Formalisation et analyse algébrique et combinatoire de scénarios d'attaques généralisées

Cécilia Gallais

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

ARTS ET MÉTIERS ParisTech

2017-ENAM-0064

École doctorale n° 432 : Science des Métiers de l'ingénieur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École Nationale Supérieure d'Arts et Métiers

Spécialité " Mathématiques appliquées et application des mathématiques "

présentée et soutenue publiquement par

Cécilia GALLAIS

le 18 décembre 2017

Formalization and algebraic and combinatorial analysis of generalized attack scenarios

Formalisation et analyse algébrique et combinatoire de scénarios d'attaque généralisés

Directeur de thèse : **Eric FILIOL**

Jury

M. Jean-Marc STEYAERT, Professeur, Laboratoire d'informatique, Ecole Polytechnique	Président
M. Maroun CHAMOUN, Professeur, Faculté d'ingénierie, Université Saint Joseph	Rapporteur
M. Thomas ENGEL, Professeur, CSC, Université du Luxembourg	Rapporteur
M. Johann BARBIER, CTO, Evolution XY	Examinateur
Mme. Antonella SANTONE, Professeure, Département d'ingénierie, Université de Sannio	Examinateur

T H È S E

Arts et Métiers ParisTech - Campus de Angers
N&S Numérique et Société

Acknowledgements

I would first like to thanks my thesis director, Eric Filiol, for the continuous support of my Ph.D study and related research, as well as his patience, motivation, and perseverance. His pieces of advice throughout this thesis have greatly contributed to the work presented in these pages.

This work would not have been possible without the support of TEVALIS and ESIEA Laval, which enabled me, thanks to an employment contract and various financial aids, to devote myself serenely to the development of my thesis.

I would like to thanks the ENSAM school for accepting me in its ranks despite my unusual situation.

I would particularly like to thanks my reviewers, Mr Maroun Chamoun and Mr Thomas Engel, for accepting to review this thesis in such short notice.

I would like to thanks all my TEVALIS co-worker, first Nicolas J. for his support, encouragement and kindness, for answering even my dumbest questions, and for all his sports anecdotes; Lahcen for its original approach to life and things in general, for improving my computing abilities (which was much needed) but not so much for kicking my feet by inadvertence every time he forgets that my desk was in front of him and that he has longer legs than he thinks; Jean-Pierre for its work on the methodology, its work ethic and its attempt to make funny jokes; Nicolas for its hilarious stories and strong opinion (our pre-dinner drinks would not have been the same without him); Christophe for its strong energy; the very discreet Jean-Paul for teaching me a lot about 3D impressions during our English lessons; Zuheir for reminding me that our work could interest other people; Marion, who was sadly not there very long, for being the other girl on board; and Gwen, who was there at the very beginning, for being the first one that put me at ease. What a team.

I would also like to thanks the operational cryptology and virology laboratory and all its members for their incredible welcome. Special kudos to Jean-Pierre who managed to be in two sections here thanks to its life decision, to Paul whose desk was not allowed in the premises of the laboratory for quite a time because of a lack of space (but hey, I have enough for you in these pages), to Richard

and its amazing stories, and finally and not for the least, to Arnaud and Nicolas (both doctors now, sorry Nicolas) who supported me for long days in their office and who quickly replaced me by a plant named in my honor. Thanks guys.

Finally, I would like to thanks my dearest ones. First, my high school girls, Cindy, Emilie, Elodie, Maud, Mathilde, Sabrina and Tiphaine, thank you for telling me that I will make it and that it will be cool to call me "Doctor Gallais". Then my college girls, Audrey and Barbara, thank you for your support and all these crazy karaoke nights on Disney and Celine Dion songs. Special kudos to Barbara who manages to drag me in salsa lessons despite my disinterest in Latin music. Special thanks to Duygu, my Enigma girl, with who I discover cryptography and security for the first time. My crypto gang, Adrien, Malice, Tiffany, Valentin and Vincent, thank you for all our late board games nights. Sarah and Mathieu, thank you for getting me out on Tuesday evenings between cinema and restaurant. My traveling bestie, Zoé, thank you for hosting me in the most exotic places and for suggesting a rereading of this thesis as the English speaker that you are ("but not too many pages in once"). My grandparents, Eugénie, Marie and Victor, thank you for asking for my help and keeping me busy when stress took over. My thoughts go to my grandfather Pierre who saw me start this thesis but who unfortunately did not see me finish it. Special thanks to my sister, Mélanie, for rereading my articles with life changing opinion ("I understand nothing after the introduction but English is okay") and for giving me some more down to earth problems during this thesis (I still want my jumper back when you come back). To my brother, Alexis, thank you for keeping my plants alive (I still do not have any idea of present for your last birthday and next Christmas by the way). My parents, Marie-Anne and Laurent, thank you for hosting me during my last days of writing. And finally, special thanks to Cachou for keeping me company until very late at night, even until the early morning.

Thank you to all of you.

Contents

List of Figures

Chapter 1

Introduction

In 2005, 1.5 million to 2 million customers were deprived of electricity for several hours in Moscow and nearby regions due to a fire and an explosion in a local south-eastern substation [5]. The year 2007 saw a cyber attack against Estonia resulting in the temporary disabling of many of the nation state's critical infrastructures [145]. In 2010, the computer worm Stuxnet was responsible for substantial damages to Iran's nuclear program by targeting its Supervisory Control And Data Acquisition (SCADA) systems [9]. In 2012, the modular computer malware Flame was used for targeted cyber espionage in Middle Eastern countries [10]. In 2013, the company Target suffered a massive cyber attack which caused one of the largest data breaches ever reported, since more than 40 millions of customers had their credit and debit card records stolen from, as well as personal information like email and mailing addresses from some 70 million people [70] [85]. In 2014, a Chinese cyber attack targeted community health systems and compromised the personal data (names, birth dates, Social Security numbers and addresses) of 4.5 million patients [50]. The same year, the attack against Sony Pictures essentially wiped clean several internal data centers and leaked contracts, salary lists, film budgets, entire films, Social Security numbers and emails [88]. In 2015, Anthem, one of the United States of America's largest health insurers, admitted that the personal information of tens of millions of its customers and employees had been compromised because of a database breach [131]. The same year, Crimea witnessed attacks against its power lines [81] that left three quarters of its population without electricity for several days and until several weeks in certain areas [93]. In 2017, the WannaCry ransomware cryptoworm infected more than 230,000 computers in over 150 countries (see Figure 1.1, page 2). Among the victims were Britain's National Health Service (NHS), Spain's Telefnica, FedEx and Deutsche Bahn [67] [55].
These examples of attacks are just a few among many others that infrastructures have experienced in recent years.

Infrastructures are far from being secured today, therefore the main objective of this thesis is to find new ways to evaluate their security. Naturally,

Figure 1.1: Screenshot of a computer infected by the WannaCry ransomware

understanding what is an infrastructure, what are its components and how they interact with each others was the first task we conducted. This led us to the study of the various and many definitions of a critical infrastructure. The existing and current definitions of a critical infrastructure are not adapted to the attacks that can be observed these days. The problem is the same for the definition of an attack and therefore, the term “cyber attack” tends to reduce the conceptual and operational field of the person in charge of the security. Most of the approaches only consider the technical and IT domain, and omit the other domains specific to intelligence. Then, the main methodologies to identify and to manage risk (EBIOS [33] or some similar methodologies) take into account a definition of a critical infrastructure which is restrictive, static and local. The model of attacker and attacks is also extremely narrowed as the technical approaches and the attacker’s angles of attack tend to be restricted to the IT domain only, even if the “cyber” angles may not exist or may only be a small part of an attack scenario.

Therefore, it is necessary to have a new definition of a critical infrastructure, more complete and which is made according to the attacker’s point of view. The security of critical infrastructures is then evaluated by assessing the threats and vulnerabilities. This thesis aims to develop accurately new models

of infrastructure and attack which will based on graph theory, with or without the cyber part. This graph-based representation is already used a lot to describe infrastructure, it will be enriched in order to have a more exhaustive view of an infrastructure environment. The dependencies with other entities (people, others critical infrastructures, etc.) have to be taken into account in order to obtain pertinent attack scenarios. This enriched representation leads to more realistic models of attacker and infrastructure. The main objective is the research of optimal paths or other mathematical structures which can be translated into attack scenarios. This global approach provides a finer (and therefore more realistic) definition of the notion of security, see as the lowest cost of the attack path for example. Therefore, the main objectives of this thesis are:

1. the design of a realistic model of attacker,
2. the design of a general methodology for the assessment of the security,
3. the implementing of the models in the form of a demonstration tool,
4. the validation of the proposed models and algorithms on an existing infrastructure.

The research program is structured in five stages. The first two steps aim to define the models and objects representing the security infrastructures as well as the attackers they are confronted with. The major difficulty encountered in developing a relevant infrastructure model lies in its ability to describe the infrastructure. Indeed, the richer the model is, the more it can describe the infrastructure accurately and the adversaries that attack it. The counterpart of developing a relevant model is its exponential characteristic. In these security models, we therefore expect that the problem of finding the vulnerabilities of a security infrastructure is equivalent to difficult problems, i.e. NP-hard or even NP-complete. The locks to be lifted will therefore consist in the design of heuristics to answer these problems in finite time with an "acceptable" response and sub optimal solution corresponding to admissible attack scenarios.
The third step is to define a generic methodology for assessing the safety of a security infrastructure. This step leads to the design of vulnerability heuristics. This task is not discussed here as it was performed by other people.
In order to validate the proposed models and methodology, a research demonstrator is developed in the form of an evaluation platform.
Finally, the last step will be to evaluate an existing system from the platform by implementing the proposed methodology. The objective of this last step is to validate the models and the methodology and and to propose improvements if necessary.

This thesis is part of a company project called InfraSec which aims to help infrastructure face multiform threats. Concretely, InfraSec is a security audit tool designed to enable companies to measure their risk exposure and to anticipate attacks by identifying attack patterns. It is the combination of an intelligence

methodology and a modeling and calculation tool. The audit methodology collects information on the ecosystem of the audited firm. This information is then injected into the eponymous tool to model the infrastructure with all its components (human, technical, organizational, etc.). The result is a clear and relevant mapping which, combined with complex algorithmic calculations, allows the identification of the vital components of the targeted infrastructure as well as the most efficient attack patterns (in terms of difficulty, cost and time). The InfraSec project was well received by the community, during international forums (FIC 2015) as well as in front of panels of experts (presentation to the Directorate General of Armaments MI, a technical expertise center for the French army for instance).

This thesis is divided into two parts. The first focuses on infrastructure modeling and includes Chapters 2, 3 and 4. The second is concerned with how to evaluate the security of an infrastructure using the chosen model and includes Chapters 5, 6 and 7.
Chapter 2 presents the definitions that are indivisible of the topic of infrastructures' security, in order to fully understand what is actually an infrastructure and what it faces. Some of these definitions are our own. Chapter 3 discusses some of the existing infrastructure models. Chapter 4 presents our model of infrastructure used in the InfraSec project. Chapter 5 then introduces the notion of connectivity. Finally, Chapters 6 and 7 outline attack patterns, i.e. mathematical structures that we study to see if they can be used to build operational attack scenarios whose features allow to evaluate the security of an infrastructure. Some of the main algorithms used for the demonstrator are also presented.

An index of all definitions is given at the end of the thesis, as well as a bibliography.

Part I

How do we model infrastructures?

Introduction

The many attacks presented in the introduction of this thesis are only a fraction of the attacks that have taken place in recent years. Therefore, infrastructures are still far from being protected today and it is necessary to find new ways of evaluating their security. In view of all the information that can now be found on infrastructures, a tool for representing and processing this information is necessary in order to process it as efficiently as possible. To achieve this, a model of infrastructure must be defined.

Before modeling an infrastructure, it is important to know exactly what to model. This led us to the study of the various and many definitions of a critical infrastructure and to the writing of a new definition of a critical infrastructure. Thanks to this study, the observation that the term cyber is predominant, whether it is in the definitions of a critical infrastructure or in the definitions of an attack will be made, and the consequences of this predominance on infrastructure security will be presented.

In this part, we will present the definitions of the key concepts of infrastructure security (critical infrastructure, attack, dependency and resilience among others), and explain how the predominance of the cyber term in these definitions can have serious consequences for the security of an infrastructure. We will also present some of the existing models of infrastructure, mostly the models that depict the attacks and the models that depicts the infrastructure, and we will explain why we chose a graph-based model that depicts the infrastructure for the InfraSec project. Then the chosen model of infrastructure will be presented.

To this end, Chapter 2 will attempt to define the key concepts of infrastructure security. Chapter 3 will then discuss some of the existing infrastructure models. Finally, Chapter 4 will detail the model selected for the InfraSec project, a graph-based model that depicts the infrastructure.

Chapter 2

Infrastructure security and operational notions

The attacks presented in the introduction of this thesis prove that infrastructures today are far from being secure and this partially comes from the fact that the concept of critical infrastructure used nowadays is still too poor. Before proposing new solutions to improve infrastructures security, it is necessary to know and understand what is a critical infrastructure and what they face. For this purpose, a presentation and explanation of some key notions of the infrastructures security is needed.

First, it should be noticed that every attack of this introduction are presented as cyber attack in the press and every online website. We will see in the first section of this chapter how this tendency to favor the cyber aspect of attacks can have consequences on an infrastructure security, since it provides a strong bias.

2.1 The predominance of the cyber aspect in security thought

The following surveys are made to emphasize the idea that security tends to be seen mostly through its computerized aspect.

2.1.1 How is a critical infrastructure defined according to nation states and organizations?

National plans for the protection of critical infrastructures thrive pretty much everywhere: Australia [57], Canada [17], Japan [28], Germany [117], United

States of America, the European Union [115] and even Africa with nation-states like Mauritius [107] and Kenya [94], which proves the importance of the concept of 'critical infrastructure' in contemporary security thought. Most of these plans include a definition of critical infrastructure, as defining it is the first logical step before protecting it.

The first definition of 'critical infrastructure' appeared in the United States' Presidential Decision Directive (PDD) 63 dating back to 1998 [68]. At this time a critical infrastructure consisted of physical and cyber-based systems that were essential to the minimum operations of the economy and the government. Since the appearance of this initial definition, several others have followed and despite the variety and the great number of definitions, none of them gives a complete and accurate description of what constitutes a critical infrastructure, as some important components are not mentioned.
To highlight these omissions, we compiled a survey of the definitions of critical infrastructure. The survey is principally based on the *International CIIP handbook of 2008/2009* [16] and its previous versions: 2002 [150], 2004 [40] and 2006[1]. Even if the subject of these documents is the critical information infrastructure, which can be seen in broad outline as a part of critical infrastructure and "refers exclusively to the security and protection of the IT connections and IT solutions within and between the individual infrastructure sectors" [53], critical infrastructures are mentioned and several definitions are provided. It is not a surprise since there is no official distinction between critical infrastructure and critical information infrastructure. Even the terms become interchangeable in some countries. The survey is also based on a document from the Organization for Economic Co-operation and Development [51].

Twenty five definitions are presented here, including two from African nation-states, one from Arabic nation-states, five from Asian and Pacific nation-states and organizations, five from American nation-states and organizations, and twelve from European nation-states and organizations.
The definitions are the most recent ones that can be found. But, despite the great amount of documents on critical infrastructure (especially on the protection of such), it is quite difficult to be sure that the adopted definitions are really the most recent ones.

- **The Asia-Pacific Telecommunity** - The Asia-Pacific Telecommunity, with the report of the South Asian Telecommunication Regulators Council (SATRC) named *Critical information infrastructure protection and cyber security* and adopted between the 18th and 20th of April 2012, defines critical infrastructure as "the computers, computer systems and/or networks, whether physical or virtual, and/or the computer programs, computer data, content data and/or traffic data so vital to a country that the incapacity or destruction of or interference with such systems and assets would have a debilitating impact on security, national or economic security, national public health and safety, or any combination of those matters" [124].

- **Australia** - The Attorney-General's Department website says that "critical infrastructure delivers services essential to our daily lives, such as power, water, health services, communications systems and banking" [113].
The Trusted Information Sharing Network's website provides more information as it defines Australian critical infrastructure as "those physical facilities, supply chains, information technologies and communication networks, which if destroyed, degraded or rendered unavailable for an extended period, would significantly impact on the social or economic wellbeing of the nation, or affect Australias ability to conduct national defense and ensure national security" [144].
The same definition can be found in the Critical Infrastructure Resilience Strategy [57].

- **Austria** - The Austrian federal chancellor defines critical infrastructures as "natural resources; services; information technology facilities; networks; and other assets which, if disrupted or destroyed would have serious impact on the health, safety, or economic well-being of the citizens or the effective functioning of the Government" [43].

- **Belgium** - As stipulated in the law of the first of July 2011, a critical infrastructure is an installation, system or part thereof, of federal interest, which is essential for the maintenance of vital societal functions, health, safety, security, economic or societal well-being of people, and which, if disrupted or destroyed, would have a significant impact [72].

- **Canada** - The National Strategy for Critical Infrastructure defines critical infrastructure as "processes, systems facilities, technologies, networks, assets and services essential to the health, safety, security or economic well-being of Canadians and the effective functioning of government. Critical infrastructure can be stand-alone or interconnected and interdependent within and across provinces, territories and national borders. Disruptions of critical infrastructure could result in catastrophic loss of life, adverse economic effects, and significant harm to public confidence" [17].
This definition can also be found on the Public Safety Canada's website [97].

- **Colombia** - Critical infrastructure are "the array of computers, computer systems, and telecommunications, data, and information networks, whose destruction or interference could weaken or impact on the security of a country's economy, public health, or both". [99]

- **The European Union** - The Council Directive 2008/114/CE defines critical infrastructure as "an asset, system or part thereof located in Member States which is essential for the maintenance of vital societal functions, health, safety, security, economic or societal well-being of people, and the disruption or destruction of which would have a significant impact in a Member State as a

result of the failure to maintain those functions" [116].

- **Germany** - The National Strategy for Critical Infrastructure Protection defines critical infrastructure as the "organizational and physical structures and facilities of such vital importance to a nation's society and economy that their failure or degradation would result in sustained supply shortages, significant disruption of public safety and security, or other dramatic consequences" [119].

- **Hungary** - The Hungarian definition of a critical infrastructure is based on the definition of the European Union : "critical infrastructures are the interconnected, interactive, and interdependent infrastructure elements, establishments, services, and systems that are vital for the operation of the national economy and public utilities to maintain an acceptable level of security for the nation, individual lives, and private property, as well as concerning the maintenance of the economy, the public health services, and the environment" [16].

- **Jamaica** - Critical infrastructure "include systems and assets, whether physical or virtual, so critical that the incapacitation or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination thereof". [106]

- **Japan** - Critical infrastructure is defined by the Second Action Plan on Information Security Measures for Critical Infrastructures. According to this plan, "critical infrastructure is the basis of people's social lives and economic activities formed by business that provide services which are extremely difficult to be substituted by others. If its function is suspended, deteriorated or become unavailable, it could have significant impacts on people's social lives and economic activities" [29]).

- **Kenya** - The Cabinet Secretary Interior and Co-ordination of National Government defines a critical infrastructure as "the totality of critical infrastructure assets"; the critical infrastructure assets are the "designated physical and virtual assets or facilities, whether owned by private or public entities which are designated as such under this Act as essential to the provision of vital services to Kenyans for their social and economic wellbeing, and which if destroyed, degraded or rendered unavailable, would impact on the social or economic wellbeing of the nation or affect Kenyas ability to conduct national defense and security" [94].

- **Latvia** - "Critical infrastructure is the objects, systems, or their parts, which are important in providing the performance of functions essential to society, as well as for ensuring the protection of human health, security, economic or social welfare, whose destructruction or malfunctioning may significantly influence the performance of state functions." [77]

- **Malaysia** - A critical national information infrastructure is "those assets

(real and virtual), systems and functions that are vital to the nations that their incapacity or destruction would have a devastating impact on:

1. National economic strength; Confidence that the nation's key growth area can successfully compete in global market while maintaining favorable standards of living;

2. National image; Projection of national image towards enhancing stature and sphere of influence;

3. National defense and security; guarantee sovereignty and independence whilst maintaining internal security.

4. Government capability to functions; maintain order to perform and deliver minimum essential public services;

5. Public health and safety; delivering and managing optimal health care to the citizen."

[129].

- **The Netherlands** - "Critical infrastructure includes the business enterprises and public bodies that provide the goods and services essential for the day-to-day lives of most people in the Netherlands" [120].
This definition seems to be forgotten since the article of ICCWS-14 [45]. Indeed it is no longer on the government's website. Another definition was found in "Securing Critical Infrastructures in the Netherlands". It defines the critical infrastructures according to European Commissions definitions: critical infrastructures are those physical and information technology facilities, networks, services and assets which, if disrupted or destroyed, have a serious impact on the health, safety, security or economic well-being of citizens or the effective functioning of governments [35].

- **New Zealand** - A presentation of the International Disaster and Risk Conference in Davos, defines critical infrastructure as "infrastructure necessary to provide critical services, whose interruption would have a serious adverse effect on New Zealand as a whole or on a large proportion of the population, and which would require immediate reinstatement" [127].

- **The North Atlantic Treaty Organization** - During a session about the protection of critical infrastructures in 2007, the NATO parliamentary assembly admitted there is no universally agreed definition of a critical infrastructure. But this term "is generally understood as those facilities and services that are vital to the basic operations of a given society, or those without which the functioning of a given society would be greatly impaired" [6].

- **Norway** - The report NOU 2006:6, about the protection of critical infrastructures and critical societal functions, defines critical infrastructures as

"the facilities and systems that are necessary to maintain the functions that are critical for society. These functions cover basic needs in the society and contribute to a sense of safety in the population" [101].

- **Poland** - The National Critical Infrastructure Protection Programme defines critical infrastructure according to the Act on Crisis Management : critical infrastructure shall be understood as "the systems and functional sites forming their part which are mutually related, such as building sites, facilities, installations, key services for the safety of the state and its citizens and serving to ensure efficient functioning of the public administration authorities, as well as institutions and entrepreneurs" [12].

- **Quatar** - Critical infrastructure are those "physical assets, systems or installations, which if disrupted, compromised, or destroyed, would have a serious impact on the health, safety, security, or economic well-being of Qatar or the effective functioning of the Qatari government". [130]

- **South Africa** - Critical information infrastructure includes "all ICT systems, data systems, data bases, networks (including people, buildings, facilities and processes), that are fundamental to the effective operation of the State". [110]

- **Spain** - The Law 8/2011 defines critical infrastructure as those installations, networks, systems, physical equipment, and information technologies, whose interruption or destruction would have a grave impact on the health, security, social or economic well-being of citizens or on the efficient functioning of the state institutions and of the public administration [34].

- **Switzerland** - The Federal Councils Basic Strategy for Critical Infrastructure Protection defines critical infrastructure as "infrastructures whose disruption, failure, or destruction would have a serious impact on public health, public and political affairs, the environment, security, and social and economic well-being" [112].
The same definition can be found in a more recent article written by Stefan Brem [14].

- **The United Kingdom** - The United Kingdom's critical national infrastructure is defined by the Government as "those facilities, systems, sites, information, people, networks and processes, necessary for a country to function and upon which daily life depends. It also includes some functions, sites and organizations which are not critical to the maintenance of essential services, but which need protection due to the potential danger to the public (civil nuclear and chemical sites for example)" [52].

- **The United States of America** - The USA Patriot Act of 2001 defines critical infrastructure as "systems and assets, whether physical or virtual, so vi-

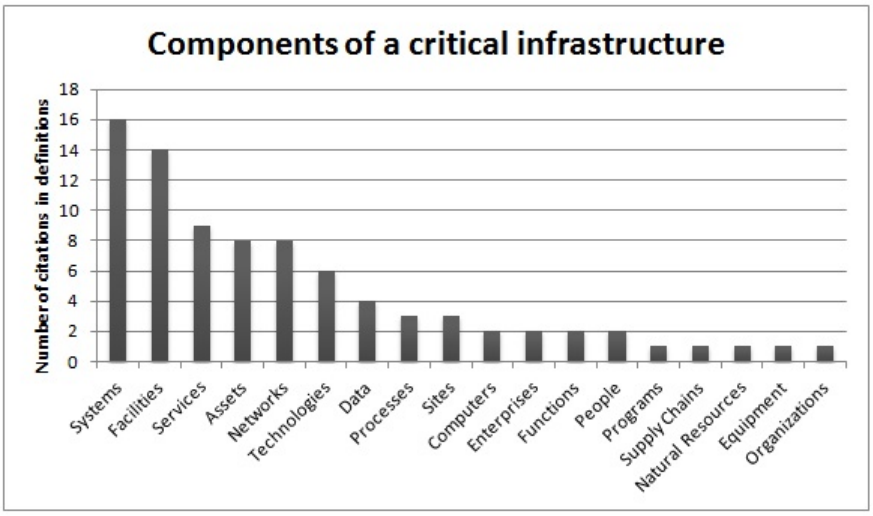

Figure 2.1: Histogram of the cited components in the definitions of a critical infrastructure

tal to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters" [22]. According to the Homeland Securitys website, this definition is still currently used [103].

Figure 2.1 (page 15) shows a histogram of the cited components and their number of citations in the previous definitions. In order to make the histogram clearer and then easier to understand, some components have been classified under the same name because they designate the same kind of elements. For example, structures, installations, equipment and buildings are counted as facilities, as well as information is counted for data.

Now that the definitions of critical infrastructure have been presented, we can study them and present our conclusion.

2.1.2 The lacks of the actual definitions of a critical infrastructure

There are more and more definitions of a critical infrastructure and they have undergone many modifications and will certainly undergo others since more and more actors understand and grasp the importance of this concept. Even companies have their own definition like Elia in Belgium [41].

The identified components of a critical infrastructure

The definitions of a critical infrastructure are usually divided into the list of its components and the consequences of its disruption, damage, or destruction. The list of the components is the part of the definition which differs the most from nation-state to nation-state.
Many components of a critical infrastructure were identified in the different definitions presented above, including assets, systems, or networks (see Figure 2.1, page 15). The list of these components tends to be reduced with time. For example, the *2005 Green paper on a European programme for critical infrastructure protection* [115] gives a more complete definition than the one from the Council Directive 2008/114/CE [116].

Missing components

Among all the components cited in the various definitions, the absence of human components is perhaps most immediately noticeable. Almost none of the survey's definitions mentions humans as part of a critical infrastructure, although humans are essential for the functioning of every existing infrastructure, critical or not. We define human components as the staff and the human factors which are defined according to the Clinical Human Factors Group (CHFG) as "the environmental, organisational and job factors, and individual characteristics which influence behaviour at work" [140].

The United Kingdom and South Africa are the only nation-states that clearly includes the human component as a component of a critical infrastructure (see Figure 2.1, page 15). And the British definition did not include it until recently according to a precedent survey published in 2014 [45] [48]. At this time the United Kingdom's national infrastructure was defined by the Government as "those facilities, systems, sites and networks necessary for the functioning of the country and the delivery of the essential services upon which daily life in the UK depends".

Some may say that a system, which is a component whose presence is acknowledged in many definitions, could be defined as being comprised of people, processes, and technology. But human components are still not clearly stated as a component of a critical infrastructure, and thus the definitions can mislead those in charge of critical infrastructure's security with respect to the importance of people.

Also of note is the lack of some 'intelligence' perspective, thus allowing a broader and more operational view as far as the cyber dimension is concerned. As an example, no existing definition takes interdependencies with external components into account, thus providing only a very narrow-minded view of an infrastructure, which is considered then only as a completely isolated structure.

Indeed, even though some of the definitions mention the concept of interdependency, such as Canada's, Hungary's, and Poland's, the interdependencies taken into account are only the ones within the critical infrastructure itself or with other critical infrastructures, but never with basic infrastructures including subcontractors, suppliers, data-centers, or others.
It would also be better if the critical infrastructure's environment were taken into account, especially the political and cultural environments. Attackers could use these environments to trigger a strike, for example, which could disturb the transport of needed resources or finished products.

