Kambourakis, Angelos Stavrou, and Jeffrey Voas. Ddos in the iot: Mirai and other botnets. *Computer*, 50(7):80–84, 2017.

[48] David M Kreps. Nash equilibrium. In *Game Theory*, pages 167–177. Springer, 1989.

[49] Bhupender Kumar and Bubu Bhuyan. Using game theory to model dos attack and defence. *Sādhanā*, 44:245, 11 2019.

[50] Andrea Landherr, Bettina Friedl, and Julia Heidemann. A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, 2(6):371–385, 2010.

[51] Feng Li, Xinteng Yan, Yunyun Xie, Zi Sang, and Xiaoshu Yuan. A review of cyber-attack methods in cyber-physical power system. In *2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP)*, pages 1335–1339. IEEE, 2019.

[52] Mei Li, Xiang Wang, Kai Gao, and Shanshan Zhang. A survey on information diffusion in online social networks: Models and methods. *Information*, 8(4):118, 2017.

[53] Hung-Jen Liao, Chun-Hung Richard Lin, Ying-Chih Lin, and Kuang-Yuan Tung. Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, 36(1):16–24, 2013.

[54] Yantao Liu and Yasser Morgan. Security against passive attacks on network coding system–a survey. *Computer networks*, 138:57–76, 2018.

[55] Linyuan Lü, Duanbing Chen, Xiao-Long Ren, Qian-Ming Zhang, Yi-Cheng Zhang, and Tao Zhou. Vital nodes identification in complex networks. *Physics Reports*, 650:1–63, 2016.

[56] Somayya Madakam, Vihar Lake, Vihar Lake, Vihar Lake, et al. Internet of things (iot): A literature review. *Journal of Computer and Communications*, 3(05):164, 2015.

[57] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, D. Breitenbacher, A. Shabtai, and Y. Elovici. N-baiot: Network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing, Special Issue - Securing the IoT*, 17(3), 2018.

[58] Emerson Melo. A variational approach to network games. *SSRN Electronic Journal*, 05 2017.

[59] SN Mohammad. Security attacks in manets (survey prospective). *Int. J. Eng. Adv. Technol.(IJEAT)*, 6(3), 2017.

[60] Didier Müller. *Introduction à la théorie des graphes*. Commission romande de mathématique, 2011.

[61] Roger B Myerson. *Game theory: analysis of conflict*. Harvard university press, 1997.

[62] John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

[63] United Nations et al. Combatting cybercrime. Technical report, The World Bank, 2017.

[64] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, 2015.

[65] Animesh Patcha and J-M Park. A game theoretic approach to modeling intrusion detection in mobile ad hoc networks. In *Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, 2004.*, pages 280–284. IEEE, 2004.

[66] Jeffrey Pawlick, Edward Colbert, and Quanyan Zhu. A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. *ACM Computing Surveys*, 52(4), 2019.

[67] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[68] Simona Ramanauskaite and Antanas Cenys. Taxonomy of dos attacks and their countermeasures. *Central European Journal of Computer Science*, 1(3):355–366, 2011.

[69] Mudassar Raza, Muhammad Iqbal, Muhammad Sharif, and Waqas Haider. A survey of password attacks and comparative analysis on methods for secure authentication. *World Applied Sciences Journal*, 19(4):439–444, 2012.

[70] Richard Rivera, Leandro Pazmiño, Fernando Becerra, and Jhonattan Barriga. An analysis of cyber espionage process. In *Developments and Advances in Defense and Security*, pages 3–14. Springer, 2022.

[71] Giorgio Roffo and Simone Melzi. Features selection via eigenvector centrality. *Proceedings of new frontiers in mining complex patterns (NFMCP 2016)(Oct 2016)*, 2016.

[72] Neil Rowe and Julian Rrushi. *Introduction to Cyberdeception*. Springer, 2016.

[73] Neil C Rowe and Julian Rrushi. *Introduction to cyberdeception*. Springer, 2016.

[74] Sankardas Roy, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, Vivek Shandilya, and Qishi Wu. A survey of game theory as applied to network security. In *2010 43rd Hawaii International Conference on System Sciences*, pages 1–10. IEEE, 2010.

[75] Karen Scarfone, Peter Mell, et al. Guide to intrusion detection and prevention systems (idps). *NIST special publication*, 800(2007):94, 2007.

[76] Christian Schneider, Tamara Mihaljev, Shlomo Havlin, and Hans Herrmann. Suppressing epidemics with a limited amount of immunization units. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 84, 02 2011.

[77] Yudhvir Singh, Dheer Dhwaj Barak, Vikas Siwach, and Prabha Rani. Attacks on wireless sensor network: A survey. *International Journal of Computer Science and Management Studies*, 12(03), 2012.

[78] Trey Smith and Reid Simmons. Heuristic search value iteration for pomdps. *Proceedings of UAI*, 07 2012.