These omissions were previously mentioned by Eric Filiol in an article called "*Operational Aspects of a Cyberattack: Intelligence, Planning and Conduct*" and presented in "*Cyberwar and Information Warfare*" [146].

2.1.3 How is an attack (or cyber attack) defined according to nation states and organizations?

Eighteen official definitions are presented in this section, including one from Arabic nation-states, one from Asian and Pacific nation-states and organizations, six from American nation-states and organizations, and ten from European nation-states and organizations.
The definitions are the most recent ones that can be found.

- **Austria** - "The term 'cyber attack' refers to an attack through IT in cyber space, which is directed against one or several IT system(s). Its aim is to undermine the objectives of ICT security protection (confidentiality, integrity and availability) partly or totally". [121]

- **Belgium** -
"A cyber-attack is often a combination of technical possibilities and of social engineering exploiting the habits and credulity of the victim." [95]

- **Bosnia and Herzegovina** - "Cyber attacks can be planned as to target the key infrastructure of any country, overload communication systems and cause severe consequences on the security system of the country under attack". [13]

- **Canada** - "Cyber attacks include the unintentional or unauthorised access, use, manipulation, interruption or destruction (via electronic means) of electronic information and/or the electronic and physical infrastructure used to process, communicate and/or store that information. The severity of the cyber attack determines the appropriate level of response and/or mitigation measures: i.e. cyber security". [98]

- **Colombia** - A cyber attack is "an organized and/or premeditated act by one or more persons to harm or cause problems to a computer system via cyberspace". (Ministry of Defense of Colombia) [99]

- **Cyprus** - "Cyber-attacks are a very real and ever-increasing threat. Whether against individual countries, companies or most recently against the European Commission, they can paralyse key infrastructure and cause huge long-term damage". [114]

- **Czech Republic** - "Cyber attacks directed against the public as well as private sectors are increasingly frequent and sophisticated. They can cause in particular failures of communication, energy and transport networks, transport processes and industrial and financial systems, resulting in considerable material damage. The armed forces' dependence on information and communication systems may affect the defence capability of the state. Another problem closely associated with cyber attacks is political and economic espionage". [100]

- **Germany** - "A cyber attack is an IT attack in cyberspace directed against one or several other IT systems and aimed at damaging IT security. The aims of IT security, confidentiality, integrity and availability may all or individually be compromised". [118]

- **Jamaica** - A cybercrime is "a crime in which a computer is the object of the crime or is used as a tool to commit an offence". [106]

- **The North Atlantic Treaty Organization (NATO)** - A cyber attack is an "action taken to disrupt, deny, degrade or destroy information resident in a computer and/or computer network, or the computer and/or computer network itself. Note: A computer network attack is a type of cyber attack". [91]

- **The Netherlands** - "Cyber attacks consist of different activities, such as making use of malware, social engineering, overloading processes, hardware and software weaknesses, physical attacks, and electromagnetic attacks. They are performed to sabotage or steal information from a particular computer system or render it dysfunctional. Examples of attacks are Distributed Denial of Service (DDoS) attacks, Trojans, Structured Query Language (SQL) injections, Bot-Network attacks and Zero-Day exploits. These sort of attacks can be orchestrated by hackers, cyber criminals, hacktivists, competitors, other nation states, and amateurs. Since the skills required to program and carry out these attacks have become decentralized and more accessible, attackers can pick up these skills quite quickly and can make them available to others in the form of script kiddy tools. The range of cyber-attacks can vary from simple attacks that require a few clicks, to very complex ones that require thousand of coding hours, and capital investments in logistics and acquiring Zero-days. The consequences of these attacks can also range from information leakages, irregular machine activity, to more severe consequences such as machinery coming to self-destruct

and cause real damage.
The risk depends on the intent and sophistication of the attacker. Cyber attacks can be classified as acts of illegal intrusion, theft, excessive protest, sabotage, espionage, and in some instances even as acts of war".

- **New Zealand** - "A cyber attack is an attempt to undermine or compromise the function of a computer-based system, access information, or attempt to track the online movements of individuals without their permission". [59]
"Cyber risks include state-sponsored espionage, cyber vandalism or issue-motivated hacktivism, a broad range of cybercrime (e.g. scams and fraud), and deliberate or inadvertent actions by employees or contractors. Malicious cyber actors are constantly changing their methods and tactics, often re-emerging in different guises or exploiting vulnerabilities before they are patched. They can act stealthily and anonymously online, leaving few clues, and operating from any Internet-connected location globally. This makes it hard to distinguish between the actions of state-sponsored cyber intruders, organised cyber-criminal groups or an isolated computer hacker". [60]

- **Poland** - "A cyber attack is an intentional disruption of the proper functioning of cyberspace which is defined as a space of processing and exchanging information created by the ICT systems". [109]

- **Quatar** - A cyber crime is a "misconduct or crime committed using technology. Examples of cyber crime may include illegal access to systems or information, fraud, identity theft, or content-related offenses such as spam". [130]

- **Spain** - "Cyberattacks, whether in the form of cyberterrorism, cybercrimes/cyberoffences, cyberespionage or hacktivism, have become a powerful instrument for attacking individuals and public and private institutions. Factors such as their low cost and minimal risks to the attacker and their easy use, effectiveness and accessibility explain why the phenomenon is spreading. These illegal attacks are perpetrated and increasingly frequently by terrorist groups, organised crime networks, companies, States or individuals". [141]

- **Switzerland** - "Cyber-attacks are carried out on computers, networks and data. They are aimed at disrupting the integrity of the data or the functioning of the infrastructure and restricting or interrupting their availability. They also seek to compromise the confidentiality or authenticity of information by means of unauthorized reading, deletion or modification of data, connections or server services are overloaded, information channels spied upon or surveillance and processing systems are manipulated in a targeted manner".

- **The United Kingdom** - "An attack is the manifestation of a threat, which is defined as a potential cause of harm to an asset. A threat exploits a vulnerability to impact an asset". [96]

Nation State	Attack	Cyber Attack
Austria		×
Belgium		×
Bosnia and Herzegovina		×
Canada		×
Colombia		×
Cyprus		×
Czech Republic		×
Germany		×
Jamaica		×
NATO		×
The Netherlands		×
New Zealand		×
Poland		×
Quatar		×
Spain		×
Switzerland		×
The United Kingdom	×	
The United States of America	×	

Figure 2.2: Summary of the choice of term used in the definitions.

- **The United States of America** - "An attack is an attempt to gain unauthorized access to system services, resources, or information, or an attempt to compromise system integrity.
It is also, according to the same source, any kind of malicious activity that attempts to collect, disrupt, deny, degrade, or destroy information system resources of the information itself". [111]
"An attack is an actual assault perpetrated by an intentional threat source that attempts to alter a system, its resources, its data, or its operations". [83]

Figure 2.2 (page 20) shows a table which indicates for each country if an official definition of an attack or an official definition of a cyber attack was found. Now that the definitions have been presented, we can study them and present our conclusion.

2.1.4 The large predominance of the term "cyber attack"

Eighteen official definitions are presented in the previous section and sixteen of them are a definition of a cyber attack or a cyber crime (see Figure 2.2, page 20). It was possible to find an official definition of an attack, and not of a cyber attack or a cyber crime, for only two nation states: the United Kingdom and the United States of America.

We do not imply that such a definition does not exist for these sixteen nation states and organizations, but that we did not find it in national plans or other official documents, as it was for the definition of a cyber attack. The term cyber is omnipresent in the official documents and not only there. When you search the Internet for examples of attacks against critical infrastructures, the first pages of results are all cyber attacks. This omnipresence can lead to the wrong assumptions that cyber attacks are the only threats for critical infrastructures and that the other ones are irrelevant today. Many people may know that it is not the case in reality but it could not be that obvious for some people in charge. And right now, the official documents, the Internet and even most of the security's solutions that are proposed today can lead them to these wrong assumptions. Cyber attacks may be too popular for the good of infrastructures' security.

It is worth noticing that among all the cited ways an attacker can use to harm an infrastructure, some of the definitions mention social engineering when the human factor is barely mentioned in the definitions of a critical infrastructure. It turns out that many cyber attacks are in fact phishing attacks : 90% according to studies led by two companies, Solucom and Conscio Technologies, in 2015 [56]. A recent example is the WannaCry ransomware which used phishing to propagate [55].

2.1.5 What can be the consequences of a cyber oriented approach?

As stated previously, the human component is missing from almost all the definitions which are presented above, despite the fact that people are essential for the functioning of critical infrastructures. And most of all, the term cyber attack tends to be overexposed even though the observed attacks did not rely only on cyber elements. Mitnick and Simon consider humans to be the weakest link of security [86]. Their work demonstrates that, despite the use of the best possible security protection items, it is possible for an attacker to obtain access to critical information or critical objects just by using social engineering techniques.
As an example, Mitnick and Simon show how an attacker, or a manipulator in this case, can get a username and the corresponding password just by asking its owner after pretending to be part of the information security office. And with this user name and password, the manipulator has everything he or she needs to get into the company's network and to locate the elements he or she is looking for.
A recent event perfectly illustrates this. In 2016, a hacker managed to get into the FBI's servers, and was able to have access to a terabyte of data, from which he extracted the contact details of nearly 20,000 FBI employees, and 9000 internal security employees. To succeed, he did not have to use its computer skills, but rather abused the confidence of some employees of the US government [30].

If some of these stories are not persuasive enough because they are fictional, still based on Mitnicks experience, the Snowden [61] and Wikileaks [147] real cases show that humans can be a major flaw in any security scenario. In these cases, however, not much can be done as it is difficult to prevent employees from giving confidential information of their own free will, unlike the cases presented by Mitnick and Simon that can be avoided.

It is interesting to notice that a definition of an infrastructure, dating back to 1996, briefly mentions this component. Indeed the Executive Order 13010 defined infrastructure as "the framework of interdependent networks and systems comprising identifiable industries, institutions (including people and procedure), and distribution capabilities that provide a reliable flow of products and services essential to the defense and economic security of the United States, the smooth functioning of government at all levels, and society as a whole" [89]. Since then, the human component has been put aside in definitions.

The external components and the environment of the infrastructure are also missing from the definitions above. However, Filiol and Raynal have planned an attack which uses these components to delay the departure of a military ship [49]. Instead of considering the military ship as an isolated structure and focusing their attack on it, they have targeted its suppliers. They also used human factors targeting.

Indeed, instead of directly attacking the ship and its computerized systems, they preferred to target and use the components found missing in the definitions presented in this paper : the political and social environment. They triggered a strike of the employees of its oil supplier and a riot to stop the delivery of helicopter's pieces (instead of considering the military ship as an isolated structure and focus their attack on it). They also used the human component by falsely incriminating the captain of the military unit which was supposed to embark on the ship.

In these examples, the missing elements of the definitions allow the attacks to succeed. If these two components do not appear in official definitions, the danger is that they may not be taken into account in security policies; and then infrastructures may not be protected as they should be. Attackers always use the weakest link to reach their goals.

2.2 Definitions

The following notions are indivisible of the topic of infrastructure security and therefore, it is necessary to define them clearly.

2.2.1 Critical Infrastructure - Discussion towards an enlarged and more suitable definition

The existing definitions of a critical infrastructure appear restrictive, static, and local as they are mainly dictated by the defender point of view. So, in order to have a more complete and realistic definition, the following statements are dictated by the attacker point of view.

A **critical infrastructure** can be a company, an institution, an organization, facilities [122], services, or equipment, whether regional, national, or international, which, if disrupted, damaged, or destroyed, would have a serious impact on the health, safety, security, or economic well-being of citizens or on the effective functioning of governments and others infrastructures depending on it.
It includes any element which would have a serious impact on the health, safety, security, or well-being of a population (including employees) or could lead to the disruption, damage, or destruction of the critical infrastructure and have a serious impact on its effective functioning.

Components of a critical infrastructure

More precisely, a critical infrastructure includes people, which, if co-opted, diverted, or eliminated, could lead to the disruption, damage, or destruction of the critical infrastructure.
It also includes (non-exhaustive list):

1. installations (such as access, buildings, sites),
2. facilities,
3. structure,
4. property,
5. equipment (for example, computers, printers, hard drives),
6. unobligated or unexpended balances of appropriations,
7. funds;
8. resources, whether physical or natural,
9. material,
10. networks, whether physical (like electricity or water) or virtual (such as the Intranet or the Internet),
11. information/data, whether physical or virtual (confidential data, like passwords or access codes, procedures, organization charts, contracts),

12. information and communication technology facilities,
13. services,
14. processes,
15. the corporate image,
16. systems or part thereof,
17. others infrastructures with which strong dependencies exist (suppliers of services or products, for example) [49].

These elements can also be found in the political and cultural environment of the infrastructure.

The elements of an infrastructure are defined as **components**, or sometimes considered as **assets**. However, in some context, people are not considered assets (the national infrastructure protection plan of the United States of America for example[105]). Then, the term component is favored to the term asset in the following pages.

Some components of the infrastructure can be particularly critical. They are in a way the weakest components, security-wise, of the infrastructure. These components are said to be vulnerable and a potential target of an attacker. They are called critical components of the infrastructure.

Surely all the infrastructures components which can lead to its disruption, damage, or destruction cannot be identified and enumerated. Indeed, the task seems impossible since the security aspect lies in the ability of the attackers to be innovative, creative, and, in essence, unpredictable. They can turn a component that is thought to be inoffensive into a weapon. This may explain why some definitions, such as the Swiss [112] and the Dutch [35] ones, really do not go into details regarding the components of a critical infrastructure.

Distinction between a critical and a basic infrastructure

The distinction between a critical and a basic infrastructure is an element which appears clearly in all these definitions: it lies in the criticality of the consequences of their disruption, damage, or destruction. The definitions differ when it comes to the domains on which the disruption or destruction have serious consequences, although some of them appear frequently, such as public safety, public security, or the social and economic well-being of the citizens. But the notion of criticality is always there.

The criticality of an infrastructure depends as much on the infrastructure itself as on its relations with other infrastructures. So two kinds of criticality

are identified: the **inherent criticality** that occurs when an infrastructure is critical in and of itself, and the **external criticality** when an infrastructure is critical for other infrastructures due to some dependencies.
Most of the time, infrastructures are qualified as external critical infrastructures only after a first disaster because it is really difficult, even impossible, to predict what can be the consequences of its disruption, damage, or destruction on other infrastructures. So this qualification is mostly done *a posteriori*.
Identifying the inherent critical infrastructure seems to be less difficult, as some infrastructures such as the energy's suppliers come to mind instantaneously. The lists of critical sectors seem to be useful in identifying the inherent critical infrastructure. But nothing proves that all inherent critical infrastructures can be found this easily.

In that respect, the distinction between a critical and a basic infrastructure can be made most of the time only *a posteriori*.

2.2.2 Critical sectors - Where are most critical infrastructures found?

The critical sectors appear in a lot of documents related to critical infrastructures' protection. Then, it seems necessary to develop this subject.

As stated in national plans for the protection of critical infrastructures, the role of the **critical sectors** is to "facilitate identification, prioritization, assessment and protection of critical information infrastructure through information sharing and reporting" [107]. Therefore, it is not a surprise when a definition of a critical infrastructure is almost always followed by a list of critical sectors. The first mention of critical infrastructure sectors was found in the Executive Order 13010 of July 15, 1996 [20]. This list identified the sectors which were necessary to the effective functioning of the society.
The list of critical sectors tends to be specific to each nation-state or organization. A sector may be included for historical, geographic, socio-political, or cultural reasons, which can explain the differences between the lists of critical sectors. Forty-six lists are presented in the appendix A from nation-states such as the U.S.A., Germany, Switzerland, India, and Kenya and organizations such as the European Union and the Asia-Pacific Telecommunity.

It also appears that some nation-states have a list of critical sectors while they do not have a definition of a critical infrastructure. In the case of the members of the European Union, they may not have a definition of critical infrastructure of their own because they probably content themselves with the European Union's definition. For the others nation-states, it is not that easy and obvious to explain the absence of definition.

On the contrary, some nation-states such as Austria do not have an official list of critical sectors [43], the one given in this paper was developed by some experts and it is not an official definition from the Austrian government, while it does have a definition of a critical infrastructure.

Identification of the critical sectors

Despite the great variety of lists of critical sectors, most of the nation-states and organizations seem to agree on certain critical sectors. As can be seen in the histogram below, the transport sector is mentioned in more than 95% of the lists of critical sectors that were gathered, the energy sector in more than 86%, and the sector of communication technology in more than 84%. They are closely followed by the sectors of finance and water.
In the histogram of Figure 2.3, page 26, C.T. stands for Communication Technology, I.T. for Information Technology, E. S. for Emergency Services and Gov. for Government.

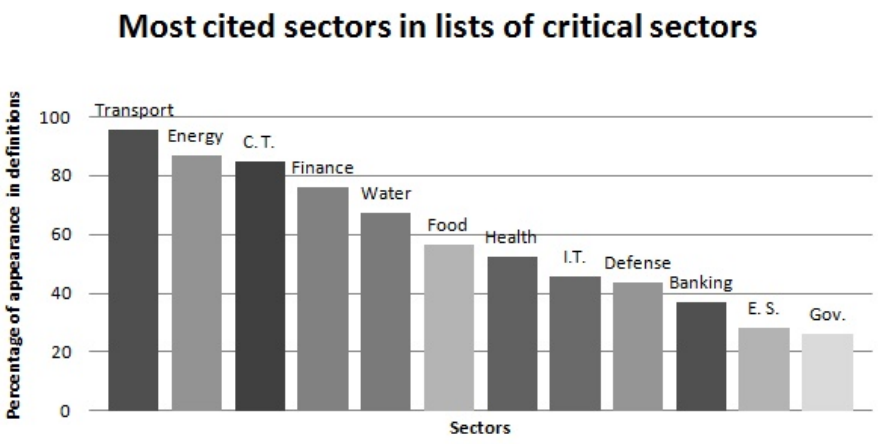

Figure 2.3: Histogram of the most cited sectors in lists of critical sectors

The importance taken by the critical sectors

Many nation-states, like Estonia [42], Finland [102], France [58], Italy [2], or Sweden [16], have a list of critical sectors but no official definition of a critical infrastructure. And a few others have proposed a definition many years after they presented a list of critical sectors.
Moreover, the list of critical sectors tends to have more modifications than the definition: sectors are added or removed, their names are changed, subsectors

are present or not, sector becomes subsector, or subsector becomes sector (emergency services, for example). And when definitions of critical infrastructure tend to be simpler, lists of critical sectors get more complex, mostly with the addition of subsectors. So, establishing the list of critical sectors seems to have prevailed over the definition of a critical infrastructure.

Therefore, the primary aim of the nation-states may be to identify quickly their critical infrastructures with the use of a list of its critical sectors and at least a list of the inherent critical infrastructures, as they may still ignore the external critical infrastructures. After all, identifying the critical infrastructures is a necessary step before protecting them. But the identification of critical infrastructures is not enough to guarantee their security, and their protection may be difficult if the private owners and operators, who own a great majority of the critical infrastructures, do not know exactly what they have to protect. Therefore, a more complete definition of a critical infrastructure is really necessary.

2.2.3 Dependency - How can an infrastructure be structured?

The highly connected nature of infrastructure, critical or not, is a major concern for anyone trying to ensure its security and to improve its resilience. The notion of dependency, more precisely the notion of interdependency, appears in a great number of documents on infrastrutures security [115] [105] [104]. These notions allow to link components of very different nature (human, technical, external, etc.).

We privileged the notion of dependency to the notion of interdependency as the fact that a component 1 is dependent of a component 2 does not necessary imply that the component 2 is dependent of the component 1.

The **dependency** is the one-directional reliance of a component on an other component. A component c_1 depends on a component c_2 if it is possible with c_2:

1. to have access to the component c_1 if c_1 is a place. For example, a room depends on its access (door, window, etc.);

2. to access, to obtain, to modify or to delete the component c_1 if c_1 is a physical object or a piece of information. For example, a safe depends on its location or its combination. And a combination depends on the person who knows it;

3. to corrupt, to exploit without its awareness or to injure the component c_1 if c_1 is a human. For example, an employee depends on its direct supervisor.

The components c_1 and c_2 are **interdependent** if c_1 depend on c_2 and if c_2 depend on c_1. "The degree of interde-pendency does not need to be equal in both directions". [105]

Classification by nature

There are different ways to categorize dependencies. They can be categorized according to their nature [133]:

1. physical,
2. cyber,
3. geographic
4. logical.

Several models based on this categorization already exist [132].

Classification by effect on the security

They can also be categorized according to how it acts on the infrastructure security:

1. a cascade failure is when infrastructure components exhibit a chain of dependencies and when the failure of one component in this chain propagates to the others. Since neither the extent nor complexity of chains of dependence is well known, cascade failure may represent a significant threat to infrastructure,
2. a single point of failure is when several infrastructure components depend on a single asset, or type of asset.

An example of a cascade failure happened in 2009 when the Cumbrian Floods destroyed a bridge carrying 312 fibre optic circuits serving 40,000 people, including police and local businesses and causing disruption to the transport sector. Another example occured in 2005 when 1.5 million to 2 million customers were deprived of electricity for several hours in Moscow and nearby regions due to a fire and explosion in a local south-eastern substation. The failure of this one substation led to a power outage in several areas on account of a cascade effect [66].

Examples of single point failure are regional convergence where multiple infrastructure components are located in the same area and constitutes a risk to resilience by magnifying the impact of localised disasters (A recent case study by Humberside has identified three major coal fired power stations and renewable energy assets, 17% of the UKs generating capacity, co-located in a region vulnerable to flooding [108]), and by increasing dependence on signals from Global

Positioning System (GPS) satellites as many infrastructure components rely on precise time signals to synchronize with other assets.

Discussion about how best to address interdependence is in its early stages, but some methods may include:

1. reducing coupling: gaining a better understanding of interdependencies, and if possible eliminating them, makes it easier to manage consequences of asset failure.

2. improving diversity: where dependence on supply from other assets is unavoidable, ensuring the availability of a range of sources can remove single points of failure.

2.2.4 Attack - What do the infrastructures face?

In broad outlines, an **attack** is an attempt to harm the targeted infrastructure. It could be by gaining unauthorized access to system services, resources, rooms, or information, by attempting to compromise a system integrity or an employee, or by damaging components.

Attacks can be direct or indirect [134]. **Direct attacks** would have for consequences the stoppage or disruption of the critical infrastructures functions or key assets through an attack on a critical component.
Indirect attacks would have for consequences damages that result from a reaction to attacks on other critical infrastructures. Many critical infrastructures depend on each other, therefore an attack on one critical infrastructure can affect other infrastructures.

Classification by nature

We distinguish three types of attack: physical, human, and the most predominant one, cyber.

A **physical attack** is an attempt to harm physically the targeted infrastructure. It could be by gaining unauthorized access to places or physical documents, by stealing equipment, raw material or document, or by damaging equipment. Physical attack's techniques include lock picking, theft techniques as pickpocket techniques or dumpster diving.

A **human attack** is an attempt to harm the infrastructure by targeted its personel and the people who have interactions with it. It is most commonly called social engineering, also defined as a psychological manipulation of people which leads them to perform actions or divulge confidential information. Christopher Hadnagy [63] and Kevin Mitnick [86] are two of the most notable social engineers. Social engineering's techniques include pretexting, diversion

theft, phishing (phone phishing, spear phishing), water holing, baiting, quid pro quo, or tailgating.

A **cyber attack** is an attempt to damage, disrupt, or gain unauthorized access to a computer, computer system, or electronic communications network. Cyber attacks include (non exhaustive list):

1. Identity theft, fraud, extortion,
2. Malware, pharming, phishing, spamming, spoofing, spyware, Trojans and viruses,
3. Denial-of-service and distributed denial-of-service attacks,
4. Breach of access,
5. Password sniffing,
6. System infiltration,
7. Website defacement,
8. Private and public Web browser exploits,
9. Instant messaging abuse,
10. Intellectual property (IP) theft or unauthorized access,
11. network packet snifflers,
12. distribution of sensitive information,
13. man-in-the-middle attack,
14. application layer attacks,
15. scareware,
16. malvertising,
17. social engineering techniques on social networks,
18. clickjacking,
19. Advanced Persistent Threats (APT).

This list still evolves today along with the information and communications technologies.

Classification by authorship

Attacks may be classified according to their authorship and impact [71]. An attack can be sponsored by:

1. nation states. Well-known examples are the cyber attacks of Estonia in 2007 and the Stuxnet cases,
2. private organizations,
3. terrorist, political or extremist group,
4. groups of organized crime,
5. hacktivists,
6. random people for personal reasons,
7. insiders or people with privileged access.

It is worth noticing that attacks are not the only threats that face infrastructures, incidents may occur too and be as devastating as an attack for the infrastructures. But, contrary to attacks, which are deliberate events, incidents are not deliberate, they belong to the safety domain. In the context of this thesis, we are taken only attacks into account.

2.2.5 Resilience - What if an attack happen and succeed?

Even if every possible precaution has been taken, no infrastructure is safe from attacks. The possibility is always here, and therefore, it can be necessary to evaluate if an infrastructure is able to recover from such events. This ability to recover is called the resilience.

A multidisciplinary notion

The notion of resilience is usually defined as the quality of being tough and able to recover from difficulty or damage [38] and appears in several topics: ecology, social sciences or engineering.

This notion appeared first in physics. It characterizes the resistance to impact or more precisely, the mechanical property of an inert material which consists of keeping its property after an impact.
Then several analogies follow in other topics. In psychology, the resilience is the process of adapting well in the face of adversity, trauma, tragedy, threats or significant sources of stress [7]. In computer science, it is the ability of an information system to withstand a breakdown or cyberattack and return to its initial operating state after the incident. [32]. In economics, it is defined as

the policy-induced ability of an economy to withstand or recover from the effects of exogenous shocks, arising out of economic openness [15]. In armament and aerospatial industry, it is the ability of an embedded system to keep going despite of being in degraded mode operation and in a hostile environment.

Infrastructure resilience

This leads us to the infrastructure resilience. As for the notion of critical infrastructure, many nation states and organizations have defined this notion: Canada [17], France [32], Germany [25], New Zealand [60], Quatar [130], The United Kingdom [123], or The United States of America [104] [105].
The following definition is a summary of the definitions found for the previous nation states and organizations.

The **resilience** of an infrastructure is its ability to resist, absorb, adapt to, recover from, or successfully adapt from disruptions resulting from deliberate attacks, accidents, or naturally occurring threats or incidents.

It would be interesting to keep this notion in mind and see if mathematical structures can help to evaluate the resilience of an infrastructure.

The work presented in this chapter has been published a first time in the proceeding of the 9^{th} International Conference of Cyber Warfare and Security [45] and a second time in the Journal of Information Warfare [48].

Now that we have defined the main notions which are indivisible of the topic of infrastructure security, it is time to present how we are going to model them in order to evaluate an infrastructure security. The next chapter discusses some of the existing infrastructure models.

Chapter 3

Infrastructure models

The threats that face infrastructures are increasingly numerous and can take many different forms. An infrastructure must face competitors, organized gangs, states or even terrorist groups that have a wide variety of tools and strategies. An attack can involve various techniques ranging from theft of mobile devices to the espionage of individuals or the undermining of the image and the credibility of the targeted infrastructure. Furthermore, the final target of an attack is not always obvious. Attackers may favor attacks based on a domino effect. Indeed, it is often preferable to target less protected secondary objectives that will have, by rebound, disastrous consequences on the main objective.
In this context, there is an important need to represent data that affects the security of an infrastructure. Indeed, in order to ensure the security of an infrastructure, it is necessary to know the components that have to be protected and all the peripheral elements making it possible to access them. Infrastructures have more and more sensitive elements, often connected to the information system, accessible from anywhere. It becomes impossible to analyze the mass of information without a tool of representation.
In order to create this tool, a model of infrastructure has to be defined. Many models of infrastructure exist to respond to a great variety of problematics. In this chapter, we will describe a few of them.

Graphs are commonly used to model infrastructures and to respond to an impressive list of various problematics: information extraction [73] [65], decoding of low-density parity-check codes [62], modeling of gene regulatory networks, gene finding and diagnosis of diseases [78], testing an application, supporting the systematic management of hospital information systems [152], detecting salient objects [64] [79] [18], understanding the mechanisms by which failures, ideas, and diseases propagate within networks [149] [92] [82] [8], or modeling the topological structure of internetworks and studying problems ranging from routing to resource reservation[153] to cite a few.

There are many application fields for graph-based models and as well, many

graph-based models exist. It is also true for the sole application field that is infrastructure security. In order to give a brief overview of existing graph-based models for infrastructure security, we divide them into two categories: the models which depict ways in which an attacker can harm an infrastructure and models which depict infrastructures.

3.1 Models that depict the attacks

The first type of infrastructure models that are presented are models that depict ways in which an attacker can exploit vulnerabilities to break into the infrastructure. Two specific models are particularly described here: attack tree-based models and attack graph-based models.
Due to our approach that favors the attacker's point of view, fault trees [21] and threat trees (which are very similar to attack trees) [3] are not described here.

3.1.1 Attack tree-based models

Attack trees are widely used in infrastructure security [143] as they are useful to systematically categorize the different ways in which an infrastructure can be attacked. The term "attack tree" is first introduced by Bruce Schneier in 1999 [135] [136].

An **attack tree** is a tree (an undirected graph with no cycle) whose nodes represent different attacks, sometimes called atomic attack [137], and whose arcs link a goal to its subgoal(s).
The **root node** of the tree is the global goal of an attacker. The children of the root node are subgoals of this global goal, and children of these nodes are subgoals of these subgoals, and so on. The nodes, except for the root node, are either conjunctive or disjunctive. If the nodes are **disjunctive**, this means that satisfying one sub-goal suffices. If the nodes are **conjunctive**, this means that all sub-goals have to be fulfilled.

Figure 3.1 (page 36) shows an example of an attack tree, previously presented by Bruce Schneier [135]. In this tree, the global goal of the attacker is to open a physical safe. The nodes corresponding to the subgoals "Find the combination written down", "Get the combination from the safe owner" and "Try several combinations until find the good one" are children of the node corresponding to the "Learn the combination" goal. They are disjunctive nodes (also called OR nodes). In the contrary, the goal "Eavesdrop" has two conjunctive subgoals (whose corresponding nodes are called AND nodes): "Listen to conversation" and "Get target to state the combination". Indeed, in order to eavesdrop on someone saying the safe combination, attackers have to eavesdrop

on the conversation and get safe owners to say the combination.

It is also possible to **assign values to the nodes**, whether they are Boolean or continuous, in order to evaluate the security of the main goal. Usually, the values are first assigning to the leaf nodes (and the leaf nodes only) and then calculations are made from these values to assign the resulting values to the remaining nodes.

For instance, it is possible to assign the Boolean values I and P, which means that the attack is impossible and possible respectively, to the previous example of attack tree (see Figure 3.1, page 36). First, these Boolean values are assigned to the leaf nodes only and then, the values of the remaining nodes are calculated. To do so, we say that the value of an OR node is possible if any of its children are possible, and impossible if all of its children are impossible. The value of an AND node is possible only if all children are possible, and impossible otherwise. The possible attacks are shown by dashed lines in Figure 3.1 (page 36). In this case, there are two possible attacks: cutting open the safe, or learning the combination by bribing the owner of the safe.
The following Boolean values can also be studied: easy and difficult, expensive and inexpensive, intrusive and nonintrusive, legal and illegal, special equipment required and no special equipment.

As said previously, it is also possible to assign continuous values to the nodes. Figure 3.2 (page 37) shows the attack tree with different costs assigned to the leaf nodes, instead of just having an “expensive” or “inexpensive” Boolean value for instance. Like Boolean values, continuous values can propagate from the leaf nodes to the root node as well. The disjunctive nodes have the value of their cheapest child and the conjunctive nodes have the value of the sum of their children. In Figure 3.2 (page 37), the dashed lines represents the cheapest attack. The probability of success of a given attack or the likelihood that an attacker will try a given attack are another examples of continuous values that can be assigned to the nodes of an attack tree.

These Boolean and continuous values can be combined to learn even more about an infrastructure's vulnerabilities. For example, Figure 3.3 (page 38) shows the cheapest attack requiring no special equipment.

Attack trees provide many advantages. First, they may highlight that the areas people usually think of as vulnerable are not. Secondly, they may also highlight that the areas people think of as vulnerable usually are not. Second, attack trees capture knowledge in a reusable form. Once an attack tree has been completed, it is possible to use it in other situations. And finally, the graphical, structured tree notation is helpful to partially automate the threat analysis process.

To conclude, attack trees provide a formal methodology for analyzing the

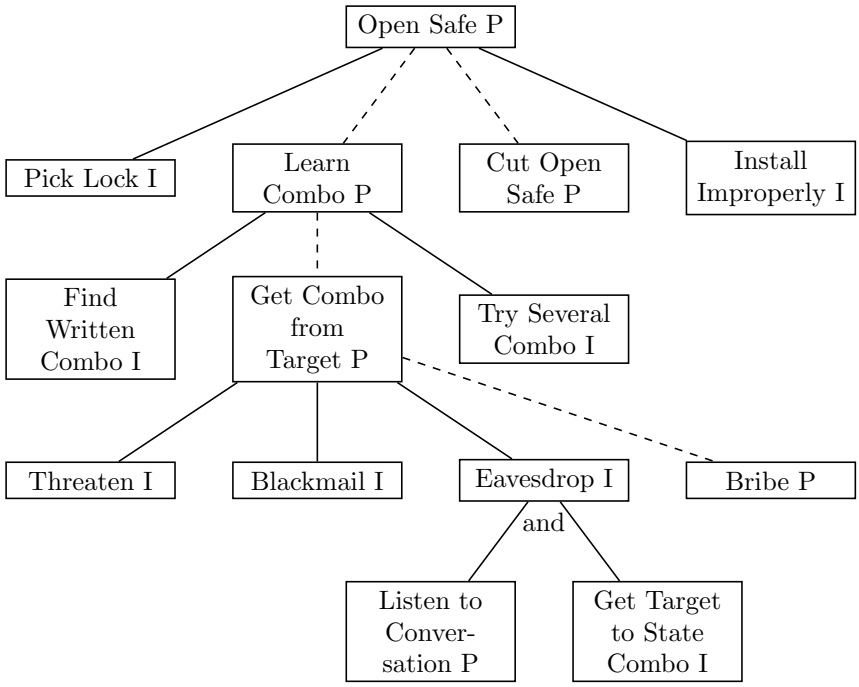

Figure 3.1: An example of attack tree with Boolean values (Bruce Schneier)

security of infrastructures and parts thereof. The basic concepts of attack trees have been formalized by Mauw and Oostdijk [84] and have been used in a great variety of applications: analysis of conventional information systems, analysis of threats against tamper resistant electronics systems (e.g. avionics on military aircraft), computer control systems (especially relating to the electric power grid [143]).

3.1.2 Attack graph-based models

Attack graphs can be seen as a generalization of attack trees. Like them, attack graphs are useful to systematically categorize the different ways in which an infrastructure can be attacked. Despite that, the attack graphs are not used as much as the attack trees.

An **attack graph** is a graph whose nodes represent different attacks, sometimes called atomic attack [137], and whose arcs link a goal to its subgoal(s). The **source nodes** (nodes without predecessor) are global goals of an attacker. The children of the source nodes are subgoals of these global goals, and children of these nodes are subgoals of these subgoals, and so on. The nodes, except for

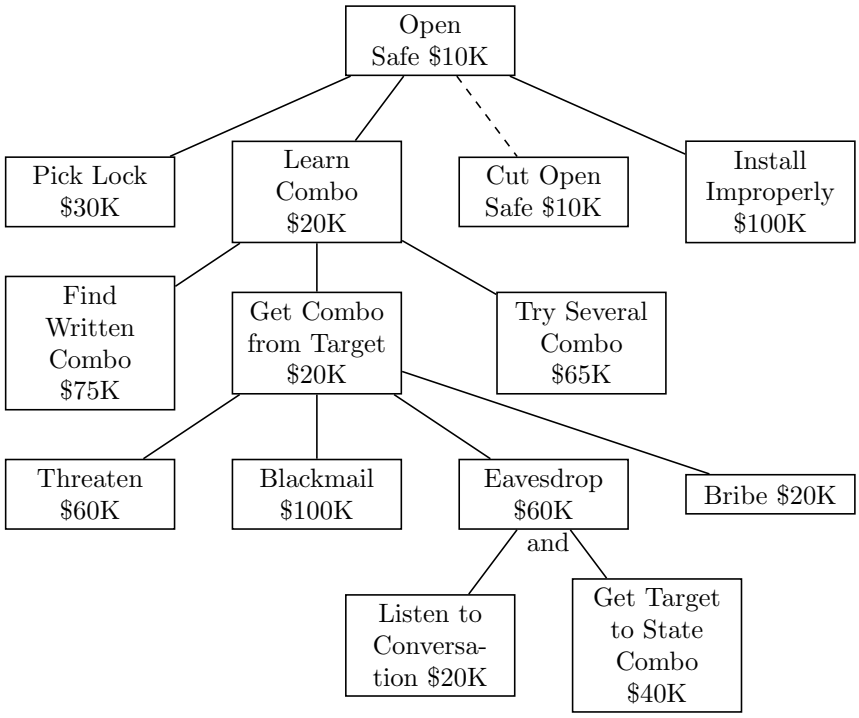

Figure 3.2: An example of attack tree with continuous values (Bruce Schneier)

the source nodes, are either conjunctive or disjunctive. Again, if the nodes are **disjunctive**, this means that satisfying one sub-goal suffices. If the nodes are **conjunctive**, this means that all sub-goals have to be fulfilled.

Figure 3.4 (page 39) shows an example of attack graph based on the example of attack tree previously presented in section 3.1.1. Another main goal and its associated subgoals were added. In this graph, the global goals of the attacker are to open a safe and to get a confidential document.
In this case, we assume that the document is not in the safe, otherwise the graph is a tree.

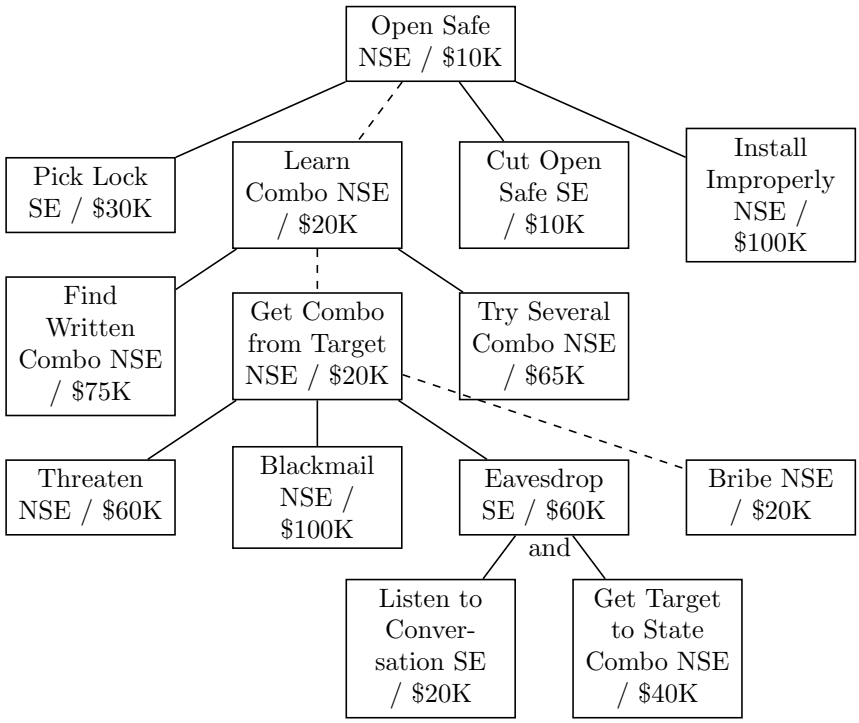

Figure 3.3: An example of attack tree with Boolean and continuous values (Bruce Schneier [135])

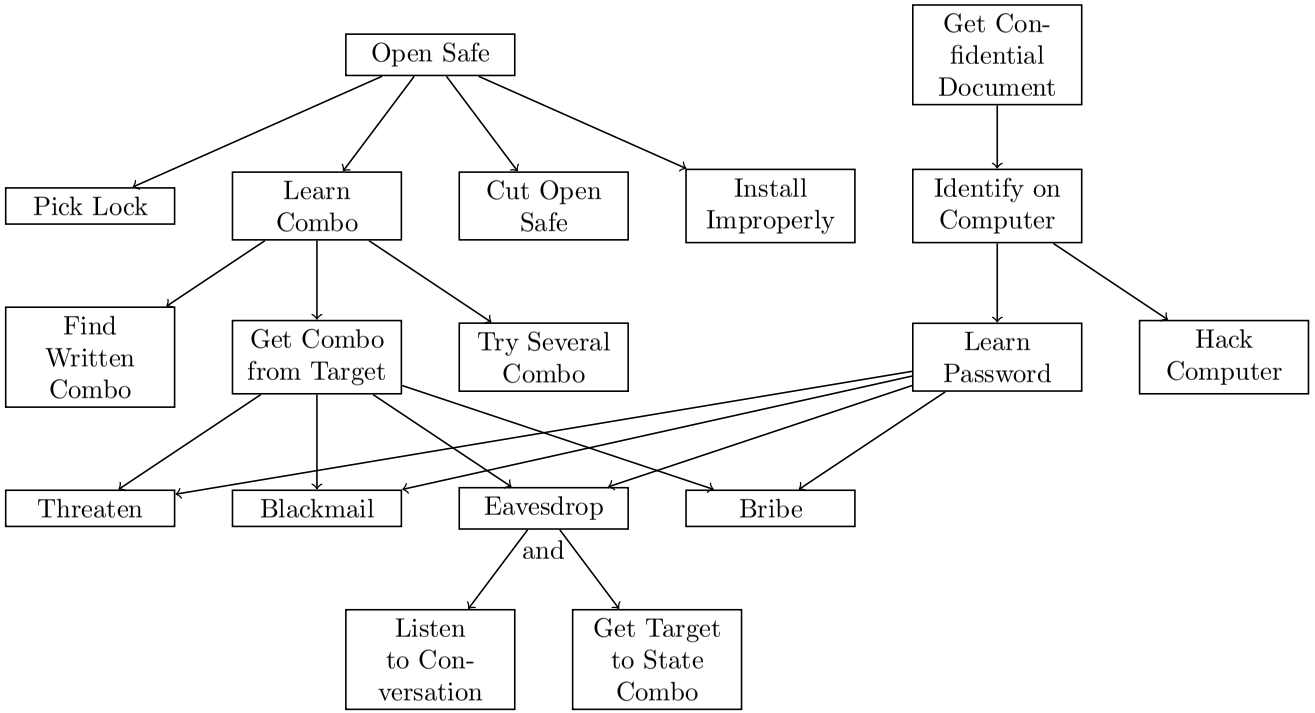

Figure 3.4: An example of attack graph

It is also possible to **assign Boolean and/or continuous values** to the nodes in order to evaluate the security of the main goals. As before, the values are first assigning to the leaf nodes and then propagating to the remaining nodes of the attack graph.

Attack graphs provide many of the advantages of the attack trees. First, they may highlight areas of the attack trees that people think of as not vulnerable when they are in fact vulnerable, and also highlight that the areas people think of usually as vulnerable are not. And finally, the graphical, structured tree notation is helpful to partially automate the threat analysis process.
Furthermore, they help to understand whether given critical resources can be compromised through multi-step attacks and allows to have a perspective on the whole infrastructure.

To conclude, attack graphs also provide a formal methodology for analyzing the security of infrastructures and parts thereof. Attack graphs are formally defined by Sheyner et al. as a tuple $G = (S, \tau, S_0, S_8)$, where S is a set of states, $\tau \subseteq S \times S$ is a transition relation, $S_0 \subseteq S$ is a set of initial states, and $S_8 \subseteq S$ is a set of success states [137].

As said previously, the attack graph-based models are not used as much as the attack tree-based models. We suppose that the reason behind it is the following: any attack graph can be transformed into an attack tree. Indeed, you just have to link all the source nodes of an attack graph to a root node representing the goal "Harm the infrastructure" to do so. Furthermore attack trees tend to be easier to manipulate and they can be easily reused, which is more difficult for an attack graph. We think that explains why attack trees are privileged in comparison to the attack graphs.

However, as qualitative models, attack graph-based models and attack tree-based models still adopt a binary view towards security, that is, an infrastructure is either secure (critical components are not reachable) or insecure. This is a limitation because it is usually desirable to find a relatively superior option among secure configurations [148].

3.2 Models that depict the targeted infrastructure itself

For now, we have presented only models which describe the different ways an attacker can harm an infrastructure. In this section, we are interested in representing the infrastructure itself. Again, we focused the study on the graph and tree-based models.

Contrary to the models which depict the different ways an infrastructure can be harmed, there is no specific name to designate the structure whose models are based on. It is probably because, unlike attack tree and attack graph-based models, these models are not limited to the field of infrastructure security.

3.2.1 Graph-based models

Graphs have been used for a very long time to model infrastructures, since we can go back to one of the first problems of graph theory: the Seven Bridges Problem of Königsberg [151]. In this case, seven bridges of the city of Königsberg are modeled in order to find a path that passes only once on each bridge.

In a graph-based model which depicts an infrastructure, an infrastructure is represented by a **graph** whose nodes model components of the infrastructure and whose arcs model relationships between those components.
The **source nodes** (nodes which have no successor) tend to represent key components of the modeled infrastructure. The components can be people, server, service, door, safe, data, power pole, etc. For example, in the model used by the Maltego tool [126], nodes model the following components, or entities like they call them:

1. People (names, email addresses, aliases),
2. Groups of people (social networks),
3. Companies,
4. Organizations,
5. Web sites,
6. Internet infrastructure (such as domains, DNS names, netblocks, IP addresses),
7. Affiliations,
8. Documents and files.

The relationships between those components can be physical or abstract. However, in practice, the modeled relationships tend to be more abstract (hierarchical link, belonging link, etc.) than physical (electrical wires, etc.). In the Maltego tool, arcs link a company to its employees, employees to their social media accounts (twitter, instagram and others), or a domain name to its IP address to cite a few.

Figure 3.6 (page 43) shows an example based on the attack graph case in order to show the differences between the two approaches. Two key components are representing by two source nodes. The first one is a safe and the

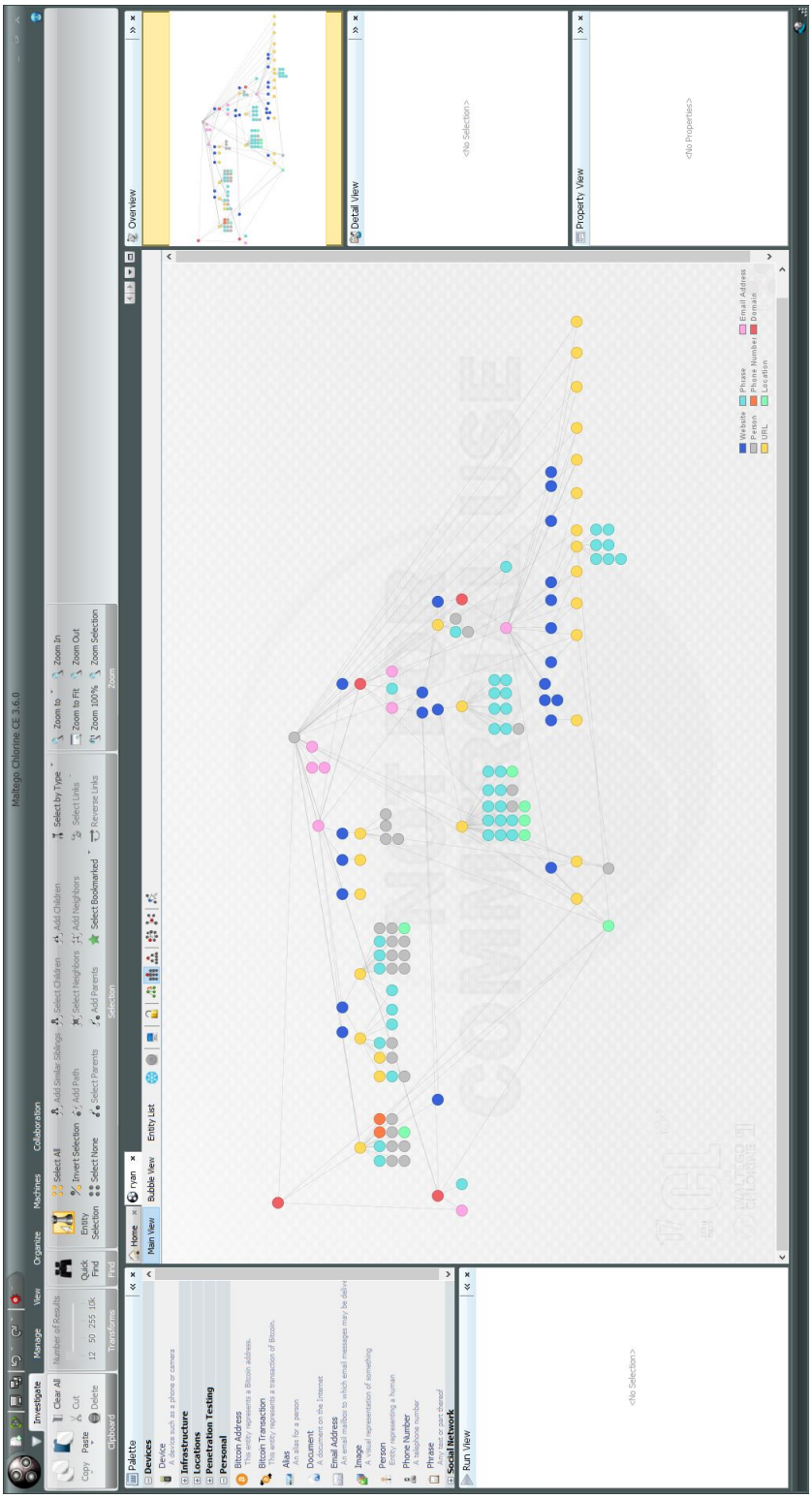

Figure 3.5: Screenshot of Maltego

second one is a confidential document. Then, every components linked to these two components are represented. And every components linked to the previous components are represented. And so on. These components are defined as peripheral components.

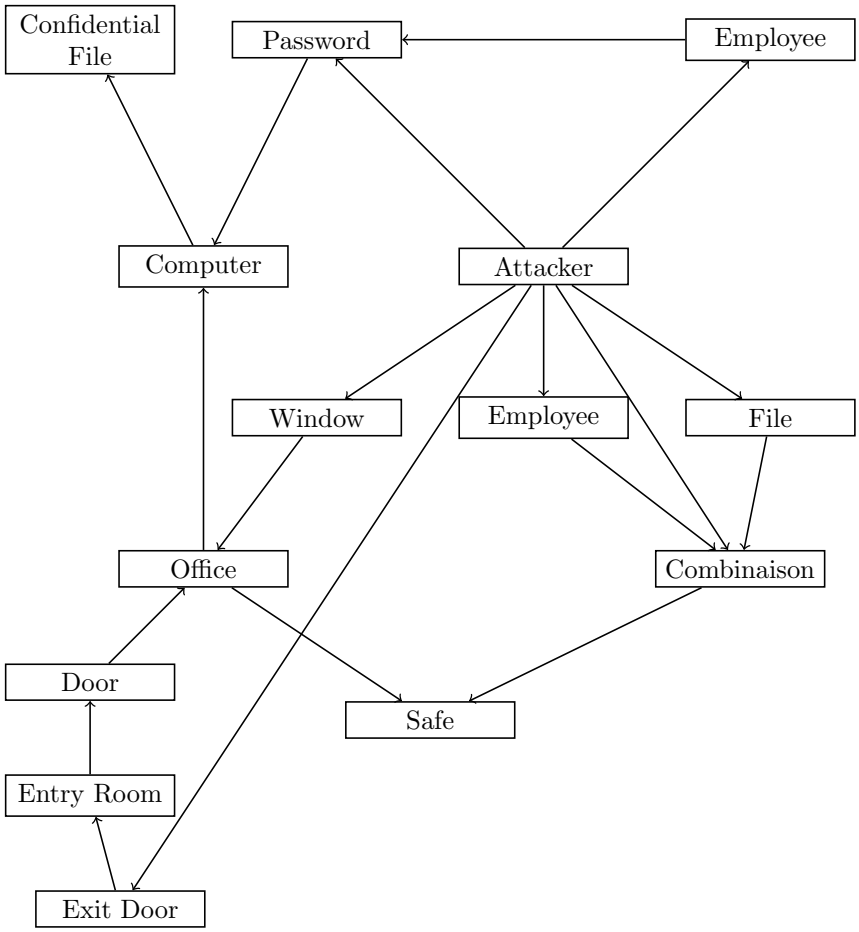

Figure 3.6: An example of a graph-based model of a very simple infrastructure

It is possible to **assign values to the nodes and to the arcs**. The values associated to the nodes can be Boolean or continuous values, and are useful to give a precise description of the component represented by the node, whereas the values that can be associated to the arcs are continuous. Usually, the values are assigning to all the nodes and arcs. Contrary to the attack graph-based models, there is no propagation of these values. But it is possible to assign values to the arcs according to the values of the nodes.

In order to evaluate the security of an infrastructure, Boolean values as protected/unprotected or possible/impossible, and continuous values as cost, time, or effort are privileged.

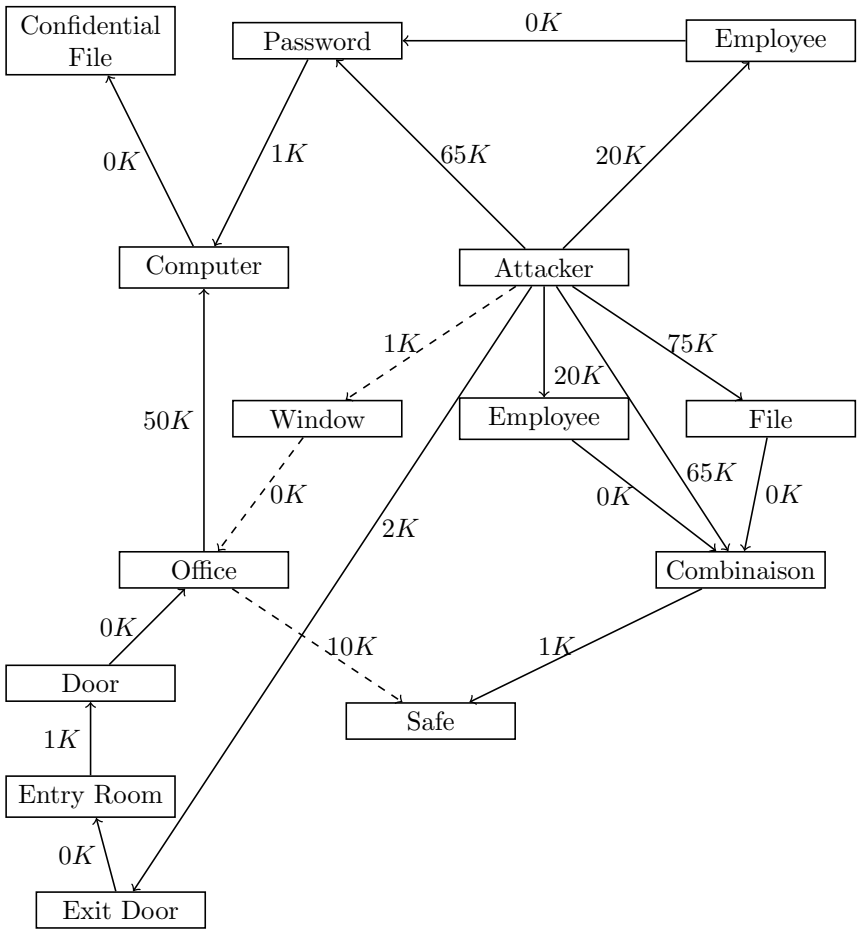

Figure 3.7: An example of a simple infrastructure modeled by a graph

In order to evaluate the security of an infrastructure, the continuous values can also be added up, maximized, or minimized. For example, in Figure 3.7 (page 44) a cost is assigned to every arc. The question we want to answer is the following: what is the cost of the cheapest attack? If this cost is considered too low, the infrastructure is considered as not secured since the security of its key components is compromised. The cost of an attack against a key component of the infrastructure is calculated by adding up all the values assigning to the arcs belonging to a path from a node representing the attacker to a source

node representing the key component. In Figure 3.7 (page 44), the cost of the attack represented by the path " Attacker - Employee - Password - Computer - Confidential Document " is $20K + 1K = 21K$. The cost of the cheapest attack representing by the path "Attacker - Window - Room - Office - Safe " (in dashed lines) is $1K + 10K = 11K$. If the value of the content of the safe is more than $11K$, then this infrastructure is not secured.
There are almost as many ways to evaluate the security of an infrastructure as models of infrastructure: the way we have presented here is just one among many others.
As instance, an effort associated to the abilities of attackers can also be assigned to the arcs of the graph-based model. In this case, the question we want to answer is: what is the easiest attack? And the effort of an attack is calculated by maximizing all the values assigned to the arcs belonging to a path from a node representing the attacker to a source node representing the key component.