[79] Sylvain Sorin. Stochastic games with incomplete information. In *Stochastic Games and applications*, pages 375–395. Springer, 2003.

[80] Leo Spizzirri. Justification and application of eigenvector centrality. *Algebra in Geography: Eigenvectors of Network*, 2011.

[81] Romilla Syed. Cybersecurity vulnerability management: A conceptual ontology and cyber intelligence alert system. *Information & Management*, 57(6):103334, 2020.

[82] Masayuki Takahashi. Equilibrium points of stochastic non-cooperative n-person games. *Hiroshima Mathematical Journal*, 28, 01 1964.

[83] Vijay K Tiwari and Rajeeva Dwivedi. Analysis of cyber attack vectors. In *2016 International Conference on Computing, Communication and Automation (ICCCA)*, pages 600–604. IEEE, 2016.

[84] Juliana Tolles and ThaiBinh Luong. Modeling Epidemics With Compartmental Models. *JAMA*, 323(24):2515–2516, 06 2020.

[85] Petr Tomášek, Karel Horák, Aditya Aradhye, Branislav Bošanský, and Krishnendu Chatterjee. Solving partially observable stochastic shortest-path games. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4182–4189. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Main Track.

[86] Stojan Trajanovski, Yezekael Hayel, Eitan Altman, Huijuan Wang, and Piet Mieghem. Decentralized protection strategies against sis epidemics in networks. *IEEE Transactions on Control of Network Systems*, 2:406–419, 2015.

[87] Stojan Trajanovski, Fernando Kuipers, Yezekael Hayel, Eitan Altman, and Piet Mieghem. Designing virus-resistant, high-performance networks: a game-formation approach. *IEEE Transactions on Control of Network Systems*, 12, 2017.

[88] Olivier Tsemogne, Yezekael Hayel, Charles Kamhoua, and Gabriel Deugoue. Partially observable stochastic games for cyber deception against network epidemic. In *11th International Conference GameSec*, pages 312–325, 2020.

[89] Sun Tzu. The art of war. In *Strategic Studies*, pages 63–91. Routledge, 2008.

[90] M Uma and Ganapathi Padmavathi. A survey on various cyber attacks and their classification. *Int. J. Netw. Secur.*, 15(5):390–396, 2013.

[91] Piet Van Mieghem, Jelena Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Transactions on Networking*, 17(1):1–14, 2009.

[92] Jan Vykopal, Tomas Plesnik, and Pavel Minarik. Network-based dictionary attack detection. In *2009 international conference on future networks*, pages 23–27. IEEE, 2009.

[93] Zelin Wan, Yash Mahajan, Beom Woo Kang, Terrence J Moore, and Jin-Hee Cho. A survey on centrality metrics and their network resilience analysis. *IEEE Access*, 9:104773–104819, 2021.

[94] Cliff Wang and Zhuo Lu. Cyber deception: Overview and the road ahead. *IEEE Security & Privacy*, 16(2):80–85, 2018.

[95] Qiyao Wang, Zhen Lin, Yuehui Jin, Shiduan Cheng, and Tan Yang. Esis: emotion-based spreader–ignorant–stifler model for information diffusion. *Knowledge-based systems*, 81:46–55, 2015.

[96] Christabel Wayllace, Ping Hou, and William Yeoh. New metrics and algorithms for stochastic goal recognition design problems. In *IJCAI*, pages 4455–4462, 2017.

[97] Gabriel Weimann. *Cyberterrorism: How real is the threat?*, volume 119. United States Institute of Peace, 2004.

[98] Kaze Wong, Angus Wong, Alan Yeung, Wei Fan, and Su-Kit Tang. Trust and privacy exploitation in online social networks. *IT professional*, 16(5):28–33, 2014.

[99] Dingyu Yan, Feng Liu, and Kun Jia. Modeling an information-based advanced persistent threat attack on the internal network. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2019.

[100] Guixin Ye, Zhanyong Tang, Dingyi Fang, Xiaojiang Chen, Willy Wolff, Adam J Aviv, and Zheng Wang. A video-based attack for android pattern lock. *ACM Transactions on Privacy and Security (TOPS)*, 21(4):1–31, 2018.

[101] Arif Zaman and George Marsaglia. Random selection of subsets with specified element probabilities. *Communications in Statistics-Theory and Methods*, 19(11):4419–4434, 1990.

[102] Shmuel Zamir. Bayesian games: Games with incomplete information. *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*, pages 119–137, 2020.

[103] Junfeng Zhao, Jing Wang, and Lei Yin. Detection and control against replay attacks in smart grid. In *2016 12th International Conference on Computational Intelligence and Security (CIS)*, pages 624–627. IEEE, 2016.

[104] Xiaohui Zhao, Fang'ai Liu, Jinlong Wang, Tianlai Li, et al. Evaluating influential nodes in social networks by local centrality with a coefficient. *ISPRS International Journal of Geo-Information*, 6(2):35, 2017.