To conclude, graph-based models allow to have a good understanding of an infrastructure and how it works. There are as many models of infrastructure as ways of evaluating the security of an infrastructure: study of cascading failures in power grids [75] [31], development of strategies for efficient operation and control of a water distribution network [36] or modeling the topological structure of internetworks [153]. In this regard, graph-based models are more versatile than the attack graph-based models.
Graph-based models may highlight components of an infrastructure that people think of as not important when they are in fact vital for the infrastructure security. And these components tend to be vulnerable as they are not considered as important.

3.2.2 Tree-based models

Trees are undirected graphs with no cycle. These restrictions do not allow to represent the entire targeted infrastructure in most cases since only the components linked to the component modeled by the root node of the tree are taken into account. Then, not surprisingly, tree-based models are not used as much as graph-based models. We present them anyway, as this structure is going to be mentioned in the following chapter.

In a tree-based model which depicts an infrastructure, part of an infrastructure is represented by a **tree** whose root node models the key component targeted by an attacker, whose nodes model components of the infrastructure related to the targeted key component (and that can be used to compromise the security of the key component), and whose arcs model relationships between those components.
As previously said, the components can be (non exhaustive list) people, server, service, door, safe, or data, and the relationships between those components can be physical or more abstract.

Figure 3.8 (page 46) shows an example based on the attack tree model in order to show the differences between the two approaches. The targeted key component is a safe. Then every components linked to this component are represented. And every components linked to the previous components are representing. And so on.

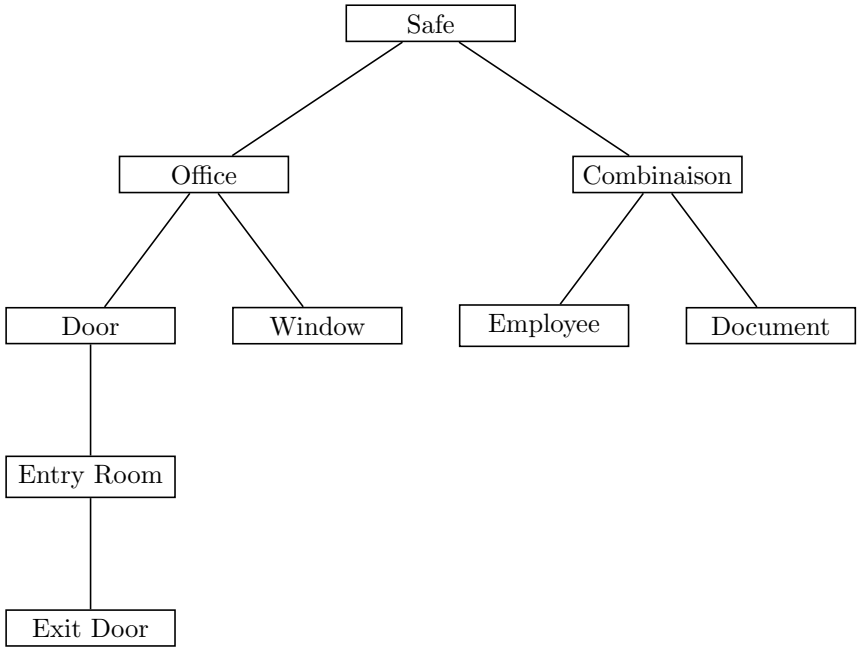

Figure 3.8: An example of tree-based model of a very simple infrastructure

It is also possible to **assign Boolean and/or continuous values** to the nodes and to **assign continuous values** to the arcs. Again, the values can be added up, maximized, or minimized in order to evaluate the security of an infrastructure.

Despite of the constraint of a sole source node (the root node), tree-based models manage to keep some advantages of the graph-based models. First they are also versatile in their ways of evaluate the security of key components of an infrastructure. Secondly, they highlight components of an infrastructure that people think of as not important when they are in fact vital for the infrastructure security. And these components tend to be vulnerable as they are not considered as important.
However, they indeed allow to model only a part of an infrastructure and that is why graph-based models are privileged.

We conclude this chapter by explaining our choice of infrastructure model: a graph-based model which depicts the infrastructure itself. First, the aim of this thesis has always been to depict an infrastructure the more precisely possible in order to evaluate its security. This is why definitions of a critical infrastructure were studied and a new definition has been provided. Secondly, these models are more versatile. To evaluate the security of infrastructures, models which depict attacks are constrained to research only the shortest paths or to simply research paths. Instead, the models which depict infrastructures have more possibilities as structures other than paths could have been studied. Finally, tree-based models were discarded, mostly as they can be seen as a particular form of graph whose structure has been considered too poor to represent an infrastructure precisely because of the existence of a unique root, whereas an infrastructure may have may have several critical assets to protect. Therefore, the following InfraSec model is a graph-based model which depict the targeted infrastructure.

Chapter 4

The proposed infrastructure model

In this chapter, the chosen model for the InfraSec tool is presented. First, we give a brief overview of the graph-based model, then we explain in more details how this model represents an infrastructure and its environment. This work was made with the collaboration of the auditor of the company TEVALIS.

4.1 A brief overview of the model

The InfraSec model is a model based on the graph theory. In a nutshell, an infrastructure is modeled by a directed graph whose vertices represent the components of the targeted infrastructure and whose arcs represent the links of dependency between two components. For example, if the security of a component c_2 depends on the security of a component c_1, then an arc exists between c_1 and c_2. The arc is denoted (c_1, c_2), as seen in Figure 4.1 (page 49).

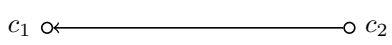

Figure 4.1: c_1 depends on c_2

The arc is directed towards c_2 as it is more natural to be directed towards a target than towards a starting point. A component is a **starting point** if the attacker has a direct access to it, and a **target** is a component that the attacker wants to reach.

In Figure 4.2 (page 50), the model of a simple infrastructure with 10 components is shown.

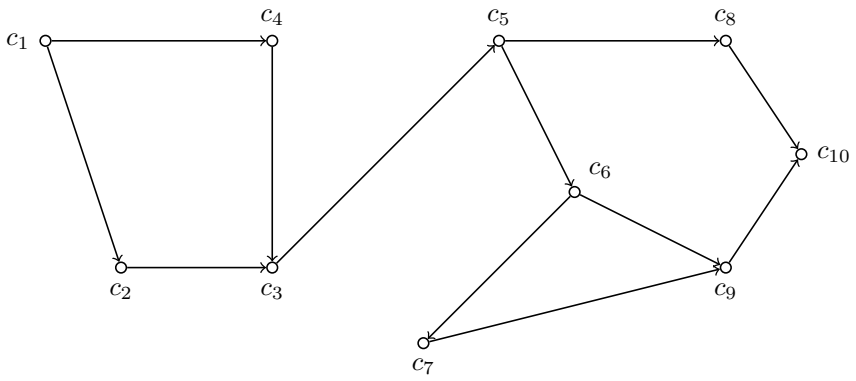

Figure 4.2: Model of a very simple infrastructure

We opted for a graph-based model because of the enlarged vision of security it brings. The model has to be adaptable to each infrastructure and its special features, so everything has to be taken into account in order to have the best possible representation of an infrastructure, and in order to have the most pertinent results possible.
Furthermore, there is still a lack of model which can adapt to all types of infrastructure. Indeed, the ones proposed are still very specific and respond often to very precise problematics.

Figure 4.3 shows an example of infrastructure modeled on the InfraSec tool.

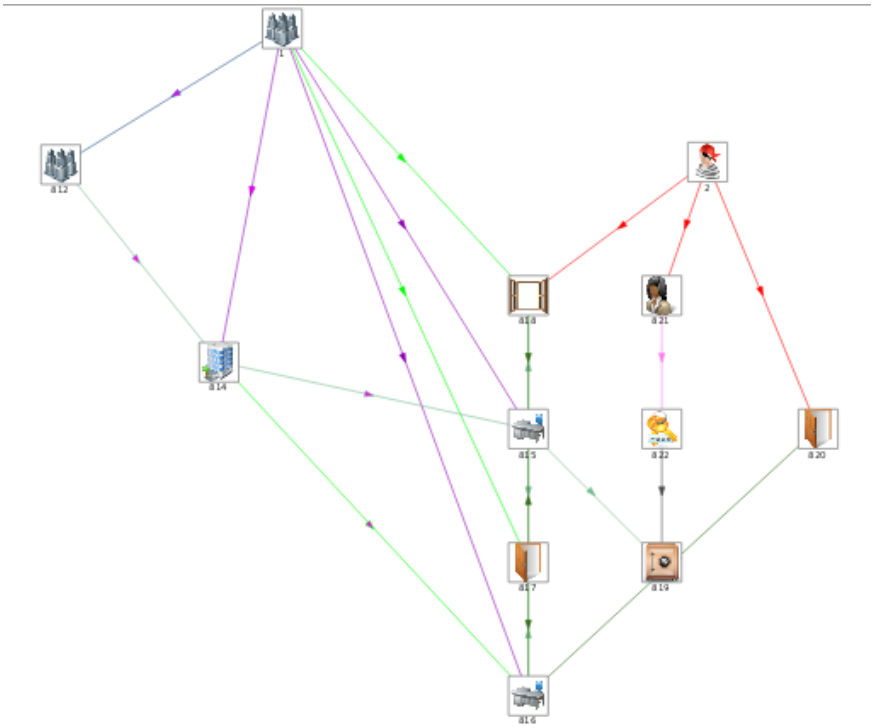

Figure 4.3: A simple infrastructure modeled on the InfraSec tool

4.2 How does a node model a component of an infrastructure?

In the InfraSec model, a node represents a component of the modeled infrastructure, whether it is an employee, a building, a network, a raw material or a production machinery [45].

4.2.1 The characteristics of a node

In order to describe the various components of an infrastructure, different characteristics are associated with a node. These characteristics are:

1. a wording,
2. a category which allows to identify the nodes with the same characteristics,
3. a group of attributes shared by all nodes:
   (a) starting point (yes/no) which allows to define if a node can be the starting point of an attack (if an attacker can have a direct access to it),
   (b) target (priority/secondary/no) which allows to identify if a node can be the target of an attack,
   (c) hidden (yes/no) which allows to withdraw a node of the graph without deleting it for good. This allows to keep a copy of the node even if it does not appear in the graph.
4. a group of attributes specific to each category of nodes,
5. a group of grades which allow to evaluate the vulnerability of the node in view of the attacker's abilities:
   (a) human vulnerability (0 to 10): the node's vulnerability against attacks targeting people,
   (b) physical vulnerability (0 to 10): the node's vulnerability to physical attacks like lock picking,
   (c) IT vulnerability (0 to 10): the node's vulnerability to IT attacks.
6. a group of grades which allow to evaluate the level of an attacker's abilities necessary to attack this node:
   (a) human abilities (0 to 10): abilities to succeed in performing social engineering attacks,
   (b) physical abilities (0 to 10): abilities to succeed in performing physical attacks like lock picking or sabotage,
   (c) IT abilities (0 to 10): abilities to succeed in performing IT attacks.

7. an open text which allows the user to describe the node freely.

All these characteristics are used to determinate the nodes' grades of vulnerability and ability.

4.2.2 A non-exhaustive list of categories

As said previously, the nodes are divided into several categories. Two nodes are in the same category if they share the same characteristics. For now, the following categories have been defined:

1. **the attacker**,
2. **the infrastructure**, (often a company in the context of the InfraSec project) which is the highest level of representation in the InfraSec tool. Every component belongs to an infrastructure, whether this structure is the target or not;
3. **the logical components**, which represent every element allowing to describe the infrastructure logically. A logical component must belong to an infrastructure. There are several subcategories of logical components:
   (a) the structure, which represents the various entities, whether they are internal or external (company, service, service provider, client, supplier, etc.),
   (b) the implantation, which represents the location and the geographic distribution of the components (geographic location, building, room, etc.),
   (c) the network, which represents the various networks of the infrastructure (water network, electrical network, ethernet, virtual private network known as VPN, bluetooth, etc.),
4. **the operational components**, which represent every element which can be targeted by an attack. An operational component must belong to a logical component. There are several subcategories of operational components:
   (a) the data, whether they are digital or not (the data must not be a product),
   (b) the person, which represents the staff of the infrastructure and also any person which could have an impact on the infrastructure,
   (c) the mean of access, which represents the entries of a physical entity (door, window, gate, road, etc.),
   (d) the information technology, which represents the IT equipment, networks and telephony,

(e) the equipment, which represents the furniture (cabinet, safe, etc.), as well as the furniture specific to the infrastructure (manufacturing machine, etc.) or to its security (key, camera, alarm, etc.),

(f) the product, which represents what is sold or purchased by the infrastructure, whether it is a data, a knowledge, a raw material or a manufactured product.

It is recognized that a category can belong to another category, the first category is then called subcategory. Subcategories also allow to have a better and more realistic description of the nodes. For example, the subcategory 'Secretary' gives more information than the category 'Person'. Then, the InfraSec tool allows the user to fill automatically the node's properties, with values determinate through self-learning and by taking into account all the nodes of the same subcategory.

4.2.3 Two layers of categories

Only some categories of nodes are taken into account by algorithms to find attack patterns. The remaining categories are used to understand and exploit the various layers in the organization of the infrastructure (services, subsidiaries, etc.). This allows to have a better understanding of the infrastructure and how it works. Nevertheless, if the attacker wants to target a specific service, he will in fact target a person, data, or equipment, but never the service itself. That is why the InfraSec tool uses the nodes and the arcs according to two layers:

1. the layer '**Analysis**' gives an incomplete view of the graph, a subgraph, which is composed of the nodes and arcs allowing the understanding of the infrastructure and how it works,

2. the layer '**Attack**' gives an incomplete view of the graph, a subgraph, which is composed of the nodes and arcs used by the algorithms to determinate attack patterns.

The layer 'Analysis' includes the infrastructure and the informational components categories which are useful to structure and organize the model, and therefore to understand it better. These categories are not taken into account to find attack patterns against the modeled infrastructure. The views of the layer 'Analysis' are:

1. display of a subgraph whose nodes and arcs are relevant to the organization of the infrastructure,

2. display of a subgraph whose nodes and arcs are relevant to the geographic distribution of the infrastructure,

3. display of a subgraph whose nodes and arcs are relevant to the information system of the infrastructure,

4. display of a subgraph defined by the user itself.

The layer 'Attack' includes the attacker and the operational components categories which are taken into account to find attack patterns against the modeled infrastructure. The views of the layer 'Attack' are:

1. display of a subgraph whose nodes and arcs can be a part of an attack against a specific node,
2. display of a subgraph whose nodes and arcs can be a part of an attack against a specific arc,
3. display of a subgraph defined by the user itself.

4.3 Evaluation of the vulnerabilities and the abilities

Identifying the most efficient attack patterns is one of the main goals of this thesis. In order to do so, the model takes into account various values associated to each node and each arc. It is necessary that these values are clearly defined in order to preserve the general coherence of the model.

4.3.1 Grades of vulnerability

The grades of vulnerability of a node are automatically calculated according to the value of the various properties of the node. To do so, the following principles are applied:

1. each vulnerability of a node has the grade 5 out of 10 as default value, the grade is then modified according to the properties of the node,
2. each property of a node is associated to three numerical coefficients which represent the impact of the property on the human vulnerabilities, the impact of the property on the IT vulnerabilities, and the impact of the property on the physic vulnerabilities, respectively,
3. each property of a node which can influence at least one grade of vulnerability of the node is associated to a list of values,
4. each element of the list of values which is associated to a property of a node, is associated with a numeric value. If this value is negative, then the grade of vulnerability is reduced and the security of the node is stronger.

Vulnerability of a node		Ability of the attacker or the targeted node	
Grade	Description	Grade	Description
0	The success of an attack against the node is impossible.		
1	The vulnerabilities can be exploited only by an exceptional attacker with unlimited resources.	10	The attacker can perform any attack and has unlimited resources.
2	The vulnerabilities can be exploited only by an exceptional attacker.	9	The attacker can perform any attack but does not have unlimited resources.
3	The exploitation of the vulnerabilities is complex and needs some important resources.	8	The attacker can perform complex attacks with some important resources.
4	The exploitation of the vulnerabilities is complex but does not need important resources.	7	The attacker can perform complex attacks but has limited resources.
5	The protection of the node was realized by security professionals.	6	The attacker is a professional, or has the same abilities, and has important resources, but lacks of experience to perform complex attacks.
6	The protection of the node matches the minimal recognized professional recommendations and is adapted to the node's context.	5	The attacker is a professional, or has the same abilities, but does not have important resources and lacks of experience to perform complex attacks.
7	The protection of the node matches the minimal recognized professional recommendations but is not really adapted to the node's context.	4	The attacker has a little knowledge and can exploit basic vulnerabilities, but he is not discreet.
8	The protection of the node is not enough	3	The attacker is resourceful and can exploit simple vulnerabilities.
9	The protection of the node can be easily bypassed.	2	The attacker has no knowledge but with time, he can exploit very simple vulnerabilities.
10	The node is not protected.	1	The attacker has no knowledge but can exploit obvious vulnerabilities.
		0	An attack is not possible.

Figure 4.4: The value scale

Be p_i the value of the property i and $coef_i$ the coefficient of property i. The computation formula is:

$$\text{Vulnerability} = 5 + \sum_{i=0}^{n} p_i \times coef_i$$

For practical reasons, the auditor is allowed to give its own grades of vulnerability. But, in order to ensure the coherence of the algorithms' results, the auditor must refer to a value scale described in Figure 4.4.

4.3.2 Grades of ability

The attack's abilities are the abilities of a person or a group of persons to perform a human, IT or physical attack. Ideally, the abilities should be calculated automatically. For now, the auditor has to define the values of these abilities according to the value scale in Figure 4.4.

4.4 How does a node model an attacker?

We do not expect to be able to model human behavior at this point, but we do believe that the model has to incorporate attacker abilities, as this can have a significant impact on security decisions [128].
The category 'Attacker' does not have subcategories and is represented by a sole node: the attacker.

4.4.1 How is defined an attacker?

An attacker is an individual or a group of individuals with various abilities which allow them to fulfill their goal(s). We divide these abilities into three categories: physical abilities (locksmith's trade, physical strength, theft techniques, etc.), social engineering abilities (pretexting, diversion theft, phishing, etc.) and technical abilities (malware, phishing, denial-of-service, etc.).
These abilities allow an attacker to reach one or several targets of the targeted infrastructure with a minimal cost and time of execution, maximal effectiveness, and without getting caught.

4.4.2 The characteristics of the attacker node

In the InfraSec model, an attacker is modeled by a node and is considered as a component of the infrastructure. The node is linked to every starting point (node whose property 'Entry' is yes). It allows to quantify the initial effort an attacker has to make to attack the infrastructure.

The node 'Attacker' does not have additional properties like the others. The tool InfraSec takes into account only the three grades of ability. The characteristics of the attacker are:

1. a grade of physical abilities,

2. a grade of social engineering abilities,

3. a grade of technical abilities.

By default, the values of the grades are set to their maximum, which means that the attackers can perform every possible attacks. Nevertheless, it is possible to restrict the abilities of the attacker in order to look for specific attacks which are associated to a specific attacker's profile.

4.5 How does an arc represent a relation of dependence between two components?

The components of an infrastructure depend more or less on the others components. These dependencies, whether they are material, social, logistical, environmental or software, represent links between these components.

As a reminder, a component c_1 depends on a component c_2 if it is possible with c_2:

1. to have a physical access to the component c_1 if c_1 is a place;

2. to obtain, to modify or to delete the component c_1 if c_1 is a physical object or a piece of information;

3. to corrupt, to exploit without its knowing or to injure the component c_1 if c_1 is a human.

The components c_1 and c_2 are interdependent if c_1 depend on c_2 and if c_2 depend on c_1.

An arc represents a link of dependency between two components. If a component c_2 depends on a component c_1, then there is an arc a from the node representing c_1 to the node representing c_2. The arc is denoted (c_1, c_2) (see Figure 4.1 page 49).
If the components c_1 and c_2 are interdependent, then there is an arc a_1 from the node representing c_1 to the node representing c_2 and an arc a_2 from the node representing c_2 to the node representing c_1 (see Figure 4.5 page 59).

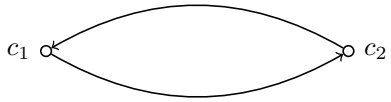

Figure 4.5: Two interdependent components

4.5.1 The characteristics of an arc

In order to describe the various links of dependency, different characteristics are associated with an arc. These characteristics are:

1. its originating node,
2. its destination node,
3. its type.

There are three types of arcs whose properties are different according to the type:

1. the hierarchical arcs which correspond to a hierarchical relationship between two nodes. The destination node belongs to the originating node. No value shall be associated to these arcs,
2. the arcs of impact which correspond to a logical relationship between two nodes. If the attacker has access to the originating node, then at least one of the properties of the destination node is modified,
3. the operational arcs that are taken into account to find attack patterns against the modeled infrastructure.

Two arcs cannot have the same originating node, the same destination node and the same type.

4.5.2 Hierarchical arcs

The hierarchical arcs correspond to a hierarchical relationship between two nodes. The destination node always belongs to the originating node, and at least one of the endpoints is an informational component. These arcs allow to structure and organize the model, and therefore to understand it better. But they are not taken into account to find attack patterns against the modeled infrastructure.

The hierarchical arcs are divided into three categories:

1. the belonging arcs, which represent the functional hierarchy. For example, a service belongs to a society, an equipment belongs to a network,
2. the arcs of location, which represent the geographical position. For example, an equipment or a person are located in a room,
3. the arcs of possession, which represent the possession or the use. For example, an equipment is owned or used by a person.

These arcs do not have properties. No value shall be associated to these arcs.

4.5.3 Operational arcs

The operational arcs represent a link between two nodes that can be used during an attack. An arc representing an operational link will always have operational components as its endpoints.
Several values are associated to each operational arc:

1. an effort (0 to 10), which is required to compromise the link between the originating node and the destination node. According to the categories of the nodes, it is possible to answer the following questions:
   (a) What effort shall the attacker made to break the relationship between two nodes?
   (b) What effort shall the attacker made to have access to the destination node from the originating node?
2. a cost (in €), which is required to compromise the link between the originating node and the destination,
3. a time (in minutes), which is required to compromise the link between the originating node and the destination,
4. a grade of discretion (0 to 10), which indicates if the attack can be detected,
5. a grade of protection (0 to 10), which indicates if the attacker can be identified.

The objective, the discretion and the protection are criteria which can be used to make the attacks more realistic in view of the context of the targeted infrastructure.

We distinguish three types of attack:

1. the physical attacks,
2. the computing attacks,

3. the human attacks.

Therefore, it seems natural to divide the operational arcs likewise, and if it is possible to attack a component physically, computationally or humanly, an arc of each category of attack could be traced, as seen in Figure 4.6 (page 61).

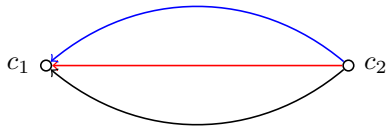

Figure 4.6: c_1 depends on c_2

But this idea was not kept since the model quickly becomes very difficult to understand, as seen in Figure 4.7 (page 61).

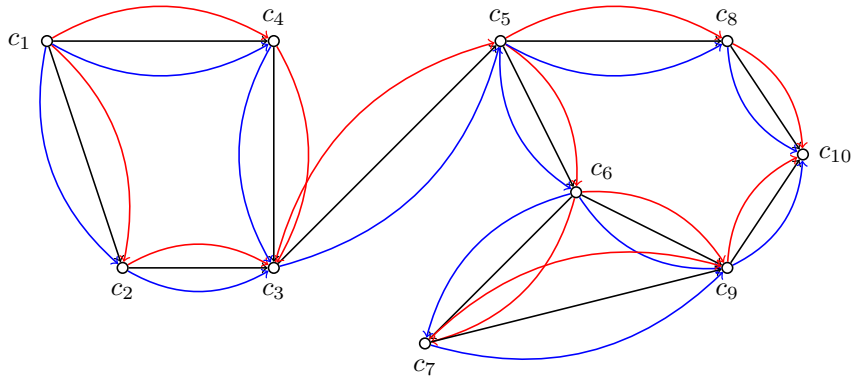

Figure 4.7: Model of a very simple infrastructure

So, instead of tracing an arc for each category of operational arcs between two nodes, only one arc is traced, and the InfraSec tool takes into account three grades of effort, one for each considered type of attacks (human, IT and physical), three costs, three times of execution, three grades of discretion and three grades of protection instead of one.

The characteristics of an operational arc are:

1. the type of the originating node, from where the attack starts,
2. the type of the destination node, the target of the attack,

3. three grades of effort, one for each type of attack
4. three average costs,
5. three average times of execution,
6. the objective: infringement of integrity, confidentiality or availability,
7. three grades of discretion: can the attack be detected?
8. three grades of protection: can the attacker be identified?

For now, only the five first attributes are taken into account in the algorithms which find attack patterns.

A grade of effort is relative to:

1. the table of equivalences ability/vulnerability given in Figure 4.4 (page 56),
2. the vulnerability of the destination node,
3. the corresponding attacker's ability.

The attacker must give an effort equivalent to an attack whose level is relative to the vulnerability. Then, the computation formula for the effort is the following:

$$\text{Effort} = 11 - \text{vulnerability}.$$

Nevertheless, a type of attack (human, IT or physical) and so the grade of effort is not taken into account if:

1. the vulnerability of the destination node is 0,
2. the ability of the attack is 0,
3. the calculated grade of effort is inferior or equal to the ability of the attacker.

Finally, InfraSec matches the arc with the most favorable grade of effort, in this case, the weakest one.

The InfraSec tool has a database which contains the costs and the time of execution of every step of an attack. Finally, the cost and the time of execution should be adapted (there is still no computation formula) to the attacker's abilities and the vulnerability of the targeted node.

4.5.4 Arcs of impact

The arcs of impact represent a causal relation between two nodes which implies a modification of the vulnerability of the destination node. If the attacker has access to the originating node, then at least one of the properties of the destination node is modified.

There is only one type of arc in this group: the arc of impact. The properties of this type of arcs are not yet clearly defined. The idea is to describe how a node can have an impact on the values of the properties of an other node. For example, in the case of a key and a door, if the attacker has access to the key, the vulnerability of the door should be maximal as the door is no longer protected.

Our best lead is to associate a percentage to an arc of impact and recalculate the values of each operational arc whose originationg point is the destination node of the arc impact according to the following computation formulas:

$$\text{Effort} = \text{Effort} - \text{Effort} \times \text{Percentage}$$
$$\text{Cost} = \text{Cost} - \text{Cost} \times \text{Percentage}$$
$$\text{Time} = \text{Time} - \text{Time} \times \text{Percentage}$$

4.5.5 Three axis of attack

At one point during the development of the InfraSec tool, the three axis of attack that are confidentiality, integrity and availability, mentioned in section 4.5.3, were not considered as characteristics of an operational arc but as a node. Indeed, it was decided that the nodes did not represent a component but an axis of attack.

As previously said, three axis of attack are identified :

1. Attack against confidentiality : it is an attack which aims to gain illegal access to information;

2. Attack against integrity : it is an attack which aims to modify or damage a component without permission;

3. Attack against availability : it is an attack which aims to make a component unavailable to something or someone which is supposed to have the right to access.

This deconstruction would allow to have more realistic results. Indeed, it is highly likely that the consequences of an attack against the integrity of a component are not the same than the consequences of an attack against the availability of this component. In other words, the confidentiality of a component does not necessarily depend on the same things than the integrity or the availability of this component.

However, this idea is still not incorporated in the InfraSec project as it was not possible to have a readable representation of complex infrastructures with so many components on the InfraSec tool. Therefore it was difficult to draw firm conclusions when analyzing the algorithms results. Furthermore, for each arc, we had a 3×3 matrix to fill, corresponding to the attack types (human, physical and numerical) and to the axis of attack (confidentiality, integrity and availability), whose input was complex and took too long for the users to analyze.

Conclusion

In order to model infrastructures, the first part of this thesis defines the main notions which are indivisible of the topic of infrastructure security, including the definition of a critical infrastructure. This gives us a better idea of what we need to model in order to best represent an infrastructure. The study of these notions also shows that the predominance of the cyber term can have terrible consequences on the security of an infrastructure.

We give an overview of how an infrastructure can be modeled to evaluate its security. We distinguish two main categories of graph-based models: the models which depict the attacks and the models which depict the infrastructure. For each category, the graph and the tree versions of the models are presented. We also explain why we favor a graph-based model which depict the infrastructure for the InfraSec project.

Then a presentation of the InfraSec model is made. We explain how exactly the graph-based model represents an infrastructure: what does a node model? What does an arc model? Which values, Boolean and continuous, are associated to them?

The major difficulty encountered in developing a richer model of infrastructure is its ability to describe it. Indeed, the richer the model is, the more it can describe the infrastructure and the adversaries that attack it. The counterpart is its exponential character. We therefore expect that, for instance, the problem of highlighting the least costly attack path is equivalent to problems that can not be solved in reasonable time (NP-hard). The locks to be lifted will therefore consist in the design of heuristics to answer these problems in finite time with an "acceptable" response.

Part II

How do we evaluate an infrastructure security?

Introduction

After the presentation of the infrastructure model, it is time to know how this model can evaluate the security of the modeled infrastructure. This model is in fact used as a help to build attack scenarios. Then, the features of these attack scenarios help us to evaluate the security of the infrastructure.

To build attack scenarios, we search for attack patterns in modeling infrastructures. We called **attack pattern** any mathematical structure related to graph theory which give us enough information to evaluate an infrastructure security.
The following mathematical structures have been studied throughout this thesis: minimum path, vertex cover, "colouring problem" and the percolation theory.

In this part, we emphasize the attack patterns which give us promising results: the minimum path and the vertex cover. The minimum path allows to evaluate an infrastructure according to the following philosophy: the cheapest, quickest and easiest the attack is, the less secured the infrastructure is. The vertex cover allows more to evaluate the resilience of the infrastructure. Indeed, vertex cover allows to identify all the components of the infrastructure that have to be targeted in order to paralyze the infrastructure. The more critical components there is, the more resilient the infrastructure is.

First, Chapter 5 discusses the notion of connectivity, which should make it possible to fill any gaps in the audit by connecting as many of the components of the audited infrastructure as possible. Chapters 6 and 7 present the two attack patterns that we have validated: the attack path for the former and the vertex cover for the latter.

Chapter 5

The connectivity property of a graph

As a principal interest, the connectivity property of a graph was studied to see what it can bring from on operational point of view when a graph represents an infrastructure.
As a reminder a graph G is said to be **connected** if, given any pair of vertices v_i, v_j of G, there is a path from v_i to v_j, and a (**connected**) **component** of G is a connected subgraph of G [151].

The idea is to know what can be the consequences of having a more or less connected graph, to know how to evaluate the connectivity property of a graph and how to make a graph more or less connected.

The study of the connectivity property of a graph as part of the security evaluation of the infrastructure it represents is still in its beginning. But we thought it was interesting to introduce this notion in this thesis.

5.1 Operational applications of the connectivity property of a graph

The attacker and the defender will have two opposing points of view on this property. Indeed, when an attacker wants the most connected graph possible in order to have more possibilities of harming the targeted infrastructure, a defender wants the least connected graph possible for the opposite reason.

The following figures illustrate these opposing points of view. The following three figures show graphs representing the same infrastructure. Figure 5.1 (page 72) shows two distinct networks of an infrastructure, denoted N_1 and N_2.

The two networks are perfectly isolated from each other, since given any pair of vertices (v_i, v_j) with $v_i \in N_1$ and $v_j \in N_2$, there is no path between v_i and v_j. The two networks are then secured from the defender's point of view since it is not possible to reach the network N_1 from the network N_2, and vice versa. Therefore the goal of the attacker, who wants the graph to be as connected as possible, is to find the element that can make a link between the two networks. This element may have already been identified by the defender or not. It could simply be a missing link between two identified elements, as seen in Figure 5.2 (page 73), or a missing component, as seen in Figure 5.3 (page 73). This missing component can be an employee or a USB flash drive for example.

This is what happened in 2010, when the Supervisory Control And Data Acquisition (SCADA) systems of the Iranian nuclear program were targeted. The computer worm Stuxnet, which was responsible for causing substantial damage, was introduced via a USB flash drive that was not initially part of the targeted infrastructure [9].

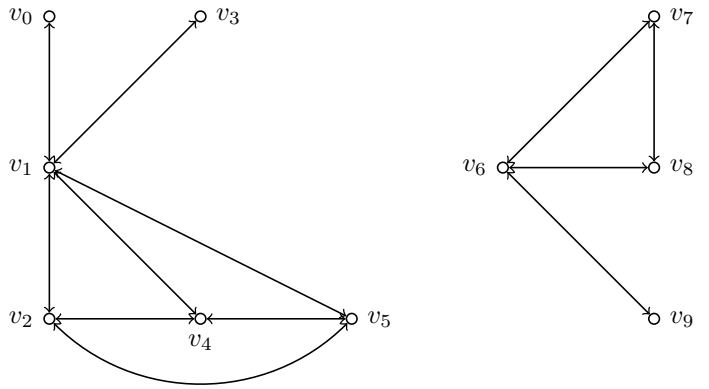

Figure 5.1: Two separate components of a graph

It is easy to know if a graph is not connected to the naked eye when the graph is of reasonable size. It is then possible to quickly evaluate where it would be possible to make the graph more or less connected. But when a graph has hundred or thousand of vertices, as it will be the case with the InfraSec project, it is not possible to do the same and tools are then needed to know if a graph is connected or not.

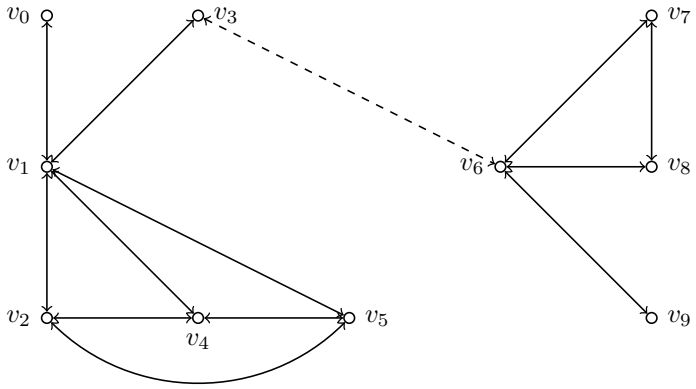

Figure 5.2: Connecting two separate components through a missing link

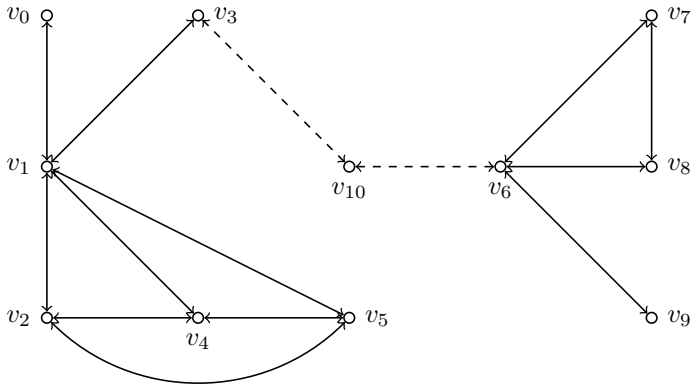

Figure 5.3: Connecting two separate components through a missing or added element

5.2 Algebraic tools to study the connectivity property of a graph

In this section, we discuss the tools that can be used to evaluate the connectivity property of a graph. We focus on the tools that can be implemented in the InfraSec project.

5.2.1 Adjacency Matrix

The first algebraic tool that comes to mind to evaluate the connectivity property of a graph is the adjacency matrix.

Be G a graph (directed or not) whose vertex-set is $V = \{v_0, v_1, ..., v_{n-1}\}$, the **adjacency matrix** is the $n \times n$ matrix $(A) = (a_{ij})$, whose ij-th entry a_{ij} is the number of arcs from v_i to v_j if G is a directed graph, and whose ij-th entry a_{ij} is the number of edges between v_i and v_j if G is a general graph. Note that if the graph G is undirected, the adjacency matrix (A) of G is symmetric, and if G is simple, the trace of the adjacency matrix (A) of G is 0 [151].

The adjacency matrix of the disconnected graph represented in Figure 5.1 (page 72) is the following matrix.

$$(A) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The adjacency matrix is of the form $(A) = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$ where C and D are two null matrices, as highlighted in the following matrix.

$$(A) = \left[\begin{array}{cccccc|cccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right]$$

It is easy to see that the graph is disconnected thanks to the two null matrices C and D in the matrix (A), but this is mainly due to the annotation of the vertices. Indeed, if the names under the vertices are associated differently, as seen in Figure 5.4 (page 75), the adjacency matrix is quite different and not as easy to read than the previous one, as shown below:

$$(A) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

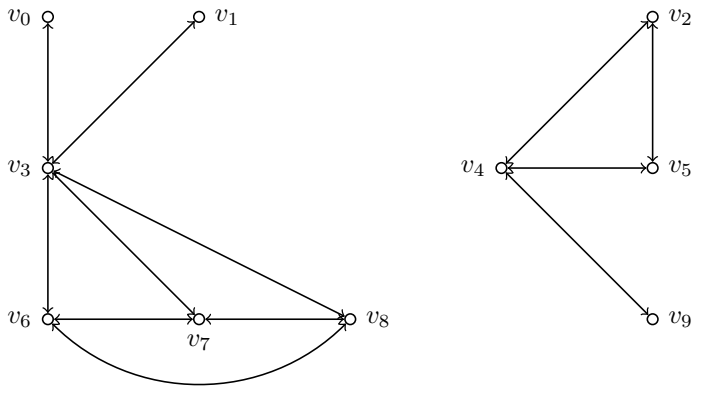

Figure 5.4: A different annotation of the vertices

Of course it is possible to rearrange the rows and columns of this adjacency matrix to fall back on the previous matrix. But if it is easy to determine that a graph is not connected when he has two connected components, what about a graph with 4, 5 or 10 connected components? That does not seem to be a viable solution, especially for graphs with many connected components, so other means need to be explored.

Calculating the k power of an adjacency matrix (A) gives the number of paths of length k between each vertex in the graph. Be G a graph, (A) its adjacency matrix and $(A)^k = (a_{ij})$, then a_{ij} is the number of paths of length k from v_i to v_j if G is directed, and a_{ij} is the number of paths of length k between v_i and v_j if G is a general graph.

For example, let's use again the disconnected graph in Figure 5.1 (page 72). The adjacency matrix is still the following matrix:

$$(A) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Let's calculate $(A)^k$ for $k = 2$ and $k = 10$:

$$(A)^2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 5 & 2 & 0 & 2 & 2 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 1 & 2 & 2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 3 & 2 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 2 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$$(A)^{10} = \begin{bmatrix} 3169 & 9672 & 8377 & 3169 & 8377 & 8377 & 0 & 0 & 0 & 0 \\ 9672 & 31469 & 26426 & 9672 & 26426 & 26426 & 0 & 0 & 0 & 0 \\ 8377 & 26426 & 22523 & 8377 & 22522 & 22522 & 0 & 0 & 0 & 0 \\ 3169 & 9672 & 8377 & 3169 & 8377 & 8377 & 0 & 0 & 0 & 0 \\ 8377 & 26426 & 22522 & 8377 & 22523 & 22522 & 0 & 0 & 0 & 0 \\ 8377 & 26426 & 22522 & 8377 & 22522 & 22523 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 895 & 729 & 729 & 380 \\ 0 & 0 & 0 & 0 & 0 & 0 & 729 & 638 & 637 & 349 \\ 0 & 0 & 0 & 0 & 0 & 0 & 729 & 637 & 638 & 349 \\ 0 & 0 & 0 & 0 & 0 & 0 & 380 & 349 & 349 & 197 \end{bmatrix}$$

In the case of a disconnected graph, there are still coefficients that are equal to 0, regardless of the power k. In this example, we calculate up to the power equal to the size of the graph (10), since it is the size of the longest path possible without loop. Therefore, if there are coefficients that are equal to 0 in all the results of $(A)^k, k \in [1, |V|]$ for the graph $G = (V, A)$, then the graph is said to be disconnected. And the attacker should find a way to either find a new link between existing vertices, as seen in the following example, or find another element which links two components of the graph.

In the case of a connected graph, like in Figure 5.2 (page 73) where a link between two existing vertices is added to Figure 5.1 (page 72), the adjacency matrix is the following matrix:

$$(A) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

We calculate $(A)^k$ for $k = 5$ and $k = 10$:

$$(A)^5 = \begin{bmatrix} 6 & 38 & 24 & 6 & 24 & 24 & 9 & 1 & 1 & 0 \\ 38 & 84 & 86 & 47 & 86 & 86 & 8 & 10 & 10 & 9 \\ 24 & 86 & 68 & 26 & 69 & 69 & 13 & 3 & 3 & 2 \\ 6 & 47 & 26 & 8 & 26 & 26 & 28 & 8 & 8 & 2 \\ 24 & 86 & 69 & 26 & 68 & 69 & 13 & 3 & 3 & 2 \\ 24 & 86 & 69 & 26 & 69 & 68 & 13 & 3 & 3 & 2 \\ 9 & 8 & 13 & 28 & 13 & 13 & 18 & 26 & 26 & 19 \\ 1 & 10 & 3 & 8 & 3 & 3 & 26 & 14 & 15 & 7 \\ 1 & 10 & 3 & 8 & 3 & 3 & 26 & 15 & 14 & 7 \\ 0 & 9 & 2 & 2 & 2 & 2 & 19 & 7 & 7 & 2 \end{bmatrix}$$

$$(A)^{10} =$$

$$\begin{bmatrix}
3327 & 9986 & 8635 & 4010 & 8635 & 8635 & 1676 & 913 & 913 & 683 \\
9986 & 33242 & 27256 & 11662 & 27256 & 27256 & 6519 & 2589 & 2589 & 1676 \\
8635 & 27256 & 22985 & 10166 & 22984 & 22984 & 4738 & 2149 & 2149 & 1531 \\
4010 & 11662 & 10166 & 5253 & 10166 & 10166 & 2626 & 1748 & 1748 & 1243 \\
8635 & 27256 & 22984 & 10166 & 22985 & 22984 & 4738 & 2149 & 2149 & 1531 \\
8635 & 27256 & 22984 & 10166 & 22984 & 22985 & 4738 & 2149 & 2149 & 1531 \\
1676 & 6519 & 4738 & 2626 & 4738 & 4738 & 3473 & 1785 & 1785 & 950 \\
913 & 2589 & 2149 & 1748 & 2149 & 2149 & 1785 & 1338 & 1337 & 835 \\
913 & 2589 & 2149 & 1748 & 2149 & 2149 & 1785 & 1337 & 1338 & 835 \\
683 & 1676 & 1531 & 1243 & 1531 & 1531 & 950 & 835 & 835 & 560
\end{bmatrix}$$

In this case, we see well that adding a link between two nodes who was not in the same component was enough to transform the initial disconnected graph into a connected one since the initial graph has only two components.

Note that the adjacency matrix shows that there can have several paths between two vertices of a graph. For example, according to the last matrix, there are 9986 paths of length 10 between the vertex v_0 and the vertex v_1. This means that an attacker can adapt its attack scenario according to the reactions of the targeted infrastructure during the attack.

Adjacency matrices are a good tool to study the connectivity property of a graph, but as such they are not very easy to use and would require too long calculations in the case of very large graphs since we have the calculation of the graph size power matrix in the worst cases. Of course the Strassen algorithm which allows to optimize the multiplication of two matrices can be used [24] along with the exponentiation by squaring which is a method used for fast computation of large square matrix (also referred to as square-and-multiply algorithms) in order to improve the complexity. However, this may not be enough and therefore, it is necessary to search for other means to study this property.

First, instead of calculating the matrix at the power of the graph size, it is possible to simply calculate the matrix at the power of the graph diameter. Be G a graph, the diameter of G is the maximum value of the distance function and is denoted $d(G)$. The distance function is defined as the number of edges traversed in the shortest walk joining two vertices of G, v_i and v_j, and is denoted $\delta(v_i, v_j)$. Therefore, it could be more interesting to calculate $A^{d(G)}$ instead of A^n with n the number of vertices of G, since $d(G) \leq n$. For that, it must be not constraining to calculate the diameter of a graph. To do so, we had to find the length of the shortest walk between each pair of vertices of the graph. The FloydWarshall algorithm can be used for this since it finds the lengths of the shortest paths between all pairs of vertices of the graph by comparing all possible paths through the graph between each pair of vertices. It is able to do this with $\Theta(|V|^3)$ comparisons in a graph G, where V is the set of all vertices of G [23].

v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
True	True	True	True	True	True	True	True	True	True

Figure 5.5: Results of the DFS algorithm from v_0 for the graph in Figure 5.2 (page 73)

v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
True	True	True	True	True	True	False	False	False	False

Figure 5.6: Results of the DFS algorithm from v_0 for the graph in Figure 5.1 (page 72)

Then, we have also studied several leads linked to positive matrices, irreducible matrices and primitive matrices but until now, none of them were conclusive.

5.2.2 Algorithms for traversing graphs

The next algebraic tools that comes to mind to study the connectivity property of a graph are the algorithms for traversing graphs. The best known algorithms are the depth-first search (DFS) algorithm and the breadth-first search (BFS) algorithm [24]. The first one explores from an arbitrary vertex as far as possible along each arc before backtracking, and the second one explores from an arbitrary vertex the neighbor vertices first, before moving to the next level neighbors. Both have a polynomial complexity ($\Theta(|V| + |A|)$ for a certain category of graphs, where V is the set of all vertices and A the set of all arcs).

By associating a boolean to each vertex of the graph representing an infrastructure (the boolean indicates whether or not the vertex was visited during the execution of the algorithm : it says True if the vertex was visited and False if the vertex was not visited), it is possible to know whether or not the graph is connected. For example, the table in Figure 5.5 (page 79) shows the results of the DFS algorithm from v_0 for the graph in Figure 5.2 (page 73). Since all the booleans associated to the vertices of the graph are True, it means that the graph is connected.

If the graph is not connected, then we have identified a first connected component of the graph. And by relaunching the algorithm from an un-visited vertex, we will be able to identify a second connected component, and so on, until all the vertices are marked as visited. For example, the table in Figure 5.6 (page 79) shows the results of the DFS algorithm from v_0 for the graph in Figure 5.1 (page 72). Here, since only some of the booleans associated to the

v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
False	False	False	False	False	False	True	True	True	True

Figure 5.7: Results of the DFS algorithm from v_6 for the graph in Figure 5.1 (page 72)

vertices of the graph are True, the graph is not connected. But a connected component is identified: the set $\{v_0, v_1, v_2, v_3, v_4, v_5\}$. And by relaunching the algorithm from an un-visited vertex (v_6), a second connected component is identified: the set $\{v_6, v_7, v_8, v_9\}$ (see the table in Figure 5.7, page 80).
By knowing all connected components of a graph representing an infrastructure, a defender will ensure that they remain as they are, while the attacker will find a way to connect them together.

This concludes the section on the different leads we studied to evaluate the connectivity property of a graph. We have presented the most interesting leads and we still have to implement them within the framework of the InfraSec project.

In this chapter, we have shown why it is interesting to study the connectivity property of a graph as part of the security evaluation of the infrastructure it represents. We explained that a connected graph favors an attacker since it means more possibilities for harming the infrastructure, while a disconnected graph favors the defender for the opposite reason. We have mainly developed the attacker point of view but the defender point of view is also interested. Indeed, it could be interesting to look for very dense areas of a graph representing an infrastructure in order to find isthmus or cut-vertex. The goal would be to "erase" them in order to have the less connected graph possible without jeopardizing the proper functioning of the infrastructure (see the vertex cover study in chapter 7).
To date, the only tools to evaluate the connectivity we have studied so far are the adjacency matrix and the algorithms for traversing graphs. Many of the leads have not been conclusive and we are still looking for efficient ways to evaluate the connectivity property of a graph. As said previously, we are just at the beginning of the study.

The study of the graph connectivity property should take place during the infrastructure audit, or at least just after an initial information gathering, in order to have leads on where to dig to get as much information as possible and thus have a graph as connected as possible before launching the algorithms for finding attack patterns.

Chapter 6

Shortest path

The first studied attack pattern is the minimum path in a graph, also widely known as the shortest path. Therefore, the objective here is the search of optimal paths in an attack scenario defined by the aim of the attacker. The question we want to answer is the following: What is the cheapest, quickest and easiest attack path between the attacker and its target(s)? We say that an infrastructure is not secured if it is inexpensive, fast and easy to attack it. This allows to have a realistic approach of the evaluation of an infrastructure security.

Before studying shortest path as an attack pattern, we should explain what a shortest path is and how it can be found in a graph. For that, we take interest in the shortest path problem, but we will first start with the path problem.

6.1 Path problem

The information of this section comes from *"Graphs and Hypergraphs"*, written by Claude Berge in 1976 [11].

The **path problem** is defined as finding (as quickly as possible) a path from a given vertex v_1 to a given vertex v_2 in a l-graph $G = (V, A)$.

If we consider a simple graph $G = (V, E)$, the **chain problem** is similarly defined as finding a chain from a given vertex v_1 to a given vertex v_2 in the graph G. Note that the chain problem becomes a path problem in the l-graph $G^* = (V, A)$ obtained from $G = (V, E)$ by replacing each edge in E by two oppositely directed arcs.

Well known examples of path problems are often building like the following one, strongly inspired from an example that can be found in "Graphs and Hypergraphs" [11]. A hunter, a wolf and a child arrive simultaneously at a river

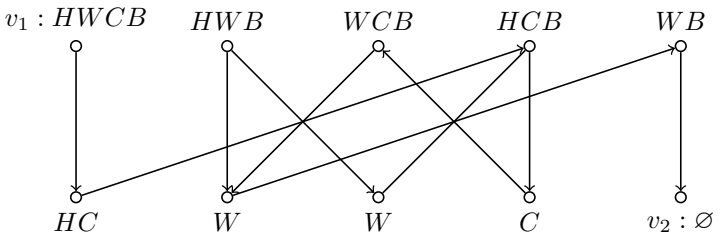

Figure 6.1: The different states of the "hunter, wolf and child" problem [11]

bank and want to go to the other side of the river. The ferry boat can have one of them on board at each passage as it has only two seats including one for the boatman. Furthermore, the boatman cannot leave the hunter and the wolf alone together, whether on the left bank or on the right bank, nor can he leave the wolf and the child alone together. Then how should the boatman arrange their passage across the river?
A graph of the various states can be constructed, see Figure 6.1 (page 82). In the graph, the state v_1 represents the hunter H, the wolf W, the child C and the boatman B all on the right bank, and the state v_2 represents all of them on the left bank. In order to solve this problem, we want to find a path from v_1 to v_2.

Two families of algorithms are considered to solve path problems: the **systematic algorithms** and the **local algorithms**.

Be $G = (V, A)$ a graph, a **systematic algorithm** for finding a path from a vertex $v_1 \in V$ to a vertex $v_2 \in V$ consists of an algorithm which find all the elementary paths of G and see if one of them have for initial endpoint v_1 and terminal endpoint v_2. It is applicable when the graph is already known.

Indeed it is always possible to find all the elementary paths starting from vertex v_1 by constructing all the different arborescences, or tree structures, rooted at v_1. Such an arborescence for the above example is shown in Figure 6.2 (page 83). We can see that there is two paths from v_1 to v_2.

Several systematic algorithms have been proposed over the years. Such algorithms are guaranteed to find a solution, if one exists, or to prove that the problem is insoluble. But they take a very long time to do so.

Another way to solve path problems are the local algorithms. Ideally, a local algorithm will not trace through the entire graph.
A **local algorithm** is a distributed algorithm that runs in constant time, independently of the size of the graph [142].

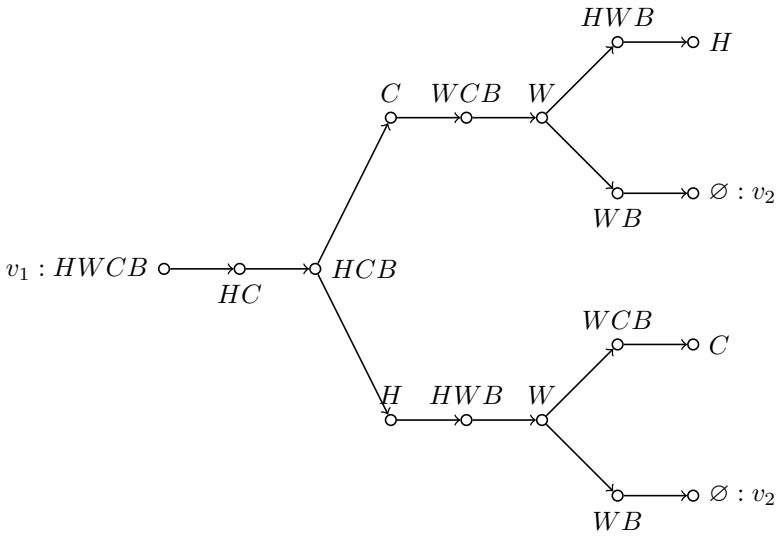

Figure 6.2: Arborescence rooted at a

The Trémaux algorithm [80] and the algorithm of P. Rosensticht and J.C. Bermond are well-known local algorithms.

6.2 Minimum path problem

The information of this section comes from *"Graphs and Hypergraphs"*, written by Claude Berge in 1976 [11], the third edition of *Introduction to graph theory* written by Robin J. Wilson in 1985 [151].

The **minimum path problem**, also known as the shortest path problem is the following: consider a graph $G = (V, A)$, and for each arc $a \in A$, a number $l(a) \geqslant 0$, called the length of a, find an elementary path μ from $v_1 \in V$ to $v_2 \in V$ that minimizes:

$$l(\mu) = \sum_{a \in \mu} l(a).$$

Again, two families of algorithms are considered to solve path problems: the **systematic algorithms** and the **local algorithms**.

Be $G = (V, A)$ a graph, a **systematic algorithm** for finding the shortest path from a vertex $v_1 \in V$ to a vertex $v_2 \in V$ consists of an algorithm which find all the elementary paths of G (by constructing all the different arborescences rooted at v_1 for example) and see if one of them have for initial endpoint v_1 and

terminal endpoint v_2. If there are several of them, the minimum one is chosen for a given function of cost, or time, or effort, or one which combines these three criteria.

And there are the local algorithms which, ideally, will not trace through the entire graph, like the Dantzig algorithm (1960) or the Dijkstra algorithm [39].

6.3 Attack path

As a reminder, an infrastructure is modeled by a weighted directed graph $G = (V, A, c, t, e)$ whose vertices $v \in V$ represent the components of the targeted infrastructure and whose arcs $a \in A$ represent the links of dependency between two components, an application of cost $c : A \to [0, \infty)$, an application of time $t : A \to [0, \infty)$ and three applications of effort $eI : A \to [0, 10]$ with $I = H, T, P$. Furthermore, an attacker is modeled by a node and is considered as a component of the infrastructure. In order to characterize the attacker, three grades of ability are defined: a grade of physical abilities, a grade of social engineering abilities, and a grade of technical abilities. We are looking for paths between the attacker and critical component(s) to build attack scenarios whose features (cost, time and effort) will be useful to evaluate the infrastructure security. These paths between an attacker and critical component(s) of an infrastructure are called attack paths.
A first study of this attack pattern was first published in 2015 [46].

We define an **attack path** aP of length n as a sextuplet (P, c_P, t_P, eP_P, eH_P, eT_P) where P is a path, i.e. a finite sequence of arcs of the form $(a_0 = (v_0, v_1), a_1 = (v_1, v_2), ..., a_{n-1} = (v_{n-1}, v_n))$ in which all the arcs and vertices are distinct, c_P is the cost of P, t_P is the time of execution of P, eP_P is the physical effort of P, eH_P is the human effort of P and eT_P is the technical effort of P. The initial endpoint of P is the attacker and the terminal endpoint is a critical component. There must be an operational link between the $i-1$th vertex and the (i)th vertex.

The cost c_P is the sum of the costs of every arc of the path P, then we have:

$$c_P = \sum_{i=0}^{n-1} c_{a_i}.$$

The time of execution t_P is the sum of the times of execution of every arc of the path P, then we have:

$$t_P = \sum_{i=0}^{n-1} t_{a_i}.$$

The physical effort eP_P is the maximum value of all the physical efforts of every arc of the path P, the human effort eH_P is the maximum value of all

the human efforts of every arc of the path P and the technical effort eT_P is the maximum value of all the technical efforts of every arc of the path P, then we have:

$$\begin{aligned} eP_P &= \max_{0 \leq i \leq n-1} eP_{a_i}, \\ eH_P &= \max_{0 \leq i \leq n-1} eH_{a_i}, \\ eT_P &= \max_{0 \leq i \leq n-1} eT_{a_i}. \end{aligned}$$

The attack path enables us to identify all the components and the associated operational links which have to be compromised. It also gives the order of the vertices which have to be compromised to reach a critical component from the attacker. But how resolving shortest path problem can help us building attack scenarios?

The security of an object is often defined as the security of its weakest link. We generalize this approach to infrastructures differently. In our context, the security of an infrastructure is not defined by the security of its weakest component but by the values of the features of an attack path. This approach is close to the search of shortest paths in a graph. Indeed, the shortest path between two nodes is not necessarily a sequence of local shortest paths. An infrastructure is then said not secured if it is inexpensive, quick and/or easy to attack it.

Before evaluating an infrastructure security, we define a **threat**, i.e. a sextuplet (target, cost, delay, physical effort, human effort, technical effort). The analyzed infrastructure is then considered vulnerable if there is an attack path which allows the attacker to reach the target while respecting the limits of cost, time and effort.
These limits of cost, time and effort will be different from one infrastructure to an other. These limits cannot be the same for a start-up than for a multinational corporation.

More precisely, the security of an infrastructure is said compromised if:

1. the cost of the cheapest attack path is below the cost of the threat,
2. the time of execution of the quickest attack path is shorter than the delay of the threat,
3. the physical effort of the easiest attack path is less important than the physical effort of the threat,
4. the human effort of the easiest attack path is less important than the human effort of the threat,
5. the technical effort of the easiest attack path is less important than the technical effort of the threat,

6. the five previous conditions are met,
7. the grade, calculated from a formula which takes into account the five features, of the "shortest" attack path is below the grade of the threat.

For the last item, we suppose that a systematic algorithm is applied to the paths between the attacker and the target. Let's say that m paths are found between the two, then the formula which takes into account the five features is the following:

$$Grade_P = coef_e \times \frac{\frac{1}{3}\times(eP_P+eH_P+eT_P)}{\sum_{i=0}^{m}\frac{1}{3}\times(eP_{Pi}+eH_{Pi}+eT_{Pi})} + coef_c \times \frac{cP_P}{\sum_{i=0} cP_{Pi}} + coef_t \times \frac{tP_P}{\sum_{i=0} tP_{Pi}}$$

where the coefficients $coef_e$, $coef_c$ and $coef_t$ are precised by the auditor to show which aspect (physical, human or technical) he wants to privilege. The sum of these three coefficients must be equal to 1. By default, the values of the coefficients are set to $\frac{1}{3}$. We assume that the attacker is equally competent in all three domains.
This formula was partially based on the weighted arithmetic mean.

Once the attack path is computed, it is up to the auditor to build an attack scenario based on the attack path, the collected information, his experience and his imagination. An example is shown in the following section.

The major lock of this step is the computational complexity of the algorithm allowing to find paths of minimum cost, time and/or effort. Once the lock is lifted, it becomes possible for us to design heuristic algorithms to respond to a problem of optimization of a constrained cost function.

6.4 A realistic example

In this section, we show how it is possible to find attack scenarios based on attack paths. The following example is based on the one studied by Eric Filiol and Frederic Raynal [49]. This example also gives a brief overview of how information on an infrastructure can be collected.

The aim of the attack is to delay the departure for a mission of the military ship located in Riencourt. The attack has to be performed in a hidden way. Indeed, directly sabotaging the ship is excluded as it would conduct to an investigation. So, secondary targets have to be determined.

The targeted infrastructure is the military ship which has about one or two thousand persons on board, from the officers, the leading seamen, the ordinary seamen to the medical staff (doctors, nurses, medical secretary). With this amount of people, some indiscretions had to be done, especially on the social networks like Facebook, Twitter or Instagram. The collection of information

with the Big Data processing must then not be unsuccessful. As an example, the reader may refer to the article of "Le Monde" about the Israeli Army [87]. For example, one of the Special Forces members was identified because his wife posted a few days before the departure on her Twitter account that she will once again miss her husband as its job sends him away for a few months. In a context of a war, finding the identity of a Special Forces member can allow the enemy to disturb its mission by threatening his family.

It is a known fact that the purpose of the mission is the extraction of expatriates from a country in a state of war. The helicopters are ones of the most useful equipment for this kind of mission. The indiscretions enable to discover that they were about to run out of helicopter pieces and that the delivery had to be done before the departure. A market study allows us to find out which was their only supplier, a company called HeliMeca. In the same time, it was discovered on the same websites that they were about to run out of specific oil, essential for the good functioning of the motors. The only possible supplier is SUD Huiles which is located in the same industrial district as HeliMeca. So, these companies became two of the secondary targets.
It is more difficult to find information on "B to B" companies as SUD Huiles and HeliMeca on the Big Data, except for their employees which can be identified thanks to professional social networks like LinkedIn or Viadeo. Therefore, other methods had to be applied and the social engineering is one of them. The social engineering allows us to discover the existence of tensions between the director of SUD Huiles and one of the shop stewards, which indicates it may be possible to compromise the proper functioning of this company. It was also discovered that HeliMeca have the same trade union than SUD Huiles but there is no proof of tensions within this company so it is probably a dead end in this case.

Thanks to tools as Google Maps, the study of the near environment of the military ship concludes that there is only one practical road to have access to the port where the ship is located : indeed, the delivery vehicles are larger than common vehicles and cannot take the other roads. So it could be possible to stop the delivery of the helicopter pieces and the oil by obstructing this road. So the road became another secondary target. The study also allows us to find out that the road crosses a tough district and the industrial area where SUD Huiles and HeliMeca are located.

With all this information, it is now possible to model the military ship and its environment. The focus was done on the following secondary targets: SUD Huiles, HeliMeca, the crew member and the road which is the only access to the port for delivery vehicles. Figure 6.3 (page 88) shows the modeling of the military ship.

After a preliminary study, the following targets were defined to delay the departure of the military ship: the oil supplier SUD Huiles, the supplier of specific helicopter pieces HeliMeca, the only road which gives access to the port where

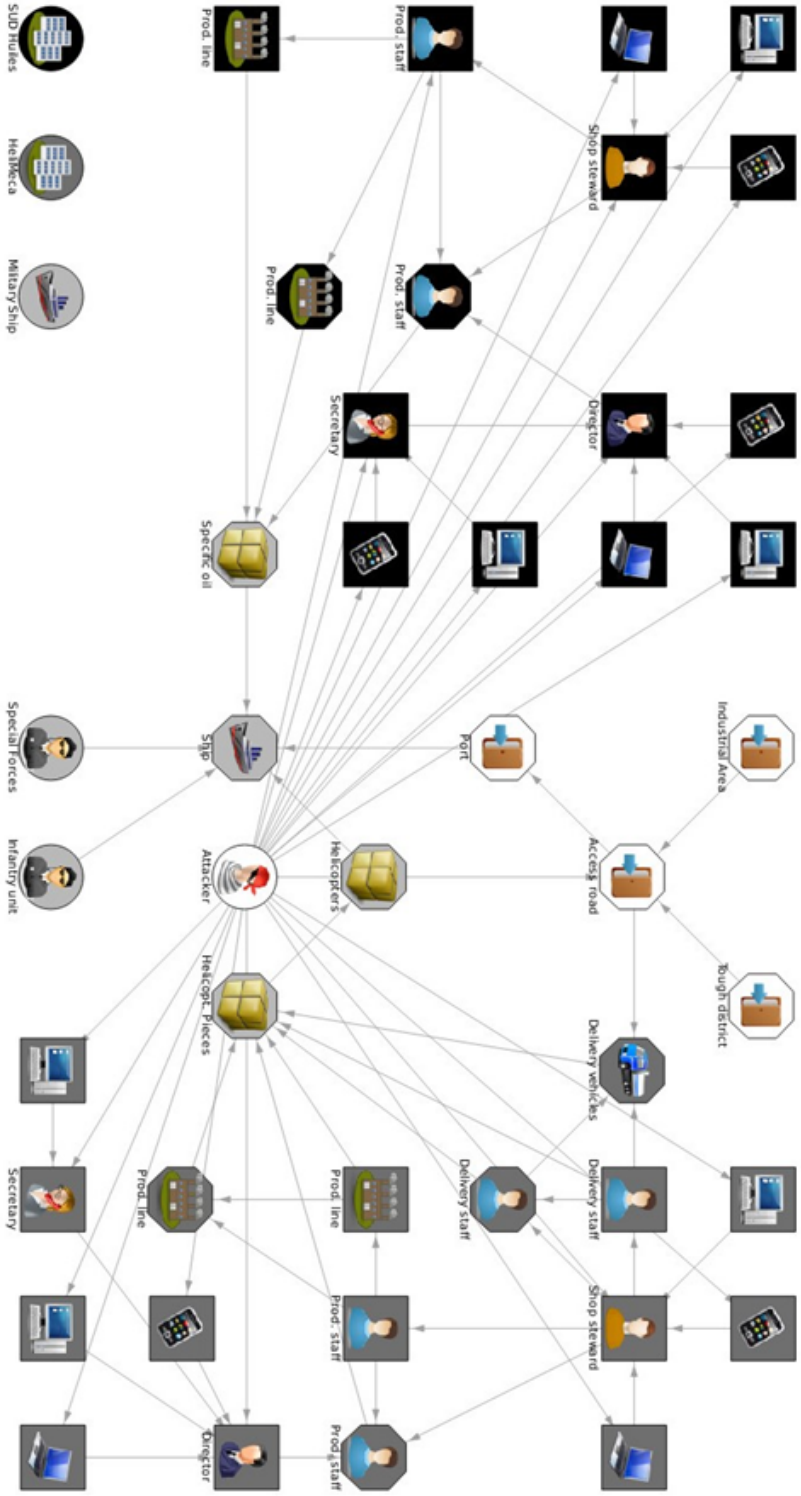

Figure 6.3: Modeling of a military ship

the military ship is located and the Special Forces member.
An attack scenario details precisely the way an attack is launched against a target. It includes the starting points and the steps necessary to reach it.

The search for the attack path is done by an algorithm based on the Dijkstra algorithm. Some modifications had to be made to fit the special features of this model of infrastructure. In our case, we search for the path whose value of dependency is the lowest.

The attack paths computed for SUD Huiles, HeliMeca, the Special Forces member and the road are mostly coherent with the attack scenarios :

1. Special Forces member: confidentiality of the Special Forces member;
2. SUD Huiles : integrity of the director computer, integrity of the director, availability of the production staff, availability of the oil;
3. Access road: availability of the access road (the attacker does have a direct access to it);
4. HeliMeca: availability of the access road, availability of the delivery vehicles, availability of the helicopter pieces.

Some of the functional links/dependencies were also unidentified by the target at first and were found by the attacker (which leads to a more connected graph and more possibilities of attack paths).

Once the attack path is computed, it is up to the auditor to build an attack scenario based on the attack path, the collected information, his experience and his imagination. The following scenarios are deduced from the previous attack paths.

The attack scenario against a Special Forces member is based on the confidentiality of his identity. Indeed, as the identity of a Special Forces member is known, it is possible to threaten his family and to make him lose his self-control, which will necessarily have repercussions on his job.

The SUD Huiles attack scenario starts with a phishing attack launched against its director. If the phishing is a success, the attacker is able to take control of the director computer to put incriminating documents on the common network, in order that one of the employees found them. As the document implied that there will be some dismissals, the existing tensions within the company escalated until it triggered a strike when the director denied the veracity of the document. The strike stopped the production and consequently the delivery of oil.

The most probable access road attack scenario is to cause a car accident to stop the traffic to and from the port. It is not very efficient as it lasts only for

a few hours.

The attack scenario against HeliMeca is based on the availability of the access road. Therefore this scenario is unlikely to succeed as it was seen that it is not possible to obstruct the access road more than a few hours.

Some of the attack paths do not give exploitable attack scenarios. Therefore, a new attack path is computed. In the case of HeliMeca, the following attack path is obtained: integrity of the shop steward computer, integrity of the shop steward, availability of the production staff, availability of the helicopter pieces. According to the information or the lack of it, it does not seem very efficient to target the integrity of the shop steward. But in regards to the attack scenario against SUD Huiles and the fact that the two shop stewards belong to the same union trade, it could be more efficient than the first attack scenario.
The research of other attack path can allow us to find more efficient attack scenarios, but it is also be used to increase the efficiency of the attack itself. In the case of SUD Huiles, combining the two best attack paths (one targeting the director and the other targeting the shop steward) can allow the strike to last longer.

This example of infrastructure shows well the relevance and the importance of the external and the human components in terms of security and in terms of functional economic dependencies, especially for military and governmental infrastructures as well as important companies which are more difficult to attack when it comes to their IT and physical components. Furthermore, the domino effect embodied in the notion of dependency enables to reach an extremely protected component by attacking a less protected component.

6.5 How the search of attack paths is integrated in the InfraSec tool?

The InfraSec project aims to create a tool which allows, among other things, to find attack paths between an attacker and the critical components of an infrastructure.

Firstly, we opted for a local approach of the problem by using the Dijkstra algorithm. At this time, only the cost of an attack path was taken into account to build attack scenarios and evaluate an infrastructure security. Therefore, the Dijkstra algorithm was used to find the value of the cheapest attack path between an attacker and a critical component.

The Dijkstra algorithm is an algorithm which finds the value of the minimum path between two nodes in a graph. Here we adapt the algorithm to find the

value of the cheapest attack path between an attacker and a target in a graph modeling an infrastructure.

Algorithm 1 Adapted Dijkstra algorithm to find the value of the cheapest attack path between an attacker and a target in a graph

Require: A graph $G = (V, A, c)$ where $c : A \to [0, \infty)$ is an application which associates a cost to each arc, an attacker node $v_a \in V$, a target node $v_t \in V$, and an application $\lambda : V \to [0, \infty) \cup \{\infty\}$

Ensure: $\lambda(v_t)$ is the value of the cheapest attack path between v_a and v_t at the end of the computation

1. Let $\lambda(v_i) = \infty \forall v_i \in V$ and $i \neq a$, $\lambda(v_a) = 0$. Let the current vertex $p = v_a$, $C = \{v_a\}$. Move to 2.

2. $\forall v_i \in \overset{+}{\underset{G}{p}}, \lambda(v_i) = min\{\lambda(v_i), \lambda(p) + c(p, v_i)\}$. Move to 3.

3. $p = v_i$ for $v_i \in V - C$ and $\lambda(v_i)$ minimum. $C = C \cup \{p\}$. Move to 4.

4. If $p = v_t$, then stop and $\lambda(p)$ is the cost of the minimal path from v_a to v_t. If $p \neq v_t$, move to 2.

This algorithm was chosen because it always finds the optimal path [76] in a polynomial time. But we had to give up this local approach when we opted for a more dynamic approach of the infrastructure model, which takes into account the links of impact between a component and its key(s) (actual physical key, password, combination, clues about password and combination, etc.).
Indeed, in order to know when a key become useful to use, i.e. when the cost of the attack path between an attacker and the key is amortized by the use(s) of the key, we have to know if and how many time the key can be used. And to do so, we have to know all the vertices of the attack path, which is not the case with the Dijkstra algorithm.

In order to have a more dynamic approach of the infrastructure model, we have to opt for a systematical algorithm which allows to manage the key issue. Furthermore, the time of execution and the efforts of an attack path are now also taken into account to build attack scenarios and evaluate an infrastructure security.

In addition to a path P, to a cost c_P, to a time t_P and to the efforts eP_P, eH_P and eT_P, we associate to the attack path a map M_P, a list of existing keys eK_P, a list of used keys uK_P and a list of forbidden keys fK_P.
The map M_P links a key to the accumulated costs of the edge $\in P$ it have an impact on. The list eK_P is the a list of vertices $(k_0, k_1, \ldots, k_m)$ representing

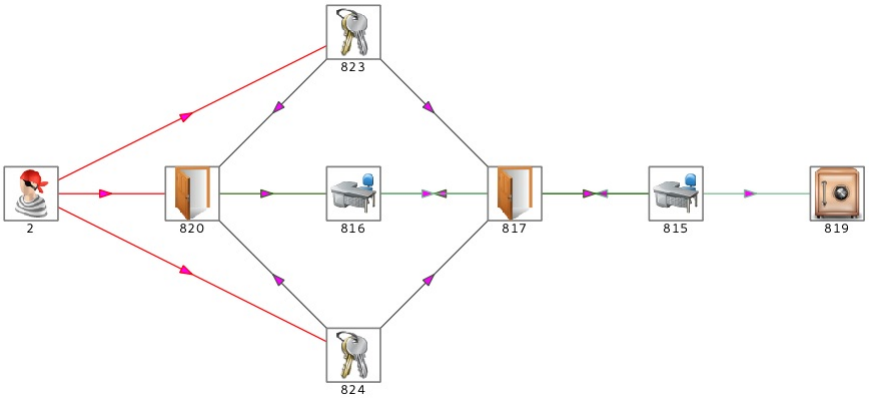

Figure 6.4: Use of the list of forbidden keys

the existing keys that may be used by the attacker to reach the target. The list uK_P is the a list of vertices $(k_0, k_1, \ldots, k_m)$ representing the keys that are used by the attacker to reach the target. The list fK_P is the list of vertices $(k_0, k_1, \ldots, k_m)$ representing the keys which must not be used by the attacker to reach the target.
The last list is useful to make sure that the attack paths obtained are not improbable. More precisely, we do not want to try to obtain a key which is in fact useless. Let's consider the example in Figure 6.4 (page 92), we should be careful that in the end of the calculations, we did not end up with an attack path which considers that the two keys are needed where only one is needed.

As a reminder, a systematic algorithm for finding an attack path from the vertex v_a representing the attacker to a vertex v_t, representing a target, consists of an algorithm which finds all the elementary paths of the graph and sees if one of them has for initial endpoint v_a and terminal endpoint v_t. To do so, we are going to construct all the different arborescences rooted at the target node v_t. We do not construct all the different arborescences rooted at the attacker node v_a as they should be more of them than the different arborescences rooted at v_t in most cases . And since systematic algorithm take a very long time to solve problems, it is a useful trick to gain some time during calculations.

To be fair, we do not exactly build arborescences rooted at the target node v_t since that implies that, starting from v_t, we are looking for all its successors, then the successors of its successors and so on. Since we have arcs directed towards target nodes in our model, it is more accurate to speak of arborescences sourced at the target node v_t which implies that, starting from v_t, we are looking for all its predecessors, then the predecessors of its predecessors and so on.

First our systematical algorithm for finding attack paths between the at-

tacker node v_a and the target node v_t will assign some initial values to the various characteristics of an attack and will update them step by step. The initial values of an attack path are assigned as follows:

1. $P = \emptyset$,$M_P = \emptyset$, $eK_P = \emptyset$, $uK_P = \emptyset$, $fK_P = \emptyset$
2. $c_P = 0$, $t_P = 0$,
3. $eP_P = -1$, $eH_P = -1$, $eT_P = -1$.

Then, a list L of attack paths is initialized to $\emptyset$.

Once this initial step is done, we use a recursive algorithm to update the attack path, predecessors by predecessors. But before we give the recursive algorithm, we are going to explain how exactly the update is done with two algorithms: a basic one and one wich takes into account the use of key. These two algorithms are given page 94 and page 95 respectively.

It is now time to present the recursive algorithm which finds all possible attack paths between two nodes whose cost, time of execution and efforts are below fixed limits. It is inspired from the depth first search algorithm and starts from the target node. When we arrive at the attacker node, we save the current attack path if the values of its features are below the limits. The algorithm ends up when there is no predecessor anymore.

The recursive algorithm calculatePath (page 96) requires the actual node v_c, the attacker node v_a, a graph G which models the targeted infrastructure, an attack path aP and a list of attack paths L.

The advantage of using a systematic algorithm is that we build all the attack path sourced in a target node v_t. An attack path $aP = (P, c_P, t_P, eP_P, eH_P, eT_P, M_P, eK_P, uK_P, fK_P)$ is not saved only if:

1. the initial endpoint of P is not v_a,
2. the values of $c_P, t_P, eP_P, eH_P, eT_P$ are not below the limits fixed by the auditors,
3. the attack path end up improbable (use of two keys when only one is necessary for example).

Therefore, the "shortest" attack path should be among the saved attack paths. But since we have to opt for a systematical algorithm, it may cause long calculations in case of an infrastructure with a high number of components, especially if the representing graph of the infrastructure is dense.
The infrastructures tested on the InfraSec tool did not exceed hundreds of components and the calculations were made easily within a minute.

Algorithm 2 updateWithoutKey Algorithm, a basic update algorithm of an attack path

Require: An attack path aP, a current node v_c, a predecessor node v_i, the arc a from the node v_i to the node v_c and a set of keys K

Ensure: Create a new attack path nP between the attacker node v_a and the target node v_t, an updated version of the attack path aP, with $nP = (N, c_N, t_N, eP_N, eH_N, eT_N, M_N, eK_N, uK_N, fK_N)$ and $aP = (P, c_P, t_P, eP_P, eH_P, eT_P, M_P, eK_P, uK_P, fK_P)$

$N = P$
Add a to N

$c_N = c_P + c_a$
$t_N = t_P + t_a$

if $eP_P \geq eP_a$ **then**
  $eP_N = eP_P$
else
  $eP_N = eP_a$
end if

if $eH_P \geq eH_a$ **then**
  $eH_N = eH_P$
else
  $eH_N = eH_a$
end if

if $eT_P \geq eT_a$ **then**
  $eT_N = eT_P$
else
  $eT_N = eT_a$
end if

$fK_N = fK_P$
for $k \in K$ **do**
  Add k to fK_N
end for

Return N

Algorithm 3 updateWithKey Algorithm, a non basic update algorithm of an attack path

Require: An attack path aP, a current node v_c, a predecessor node v_i, the arc a from the node v_i to the node v_c, a key k_a, a set of keys K and three coefficients $coef_c$, $coef_t$ and $coef_e$

Ensure: Create a new attack path nP between the attacker node v_a and the target node v_t which takes into account the key k_a, an updated version of the attack path aP, with $nP = (N, c_N, t_N, eP_N, eH_N, eT_N, M_N, eK_N, uK_N, fK_N)$ and $aP = (P, c_P, t_P, eP_P, eH_P, eT_P, M_P, eK_P, uK_P, fK_P)$

$N = P$
Add a to N

$eK_N = eK_P$
Add k_a to eK_N
$fK_N = fK_P$
for $k \in K$ **do**
  Add k to fK_N
end for

$M_N = M_P$
$costGeneral = coef_c \times c_a + coef_t \times t_a + coef_e \times (eP_a + eH_a + eT_a)$
$M_N(k_a) = M_N(k_a) + costGeneral$

$costGeneral_{k_a} = coef_c \times c_{k_a} + coef_t \times t_{k_a} + coef_e \times (eP_{k_a} + eH_{k_a} + eT_{k_a})$
if $costGeneral_{k_a} < M_N(k_a)$ **then**
  if $k_a \in uK_P$ **then**
    $c_N = c_P, t_N = t_P, eP_N = eP_P, eH_N = eH_P$ and $eT_N = eT_P$
  else $\{k_a \notin uK_P\}$
    $c_N = c_P, t_N = t_P, eP_N = eP_P, eH_N = eH_P$ and $eT_N = eT_P$
    Recalculate c_N, t_N, eP_N, eH_N and eT_N by taking into account that the key k_a is now considered useful
    $c_N = c_P + c_{k_a}, t_N = t_P + t_{k_a}, eP_N = max(eP_P, eP_{k_a}), eH_N = max(eH_P, eH_{k_a})$ and $eT_N = max(eT_P, eT_{k_a})$
  end if
else $\{costGeneral_{k_a} \not\geq M_N(k_a)\}$
  $c_N = c_P + c_a, t_N = t_P + t_a, eP_N = max(eP_P, eP_a), eH_N = max(eH_P, eH_a), eT_N = max(eT_P, eT_a)$
end if

Return N

Algorithm 4 Recursive algorithm calculatePath

```
calculatePath(aP, G, v_c, v_a, L, coef_c, coef_t, coef_e)=
if v_c = v_a then
  Add aP to L
end if
K = buildSetK(v_c, v_a, G, aP)
for Every operational arc a = (v_i, v_c) ∈ G do
  if a ∉ P then
    Decide which type of attack is the best with the formula page 86
    bool = true
    if K = ∅ then
      nP = updateWithoutKey(aP, v_c, v_i, a, K)
      if The limits of cost, time and effort are respected then
        calculatePath(nP, G, v_i, v_a, L, coef_c, coef_t, coef_e)
      end if
    else {K ≠ ∅}
      for Each k ∈ K do
        if k ∈ uK_P then
          bool = false
          nK = K
          Remove k from nK
          for Each v_k ∈ eK_P do
            Remove v_k from nK
          end for
          nP = updateWithKey(aP, v_c, v_i, a, k, nK, coef_c, coef_t, coef_e)
          if The limits of cost, time and effort are respected then
            calculatePath(nP, G, v_i, v_a, L, coef_c, coef_t, coef_e)
          end if
          Break the loop
        end if
      end for
      if bool then
        for Each k ∈ K do
          nK = K
          Remove k from nK
          for Each v_k ∈ eK_P do
            Remove v_k from nK
          end for
          nP = updateWithKey(aP, v_c, v_i, a, k, nK, coef_c, coef_t, coef_e)
          if The limits of cost, time and effort are respected then
            calculatePath(nP, G, v_i, v_a, L, coef_c, coef_t, coef_e)
          end if
          if k ∉ eK_P then
            nK = K
            nP = updateWithoutKey(aP, v_c, v_i, a, nK)
            if The limits of cost, time and effort are respected then
              calculatePath(nP, G, v_i, v_a, L, coef_c, coef_t, coef_e)
            end if
          end if
        end for
      end if
    end if
  end if
end for
```

Algorithm 5 buildSetK algorithm

Require: The actual node v_c, the attacker node v_a, a graph G, and an attack path aP

Ensure: The set K contains all the existing keys for v_c

$K = \emptyset$

for Every arc of impact $a = (v_k, v_c) \in G$ **do**
  if $v_k \notin eK_P$ **then**
    Find the shortest path kP between v_k and v_a
  end if

  if (kP exists $||v_k \in eK_P) \& v_k \notin fK_P$ **then**
    Add v_k to K
  end if
end for

Return K

We are still currently looking for better algorithms, inspiring by the Data Mining use.

The previous algorithms, among others, were implemented in C++ for the InfraSec tool.

From a hundred vertices or more, the graphical representation on the InfraSec tool is quickly unreadable, as seen in Figure 6.7 (page 100). This is why we decided to highlight the representation of algorithm results instead. Figure 6.8 (page 101) shows the main window of the widget showing the results of the minimum attack path algorithm. We see there that the attack paths are grouped into attacks and its variants. Two variants belong to the same attacks if they have the same "dna", i.e. the path of the two attack paths are the same sequence of data, persons, means of access, information technologies, equipment and products, which are the main categories of vertices.

Figure 6.9 (page 102) shows all the features of an attack path, in this case it is the shortest one. The windows shows a graphical representation of the attack path, as well as the cost, time and different efforts required to execute the attack. The sequence of all the vertices that are supposed to be targeted is also given as well as the cost, time and effort for each corresponding arc.

To conclude the study of the minimum path as an attack pattern, we note that attack paths represent attack scenarios as sequential attacks where every

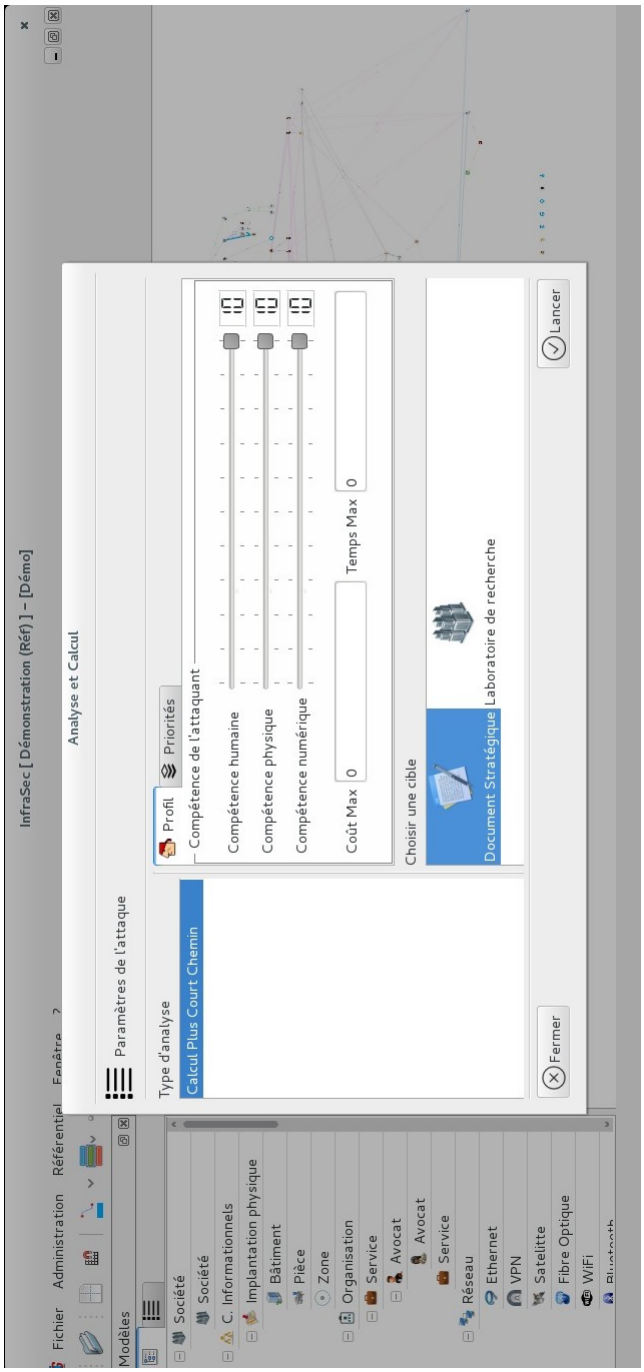

Figure 6.5: Screenshot InfraSec tool - Initialisation of the attacker feature and the target

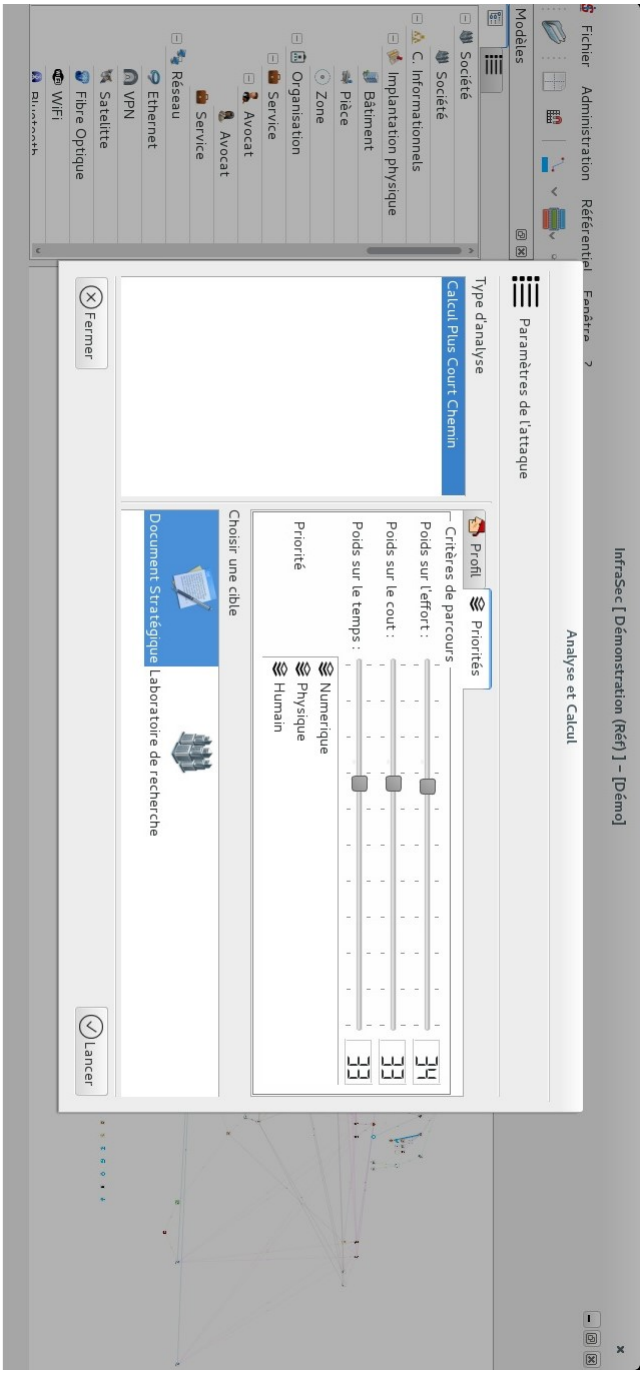

Figure 6.6: Screenshot InfraSec tool - Initialisation of the coefficients

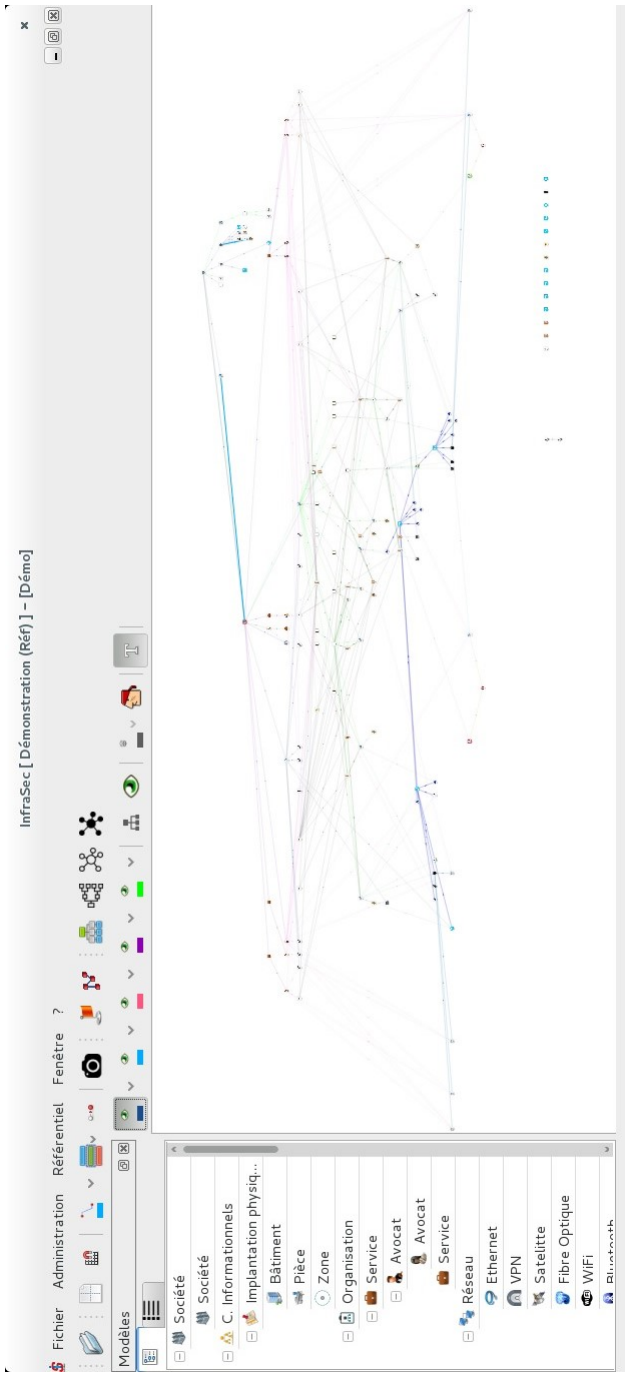

Figure 6.7: Screenshot InfraSec tool - The entire representing graph of an infrastructure with hundreds of components

Figure 6.8: Screenshot InfraSec tool - Main results of the calculation

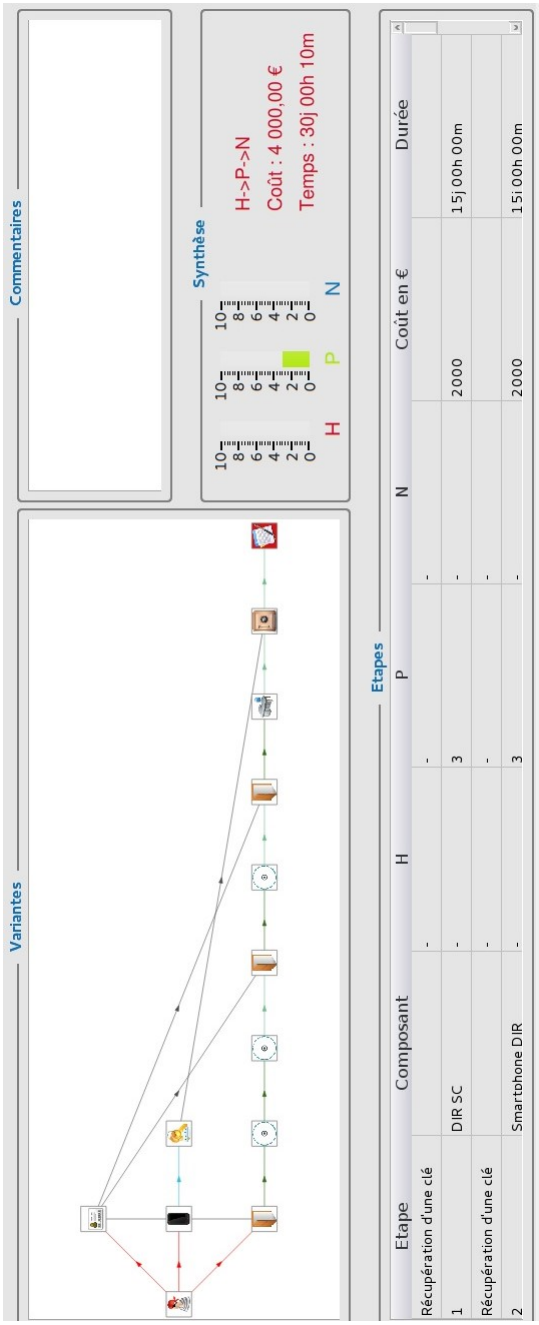

Figure 6.9: Screenshot InfraSec tool - Cheapest attack path

steps depend on the previous ones. The following studied attack pattern is different. This one allows to plan parallel attacks thanks to the identification of several critical components which have to be targeted in the attack scenarios. This parallel approach can lead to more resilient attack scenarios.

Chapter 7

Vertex cover

The study of the vertex cover as a potential attack pattern was motivated by the recent attacks against power lines in Crimea which left three quarters of its population without electricity for several days [81] until three weeks in certain areas since some repair works were delayed due to the presence of demonstrators [93]. These attacks required the destruction of only four pylons to leave most of the 1.8 million residents of the peninsula without electricity.
Furthermore, on the 25th of May 2005, 1.5 million to 2 million customers were deprived of electricity for several hours in Moscow and nearby regions due to a fire and explosion in a local south-eastern substation. The lead of a terrorist attack was ruled out here as the incident was in fact caused by aging equipments which were overburdened by high demand [5]. The failure of this one substation led to a power outage in several areas thanks to a cascade effect [66].
In order to prevent or at least to minimize the effects of this kind of attacks and failures, the components of an infrastructure whose disruption, damage or destruction can lead to its paralysis have to be greatly secured, but first of all, they have to be identified.

The option studied to identify the critical components of an infrastructure is the vertex cover, a particular structure of the graph theory [11].
A first study of this attack pattern has first been published in 2016 [47] and 2017 [54].

7.1 Definition

Be G an undirected graph. G is defined by two sets (V, A) where V is a set of vertices and A is a set of arcs. A **vertex cover** of G is a subset of V, called V', such that every arc $(v_1, v_2) \in A$ contains at least one vertex of V'. It means that $\forall (v_i, v_j) \in A$, with $i, j \in \mathbb{N}$, either $v_i \in V'$ or $v_j \in V'$ or both v_i and $v_j \in V'$. Figure 7.1 (page 106) shows examples of vertex cover, the set V' of each graph

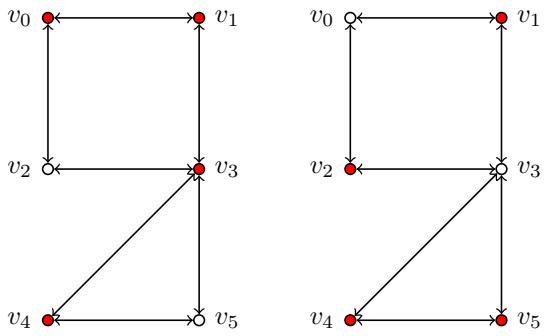

Figure 7.1: Two examples of vertex cover

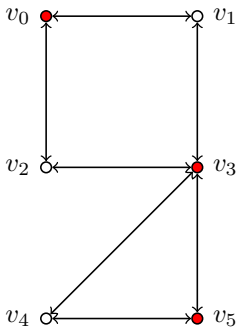

Figure 7.2: A minimum vertex cover

is in red.
The **vertex cover problem** is an algorithmic problem which consists of finding a set V' for a graph $G = (V, A)$ such as $\forall(v_i, v_j) \in A$, with $i, j \in \mathbb{N}$, either $v_i \in V'$ or $v_j \in V'$ or both v_i and $v_j \in V'$. The set V' is said to "cover" all the vertices of G.

As shown in Figure 7.1 (page 106), the same graph can have several vertex cover, therefore a vertex cover is not unique and the vertex cover problem can have several solutions.

A **minimum vertex cover** is a vertex cover of the smallest possible size. Figure 7.2 (page 106) shows one example of a minimum vertex cover for the same graph of Figure 7.1 (page 106).

The **minimum vertex cover problem** is the optimization of the vertex cover problem. It is an algorithmic problem which consists of finding a set of vertices of minimum size to cover all the vertices of a given graph. As well, the

minimum vertex cover problem can have several solutions for a same graph.

7.2 What can a vertex cover bring from an operational point of view?

Be $G = (V, A)$ a graph, G represents an infrastructure. If an attacker, or a group of attackers, corrupts, infiltrates, takes controls or steals all the components included in a vertex cover of G, he has a direct access to all of the others vertices of G.
If an attacker, or group of attackers, destroys, damages or shuts down all the components included in a vertex cover of G, all the others components of the infrastructure end up completely isolated from the others, as seen in Figure 7.3 (page 108).
Consequently, solving the vertex cover problem in a graph representing an infrastructure allows the identification of its critical components, the ones whose corresponding vertices are in a vertex cover, whose disruption, damage corruption, theft, or destruction leads to the paralysis of the entire infrastructure.
These components may not appear critical at first sight, when they were considered individually. Indeed a component may be not critical alone, if the attacker target only this component, but it may be critical if it is targeted along with a set of well chosen components.

The identification of critical components is not the only thing that the vertex can bring from an operational point of view. A vertex cover of a graph is not unique, and then it is possible to identify several sets of critical components and to associate a team of attackers to each of them. To succeed their mission, the different teams do not have to be aware of the existence of the others as well of their particular targets.
Furthermore, even inside of each team, the different members do not necessarily have to know the existence of the others and their particular target (each component whose corresponding vertex is in the vertex cover can be assigned to only one attacker). These dispositions enable to maximize the probability of success (operational redundancy) while minimizing the operational risks as even if one or some of the attackers, or an entire team, are caught, they cannot compromise the rest of the operation.

How exactly can the vertex cover allow to evaluate the security of the targeted infrastructure? This evaluation is based on the following data:

1. the size of the minimum vertex covers of the graph G representing the infrastructure,
2. the feasibility, the cost and the time of execution of the attacks against the components whose corresponding vertices are in the minimum vertex covers.

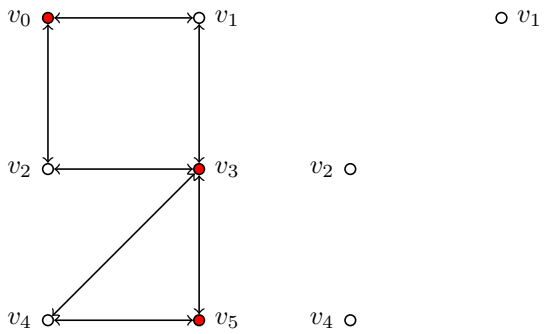

Figure 7.3: Consequence of the destruction of all the components included in the vertex cover

An infrastructure can be said resilient to an attack based on the vertex cover approach if the sum of cost and time, as well as the difficulty of the attack paths targeting all the components in a minimum vertex cover are too important according to a defined threat. The larger the size of the minimum vertex cover is, the more likely the infrastructure is to be resilient.

The "infrastructure resilience is the ability to reduce the magnitude and/or duration of disruptive events. The effectiveness of a resilient infrastructure or enterprise depends upon its ability to anticipate, absorb, adapt to, and/or rapidly recover from a potentially disruptive event" [26].
On the contrary, an infrastructure can be said vulnerable to an attack based on the vertex cover approach if the size of a minimum vertex cover is small compared to the total amount of vertices of the graph and if the attacks against the entire corresponding component are feasible, not expensive or not time-consuming.

Finally, the vertex cover of a graph is not unique, as seen in Figure 7.1 (page 106), therefore if the attack scenario against the first set of targets (corresponding to one chosen vertex cover) fails for one reason or another (because of defense mechanisms of the infrastructure or insufficient information), it is always possible to target another set of components belonging to another vertex cover, which allows the attack to be more resilient. Even if it may still be difficult since the infrastructure may have understood that it was targeted.

7.3 Adaptation of the model to fulfill the constraints of the use of vertex cover algorithms

The definition of the vertex cover requires an undirected graph when the model uses a directed graph to represent an infrastructure. Therefore, as it is highly improbable that all the components of an infrastructure are interdependent, it is not possible or rarely possible at best, to search a vertex cover in the entire graph representing the infrastructure. A solution to this impasse is to consider only a part of the graph, a subgraph whose vertices are all interdependent. Be $G = (X, E)$ a graph. A subgraph $G' = (X', E')$ of G is a graph such as X' is included in X and E' is included in E.
With this action, the initial graph may lost some vertices which are important to the pertinence of its security evaluation, but it still could be interesting to try to solve the minimum vertex cover problem in this reduced graph.
It may also be interesting to consider the entire graph while assuming that every existing links of dependency between two components are bidirectional even if they are not in reality. Some of the results may be misrepresented but still exploitable.
Only the experience will say if these possibilities can be kept to evaluate some infrastructures' security.
On the other hand, networks like the electrical power system, the Internet or the "human network" can be easily represented by an undirected graph.

7.4 A realistic example of the United States electrical power transmission and distribution system

In this section, the electrical power transmission and distribution system of the United States, also known as the "power grid" and previously studied [44], is used to illustrate how the minimum vertex cover problem can be used to plan an attack against an infrastructure in order to evaluate its security.

The U.S. electrical power transmission and distribution system is a network of substations, generating stations, transformers, transmission lines, distribution lines, pylons, and other physical components easily observable. It also includes "devices that sense and report on the state of the system, the automatic and human controls that operate the system and the intricate web of computers and communication systems that tie everything together" [27].
Therefore the graph representing the U.S. electrical power transmission and distribution system is an undirected graph G defined by two sets (V, A) as:

1. V is a set $(v_1, v_2, \ldots, v_n)$ of vertices which represent one of the lines of attack of the substations, generating stations, transformers or pylons in-

cluding in the power grid;

2. A is a set $(a_1, a_2, \ldots, a_m)$ of arcs where $a_k = (v_i, v_j)$ for $k < m$ and $i, j < n$. The arcs represent the interconnected transmission and distribution lines.

The electrical power transmission and distribution systems are a natural target as they are aging infrastructures in most industrialized nation states, which make them vulnerable to attacks. For example, the U.S. power grids suffered of several attacks over the last two years [139]. The vulnerability of the U.S. electrical power grid can be explained by geographical constraints due to the size of this nation state, by financial constraints, and by aging control systems.
The electrical power transmission and distribution system faces different kind of vulnerabilities:

1. physical vulnerability: a great majority of the pieces of equipment in the facilities of the power grid are decades old and lack upgraded technology; and some facilities are easily accessible to attacks;

2. cyber vulnerability: most of the systems are potentially vulnerable to cyber attacks, whether through Internet connections or by direct penetration at remote sites;

3. personnel vulnerability: there is a lack of skilled workers and expert engineers to replace the ones retiring, and there are also attacks from inside the infrastructure.

Furthermore "the electrical power transmission and distribution system are not designed to withstand or quickly recover from damage inflicted simultaneously on multiple components" [27]. This characteristic makes the power grid an example particularly fitted for the vertex cover approach.

The first step of the attack against the electrical power transmission and distribution system consists of mapping it, and therefore of identifying all of its components:

1. generating stations;

2. substations;

3. transformers;

4. emergency back-up generators;

5. transmission lines;

6. distribution lines;

7. pylons;

8. SCADA devices;

9. the substation automation or protection systems,
10. the energy management systems;
11. the market systems;
12. communication systems;
13. personnel of the different facilities composing the power grid.

But also various pieces of information:

1. roads close to the power grid;
2. vehicle weight, dimension and other traffic regulation;
3. response units (rescue teams, fire fighters, police, army, national guard, technicians on place, etc.);
4. maps of the different facilities of the power grid;
5. set up security systems;
6. data about past incidents;
7. data about the weather conditions.

Most of these data are openly available. A part of the collected substations can be seen in Figure 7.4 (page 112), and some of the power lines in Figure 7.5 (page 112).
The first set of elements allows to build the graph representing the electrical power transmission and distribution system when the second set is useful to evaluate the time for the infrastructure to answer to and fix the problems caused by the attack against a particular component, thanks to its accessibility and the number of men who can take care of the problems among others. As the vertex cover is not unique, these pieces of information may be very helpful to identify the "best vertex cover" if the number of available attackers is limited. It may be also important to identify the sections of the grid which correspond to redundant sources or logistic support between the three main U.S. electrical areas.

To sum up, in order to identify the vulnerabilities of the power grid and to create a "knock-on effect" to inflict maximum damages, a few dozen of relevant facilities (electrical pylons and towers, substations, etc.) are identified and different areas with difficult access for response units are spotted. A graph of the grid is then made up with these pieces of information. This graph has a sparse and very simple structure due to the nature of the electrical power grid, in particular with respect to the surrounding geography.

Figure 7.4: Substations of the United States electrical grid

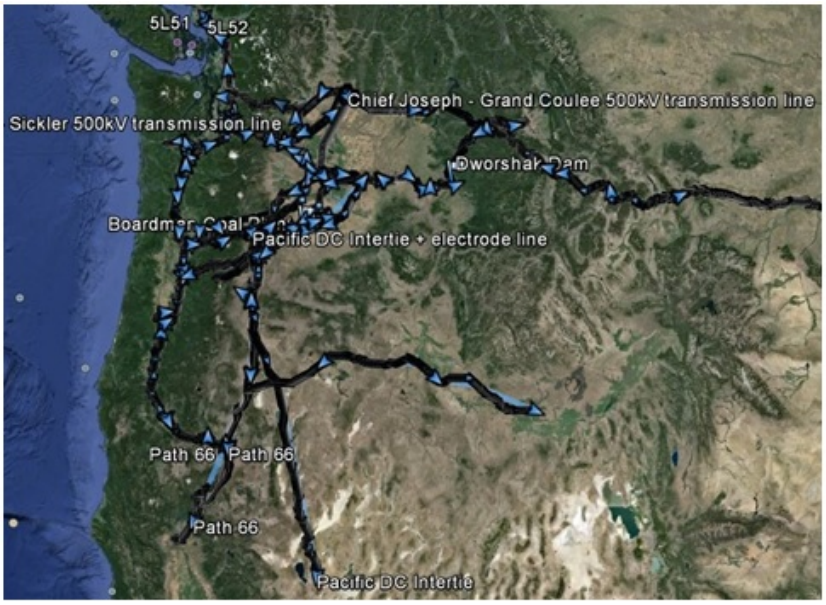

Figure 7.5: Substations of the United States electrical grid

7.5 Vertex cover algorithms and results

In order to identify the critical components of the electrical power transmission and distribution system modeled by the graph G, a minimum vertex cover algorithm has to be used on G.
The minimum vertex cover problem is a NP-complete problem [24], one of Karp's 21 ones [74], which means that it is not possible to find solutions efficiently as the algorithms used to solve them are exponential in the worst case. So, basic algorithms are unusable in practice.
A lot of works were done to improve the complexity of the minimum vertex cover problem algorithms [19] [4] [125]. The retained algorithm is the Dharwadker one of polynomial complexity for certain categories of graph [37]. In order to find a solution of the minimum vertex cover problem for a graph $G = (V, A)$ with this algorithm, G must be simple, which means that G is an undirected graph that has no loops and no more than one edge between any two different vertices.
As the electrical power transmission and distribution system is modeled by an undirected graph with no loops and no multiple edges between two vertices, it is possible to use the Dharwadker algorithm.
As a result, the size of the minimum vertex cover of the graph is proven to be 9, which is quite small considering the size of the grid. Now we have to figure out if the components corresponding to the vertices in the vertex cover are easy to attack or not. For that, shortest path problem algorithms can be used to determinate the best attack paths to take them down, corrupt them or target them if they are not directly accessible by an attacker. In the case of electric power delivery system, many key facilities are unguarded so the critical components appeared to be vulnerable.
Therefore the electrical power transmission and distribution system was proven vulnerable to a vertex cover type of attack as it is possible to get down the entire infrastructure with a reduced team of well-informer attackers.

To conclude on this, the results obtained on the electrical power transmission and distribution system of the United States are very interesting as the size of the vertex cover is very small compared to the number of components of the grid and as it is possible to attack all these components. It shows a great vulnerability of the grid as the smaller the vertex cover is, the easiest it is to paralyze the entire infrastructure. Two recent attacks confirmed the operational reality of the attacks that can be determinate thanks to the vertex cover. Indeed, several high-capacity Internet cables were attacked during the last year in California's San Francisco Bay Area [69]. And the cut of some specific fibre optic connections shut down the Internet in Humbolt [138]. In the two cases, the cables were chosen in order to maximize the effect on the entire network.

7.6 A light vertex cover

Other structures of the graph theory related to the vertex cover have also been studied.

During the study of the vertex cover as an attack pattern, it was also noticed that sometimes, the number of critical components to corrupt, infiltrate, take control or steal in order to paralyze the entire infrastructure can be inferior to the size of the minimum vertex cover. For example, in figure 7.3 (page 108), five critical components do allow to reach all the other components, but they do not form a vertex cover as not all the arcs of the graph representing the infrastructure have an extremity which belongs to the vertex cover.

We define a light vertex cover as follows: be G a graph (V, A), the subset V' of V is a **light vertex cover** of G if $\forall v \in V'$ there is at least one arc of A incident to v whose one of these ends or both of them are included in V'.

So an infrastructure whose minimum vertex cover has a big size could still be endangered if there is a light vertex cover of small size. This structure could be even more interesting if the search of a minimum light vertex cover is easier than the search of a minimum vertex cover. As we have still not found an algorithm for the search of a minimum light vertex cover, it is impossible to answer this question for now.

To sum up this section, the recent attacks in Crimea [93] and in the United States [44] and the incident in Moscow [5] show the importance of the identification of the critical components of the critical infrastructures. It was shown how the vertex cover, a particular structure of the graph theory, enables this identification and how the attack scenarios against these critical components enable to evaluate the security of the targeted infrastructure.
Through the attack scenarios, security vulnerabilities and resilience can be determined and thanks to the real-life features these scenarios bring, it is possible to know whether the vulnerabilities are operationally exploitable or not, and so, whether they may represent a real danger for the infrastructure or not.
The vertex cover also enables to maximize the probability of success (operational redundancy) while minimizing the operational risks by dividing the critical components between different attackers without them knowing the existence of the other targets as well as the other attackers.

Other structures related to the vertex cover were studied with more and less success. Only the light vertex cover seems to have the possibility to have exploitable results. The next step of the reflection will be to develop an algorithm for finding a minimum light vertex cover and see if this problem is also a NP complete problem.

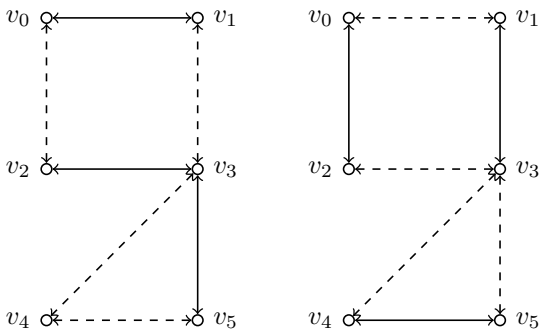

Figure 7.6: Two examples of edge cover

7.7 Aborted leads of research

Some of the studied mathematical structures did not necessarily give conclusive results as attack patterns. For example, the colouring of graph, which consists of colouring the vertices of a graph in such a way that two adjacent vertices do not have the same colour [151], was briefly considered to avoid attack path with two consecutive vertices representing a person (high risk of detection of the attack). In this section we present some of the studied mathematical structures linked to the vertex cover which ended up disappointing.

The edge cover can be seen as the opposite of the vertex cover. The idea was the following: instead of targeting the vertices, the attackers could prefer to target the arcs. In the case of the electrical power transmission and delivery system, it is the transmission lines which are poorly protected and so, easy to attack.
For a graph $G = (V, A)$ where V is a set of vertices and A is a set of arcs, an **edge cover** of G is subset of A, called A', such that every vertex of V is incident to at least one edge of A'.
To find an edge cover of a graph, this graph must have no isolated vertices (an isolated vertex is a vertex which is incident to none of the arcs).
Figure 7.6 (page 115) shows two examples of edge cover. The edge in dashed lines are in the edge cover.

The **minimum edge cover problem** is an optimization problem of finding an edge cover of minimum size. Figure 7.7 (page 116) shows two examples of edge cover. The edge in dashed lines are in the edge cover.
Unlike the minimum vertex cover problem, this problem can be solved in polynomial time. But unlike the vertex cover, taking down all the edges of a minimum edge cover does not necessarily isolate the vertices from each other, as seen in Figure 7.8 (page 116). Therefore this notion is far less powerful than the vertex cover and the resulting attacks against the infrastructure may not be as suc-

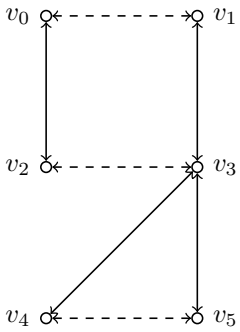

Figure 7.7: A minimum edge cover

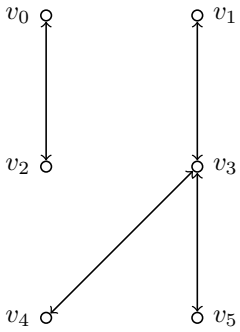

Figure 7.8: Consequence of the destruction of all the links including in the edge cover

cessful and dangerous than the attacks resulting from the vertex cover algorithm.

Another structure was also considered as disappointed considering the efficiency of the resulting attacks: the **vertex cover of the dual graph**. The construction of the dual graph often leads to graphs with loops and multiple edges, therefore it was not possible to apply vertex cover algorithm on them.

Conclusion

In this second part, we introduce the concept of connectivity and how it can be used to increase connectivity between components and provide more opportunities for the attacker to harm the infrastructure. To do so, the graph representing the infrastructure must be studied in order to know if it is connected or not. Several tools are presented: the adjacency matrix and the algorithms for traversing graphs. If a graph is not connected and its connected components are identified, elements that can link the connected components together must be searched in order to have a more connected graph and to provide more attacking opportunities which are translated buy more possible paths between the components of the infrastructure. The defender point of view is also mentioned: when an attacker wants the most connected graph possible to have more opportunities for harming the infrastructure, the defender wants the less connected graph for the opposite reason. To do so, it could be interesting to look for very dense areas of a graph in order to find isthmus or cut-vertex. The goal of the defender would be to "erase" them without jeopardizing the proper functioning of the infrastructure (see the vertex cover study).

The concept of attack pattern is also introduced and the study of two mathematical structures that fit our definition is presented. These two mathematical structures are the attack path and the vertex cover. We show how searching for these attack patterns in a graph representing an infrastructure allows us to obtain attack scenarios and how the characteristics of these scenarios allow us to determine whether an infrastructure is secure or not.
To do so, a threat is defined as a sextuplet (target, cost, delay, physical effort, human effort, technical effort). The analyzed infrastructure is then considered vulnerable if there is an attack path which allows the attacker to reach the target while respecting the limits of cost, time and efforts.
Several algorithms were implemented and integrated in the InfraSec tool in order to find attack paths in a graph representing an infrastructure.
The study of vertex cover also allows the identification of critical components that are not necessarily obvious. Once these components are identified as targets, attack paths are searched between the attacker and them. Algorithms for finding vertex cover are still yet to be implementing in the InfraSec tool.

Chapter 8

Conclusion

Starting from the observation that infrastructures today are far from being secured, the main objective of this thesis was to found new ways to evaluate their security. Understanding what is an infrastructure, what are its components and how they interact with each others was the first task we conducted. This led us to the study of the various and many definitions of a critical infrastructure: more than twenty definitions are presented. We observed that the great majority of these definitions do not mention human components as well as external components and the near environment, when IT components are sometimes too much emphasized. We do not want to minimize the vulnerabilities that can bring IT components but it is important that they should not be emphasized to the detriment of others components which are equally primordial. We had to deal with the same issue concerning the definitions of an attack, which suffer from the importance taken by the term cyber. After these observations were made, we explained how the absence of these components in security protocols can have disastrous consequences. Although we are not the first to raise this problem many have mentioned these omissions before it is still not taken into account in official definitions. Then, we thought useful to raise this issue again. We ended by proposing our own definition of a critical infrastructure since the official ones appear restrictive, static, and local. We tried to have the most exhaustive and realistic definition as possible but it is still opened to discussion as we may also have forgotten some components. There is so many different types of infrastructure that it is easy to forget some elements which are very specific to a certain kind of infrastructure. This work has been published a first time in the proceeding of the 9^{th} International Conference of Cyber Warfare and Security [45] and a second time in the Journal of Information Warfare [48]. Other important notions related to an infrastructure's security as the notions of dependency and resilience have also been defined.

An overview of existing graph-based models was then made. We distinguished two main categories: the ones which describe the attacks and the ones which describe the infrastructure. We opted for the latter as, despite the first

one, it does not require that all the critical components have to be identified first and therefore it allows us to find solutions for the identification of the critical components, which can be needed as the critical components may not be obvious.
In a nutshell, an infrastructure is modeled by a directed graph whose vertices represent the components of the targeted infrastructure and whose arcs represent the links of dependency between two components. An attacker is part of the infrastructure model. Several types of vertices and arcs are defined. The most important distinction is made between the elements used to understand how the infrastructure works and the ones used to evaluate its security. The latter are called operational elements.
As we privilege the attackers point of view, we evaluate an infrastructure security by building attack scenarios. To build these attack scenarios, we look for attack patterns. We call attack pattern any mathematical structure related to graph theory which gives us enough information to evaluate an infrastructure security. Only the operational elements are used to find attack patterns. Several mathematical structures have been studied throughout this thesis. The shortest path and the vertex cover were the ones who give us the most promising results.
The study of the path structure leads us to the definition of attack paths. An attack path is a sextuplet (P, c_P, t_P, eP_P, eH_P, eT_P) where P is a path, c_P is the cost of P, t_P is the time of execution of P, eP_P is the physical effort of P, eH_P is the human effort of P and eT_P is the technical effort of P. We look for paths between the attacker and a critical component of the targeted infrastructure. We use these features to evaluate the security of the infrastructure. For that, we define a threat as a sextuplet (target, cost, delay, physical effort, human effort, technical effort). The analyzed infrastructure is then considered vulnerable if there is an attack path which allows the attacker to reach the target while respecting the limits of cost, time and efforts.
Several algorithms were implemented and integrated in the InfraSec tool in order to find attack paths in a graph representing an infrastructure. In order to have a more dynamic model of infrastructure (with the management of keys), we have to opt for a systematical algorithm, which can cause problems in the case of an infrastructure with a high number of components.
The study of vertex cover, in addition to evaluate the security of an infrastructure, also allows the identification of critical components that are not necessarily obvious. Once these components are identified as targets, attack paths are searched between the attacker and them. Algorithms for finding vertex cover are still yet to be implementing in the InfraSec tool.
In the contrary, some mathematical structures did not give promising results. The vertex cover of the dual of the representation graph was considered as disappointed considering the efficiency of the resulting attacks when the construction of the dual graph led often to graph with loops and multiple edges, therefore it was not possible to apply vertex cover algorithm on them. The study of the "coloring problem" was quickly stopped, as it was clearly not adapted to the problem we want to solve: to avoid having attack scenarios which use a human attack just after another one. And the study of the connectivity property of a

graph was put on hold until a reasonable solution in terms of calculation time has been found.

Of the four main objectives of this thesis, three were fulfilled thorough this thesis: the design of a realistic model of attacker, the design of a general methodology for the assessment of the security, and the implementing of the models in the form of a demonstration tool. Sadly, we could not validate the proposed models and algorithms on an existing infrastructure. The best we had is a realistic example.

There is still a lot of things to do on the subject. In the continuity of what we have done, it remains:

1. to find non systematic algorithm for the attack path in a more dynamic model of infrastructure,
2. to implement algorithms for the vertex cover,
3. to automate the graph generation, since construction by hand tends to be tedious, error-prone, and impractical for attack graphs larger than a hundred nodes.

More generally, the algebraic vision needs to be further developed. There are still many mathematical structures which can be studied to see how they can help evaluate an infrastructure security. Mathematical structures like the hypergraph or structures linked to the percolation theory for example.
In the same time, the intelligence part must not be forgotten (improvement of the audit methodology). The study of the connectivity property of a graph shows us that looking for more or less obvious links between connected subgraphs of the infrastructure allows to have richer results.

And finally, to get round the NP completeness of many of the attack patterns studied to evaluate the infrastructure security, we will take interest into the isomorphism of graphs. The idea is to have a database of anonymized representing graphs and to compare the graph of the targeted infrastructure with the graphs in the database to have a first idea of its vulnerabilities more quickly.

Appendix A

Lists of critical sectors

The critical sectors of American, African, Asian, European and Pacific nation-states are presented in the following tables. For the sake of simplicity of presentation, some critical sectors were merged or renamed because they are very similar or one of them was a sub-sector of the others. For example, Railways and Aviation were merged in Transport as Electrical Power in Energy and Manufacturing in Industry.
The results are divided in several tables as they cannot be contained in only one table. First, you will find the critical sectors of the nation-states with a definition of critical infrastructure. They are presented in alphabetical order. Then, you will find the critical sectors of the nation-states without a definition of a critical infrastructure. They are presented in alphabetical order.

A.1 Nation-states and organizations which have a definition of critical infrastructure

When the information is not indicated, the list of critical sectors comes from the document where the definition is found.

In Figure A.1 (page 124), A.-P. T. stands for the Asia-Pacific Telecommunity and E. U. stands for the European Union.
The list of critical sectors of Canada comes from the "National Strategy for Critical Infrastructure" [17].

In Figure A.2 (page 125), NATO stands for the North Atlantic Treaty Organization.
The list of critical sectors of NATO comes from "162 CDS 07 E rev 1 - The Protection of Critical Infrastructures" [6].

In Figure A.3 (page 126), U.K. stands for the United Kingdom and U.S.A.

Sectors	A.-P. T.	Australia	Austria	Belgium	Canada	E. U.
Energy	X	X	X	X	X	X
Transport	X	X	X	X	X	X
Information Technology	X		X		X	X
Communication Technology	X	X	X	X	X	X
Water	X	X			X	X
Food		X			X	X
Health		X			X	X
Emergency Services	X					X
Finance	X	X		X	X	X
Banking	X	X				
Public Order	X					X
Legal Order			X			X
Safety / Security					X	X
Defense						X
Administration						X
Government	X				X	
Chemical industry						X
Nuclear industry						X
Space						X
Research						X
Vital Goods			X			
Industry					X	
Oil and Gas	X					X
SCADA systems						X
Broadcasting						X
The internet						X

Figure A.1: Critical sectors of nation-states with definition (part 1)

stands for the United States of America.
The list of critical sectors of the U.S.A. comes from "Critical Infrastructures: Background, Policy, and Implementation" [90].
The list of critical sectors of Switzerland comes from "The CIP Report: The Swiss Programme on Critical Infrastructure Protection" [14].

Sectors	Germany	Hungary	Japan	NATO	Netherlands	New Zealand
Energy	X	X	X	X	X	X
Transport	X	X	X	X		X
Information Technology	X	X	X	X	X	
Communication Technology	X	X	X	X	X	X
Water	X	X	X	X	X	X
Food	X	X		X		X
Health	X	X	X	X	X	X
Emergency Services	X			X	X	X
Rescue Services	X				X	
Finance	X	X	X	X	X	X
Banking		X				X
Insurance			X		X	
Public Order						X
Legal Order	X			X		X
Safety / Security		X			X	X
Defense		X		X	X	
Administration	X		X			
Government			X	X		X
Space					X	
Vital Goods					X	
Disaster control and management	X					
Media	X					
Culture	X					
Industry		X				
Oil and Gas			X			X
The internet						X
Waste						X
Cyber infrastructure				X		
Logistics						X
Postal services		X				
Networks						X

Figure A.2: Critical sectors of nation-states with definition (part 2)

Sectors	Norway	Poland	Spain	U. K.	U.S.A.	Switzerland
Energy	X	X	X	X	X	X
Transport	X	X	X	X	X	X
Information Technology		X	X		X	X
Communication Technology	X	X	X	X	X	X
Water	X	X	X	X	X	X
Food	X	X	X	X	X	X
Health	X	X	X	X	X	X
Emergency Services	X			X	X	
Rescue Services	X	X				
Finance	X	X	X	X	X	X
Banking	X				X	
Public Order	X					
Legal Order	X					
Safety / Security						X
Defense	X				X	
Administration		X	X			X
Government	X			X	X	
Social welfare/social services/	X					
Chemical industry		X	X		X	
Nuclear industry		X	X		X	
Space			X			
Research			X			
Disaster control and management	X					
Industry					X	X
Oil and Gas	X					
Satellite-based infrastructures	X					
Environment	X					
Waste	X				X	X
Agriculture					X	
National monuments and icons					X	
Postal services					X	
Materials					X	

Figure A.3: Critical sectors of nation-states with definition (part 3)

A.2 Nation-states without a definition

Sectors	Argentina	Brazil	Chile	Czech Republic	Denmark	Estonia
Energy	X	X	X		X	X
Transport	X	X	X	X	X	X
Communication Technology		X			X	X
Water	X	X				
Food			X			X
Health		X				X
Emergency Services						X
Finance		X				X
Banking		X				
Insurance						X
Public Order						X
Safety / Security		X				
Defense	X	X			X	
Social welfare/social services						X
Chemical industry			X			
Media	X		X		X	
Agriculture			X			
Postal services					X	X
Transectoral			X			

Figure A.4: Critical sectors of nation-states without definition (part 1)

In Figure A.4 (page 127), the lists of critical sectors of Argentina, Chile, Czech Republic and Denmark come from "Protection of 'Critical Infrastructure' and the role of Investment Policies relating to National Security" [51].
The list of critical infrastructure of Brazil comes from "International CIIP Handbook 2008/2009" [16].
The list of critical sectors of Estonia comes from "Emergency Preparedness Act" [42].

In Figure A.5 (page 128), the list of critical sectors of Finland comes from The Finnish Critical Infrastructure Protection; State Crisis Management Model And Situation Awareness [102].
The list of critical sectors of France comes from "Arrêté du 2 juin 2006 fixant la liste des secteurs dactivités dimportance vitale et désignant les ministres coordonnateurs desdits secteurs" [58].
The lists of critical sectors of Greece and Iceland come from "Protection of 'Critical Infrastructure' and the role of Investment Policies relating to National Security" [51]. The list of critical sectors of India comes from "International CIIP Handbook 2008/2009" [16].
The list of critical sectors of Italy comes from Italian Association of Critical Infrastructures' Experts" [2].

Sectors	Finland	France	Greece	Iceland	India	Italy
Energy	X	X	X	X	X	X
Transport	X	X	X	X	X	X
Information Technology	X	X				X
Communication Technology	X	X		X	X	X
Water	X	X				X
Food	X	X	X	X		X
Health	X	X				X
Emergency Services						X
Finance	X	X	X	X	X	X
Banking			X	X	X	X
Insurance	X				X	
Public Order					X	
Legal Order		X			X	
Defense		X			X	X
Administration		X				
Government						X
Chemical industry	X					
Nuclear industry					X	
Space		X			X	
Research		X				
Media		X	X			
Industry	X	X				
Oil and Gas					X	
The internet						X
Waste	X					
Agriculture			X	X		
Construction	X					
Postal services				X		X
Transectoral				X		

Figure A.5: Critical sectors of nation-states without definition (part 2)

In Figure A.6 (page 129), the lists of critical sectors of Ireland, Latvia, Lithuania and Luxembourg come from Protection of 'Critical Infrastructure' and the role of Investment Policies relating to National Security [51].
The list of critical sectors of Kenya comes from The Critical Infrastructure Protection Bill, 2015 [94].
The list of critical sectors of Korea comes from International CIIP Handbook 2008/2009 [16].

In Figure A.7 (page 130), the list of critical sectors of Mauritius comes from National Cyber Security Strategy 2014-2019 [107].
The list of critical sectors of Malaysia comes from the CNII Portal website [129].
The lists of critical sectors of Mexico, Portugal and Romania come from Protection of 'Critical Infrastructure' and the role of Investment Policies relating

Sectors	Ireland	Kenya	Korea	Latvia	Lithuania	Luxembourg
Energy	X	X	X	X		X
Transport	X	X	X	X	X	X
Information Technology		X				
Communication Technology	X	X	X	X		X
Water	X					X
Food	X			X	X	
Health					X	
Emergency Services			X			
Finance	X		X	X		
Banking	X			X		
Safety / Security		X	X			
Defense			X	X		
Administration			X			
Government			X			
Nuclear industry			X			
Disaster control and management			X			
Media			X		X	X
Oil and Gas			X			
Broadcasting			X			
Agriculture	X			X	X	
Postal services	X			X		X

Figure A.6: Critical sectors of nation-states without definition (part 3)

to National Security [51].
The list of critical sectors of Russia comes from International CIIP Handbook 2008/2009 [16].

In Figure A.8 (page 131), the lists of critical sectors of Singapore and Sweden come from “International CIIP Handbook 2008/2009” [16].
The lists of critical sectors of Slovak Republic and Slovenia come from “Protection of ’Critical Infrastructure’ and the role of Investment Policies relating to National Security” [51].

Sectors	Mauritius	Malaysia	Mexico	Portugal	Romania	Russia
Energy	X	X	X			
Transport	X	X	X	X	X	
Information Technology	X	X				X
Communication Technology	X	X	X	X		X
Water	X	X	X	X	X	
Food		X	X			
Health	X	X				
Emergency Services		X				
Finance	X	X	X	X		X
Banking		X	X	X		
Legal Order						X
Safety / Security		X				
Defense		X	X	X	X	X
Government	X	X				
Chemical industry			X	X	X	
Disaster control and management						X
Media			X			
Industry	X					
Broadcasting	X					
Waste			X			
Logistics	X					
Agriculture		X	X			
Postal services			X	X		
Transectoral			X			
Domestic and Foreign Policy						X
Science and Technology						X

Figure A.7: Critical sectors of nation-states without definition (part 4)

Sectors	Singapore	Slovak Republic	Slovenia	Sweden
Energy	X		X	X
Transport	X	X	X	X
Information Technology	X			
Communication Technology	X		X	X
Water	X		X	X
Food	X			
Health	X			
Finance	X			X
Banking	X			
Safety / Security	X			
Defense			X	
Industry				X
SCADA systems				X
The internet				X
Postal services			X	
National command systems				X

Figure A.8: Critical sectors of nation-states without definition (part 5)

Appendix B

Formalisation et analyse algébrique et combinatoire de scénarios d'attaque généralisés - Résumé français

B.1 Introduction

En 2005, 1,5 à 2 millions d'habitants ont été privés d'électricité pendant plusieurs heures à Moscou et dans les régions avoisinantes à cause d'un incendie et d'une explosion dans une sous-station locale [5]. En 2007, une cyber attaque contre l'Estonie a eu pour conséquence l'interruption temporaire de l'activité de nombreuses de ses infrastructures critiques [145]. En 2010, Stuxnet fut responsable des dommages conséquents subis par le programme nucléaire iranien en ciblant ses systèmes SCADA (Supervisory Control And Data Acquisition) [9]. En 2012, le malware Flame fut utilisé pour espionner les pays du Moyen Orient [10]. En 2013, l'entreprise Target a subi une cyber attaque qui causa la plus grande brèche de sécurié reportée puisque plus de 40 millions de clients ont vu leurs données de cartes bancaires être volées, et 70 millions de clients ont vu leurs données personnelles telles que leur email et leur adresses être volées [70] [85]. En 2014, une cyber attaque menée par la Chine a ciblé les systèmes de santé publique et compromis les données personnelles (noms, dates de naissance, numéros de sécurité sociale et adresses) de 4,5 millions de patients [50]. La même année, l'attaque contre Sony Pictures a vidé plusieurs centres de données internes, faisant fuité contrats, liste des salariés, budgets de films, films, numéros de sécurité sociale et emails [88]. En 2015, une des plus im-

portantes companies d'assurance des États-Unis d'Amérique, Anthem, a admis que les informations personnelles de dizaine de millions de ses adhérents avaient été compromises à cause d'une brèche de sécurité dans la base de données. La même année, la Crimée a subi des attaques contre ses lignes électriques qui ont laissé les trois quarts de sa population sans électricité pendant plusieurs jours, voire plusieurs semaines dans certaines régions [81] [93]. Plus récemment, le rançongiciel WannaCry a infecté plus de 230 000 ordinateurs dans plus de 150 pays. Parmi les victimes figuraient le National Health Service (NHS), Telefónica, FedEx et Deutsche Bahn [67] [55]. Ces exemples d'attaques ne sont que quelques exemples des nombreuses autres attaques que les infrastructures ont subies ces dernières années.

Les infrastructures sont toujours vulnérables aujourd'hui, ce qui montre que les solutions de sécurité proposées ne sont pas toujours suffisantes. L'objectif principal de cette thèse est donc de trouver de nouvelles façons d'évaluer la sécurité des infrastructures. Naturellement, comprendre ce qu'est une infrastructure, quels en sont les composants, et comment ils interagissent les uns avec les autres a été la première tâche que nous avons effectuée. Cela nous a menés à l'étude des diverses et nombreuses définitions d'une infrastructure critique. Nous avons remarqué que les définitions actuelles d'une infrastructure critique sont inadaptées à la réalité des attaques observées ou potentielles. Il en est de même des attaques elles-mêmes puisque le terme "cyber attaque" tend à réduire considérablement le champ conceptuel et opérationnel de celui qui est en charge de la sécurité. La quasi-totalité des approches se réduit à identifier le champ strictement technique informatique (systèmes, réseaux) et à oublier d'autres dimensions propres au renseignement. Ainsi les principales méthodologies d'identification et de gestion du risque (EBIOS [33] ou méthodologies similaires) considèrent une définition particulièrement restrictive, statique et locale de la notion d'infrastructure critique. La modélisation elle-même des attaquants et des attaques est extrêmement réduite. La principale erreur est de restreindre les approches techniques et les angles d'attaque d'un attaquant au seul champ informatique. Les angles dits cyber peuvent ne pas exister ou représenter un volet limité dans un scénario global d'attaque. En outre, l'approche classique néglige le volet opérationnel gouvernant la préparation et la conduite de la manoeuvre dans une attaque.

Il est alors nécessaire de concevoir une définition très élargie d'une infrastructure critique, laquelle doit être dictée par la vision de l'attaquant et non celle du défenseur. Cette thèse vise à développer de nouveaux modèles d'infrastructure basés sur la théorie des graphes et à modéliser de manière très élargie le concept d'attaque, incluant ou non un champ cyber. Cette représentation, déjà utilisée pour décrire la topologie des infrastructures critiques, sera enrichie pour appréhender de manière exhaustive l'environnement avec lequel elles interagissent. Les interdépendances avec d'autres entités (personnes, autres infrastructures critiques, etc.) sont un élément clef dans la construction de scénarios d'attaques sophistiqués. Cette représentation enrichie doit aboutir à de nou-

veaux modèles d'attaquants, plus réalistes et mettant en oeuvre des composants externes de l'infrastructure mais appartenant à son environnement proche. L'objectif majeur est la recherche de chemins optimaux dans un scénario d'attaque défini par l'objectif de l'adversaire. Cette approche globale apporte une définition plus fine (et donc plus réaliste) de la sécurité comme étant le coût le plus faible du chemin d'attaque pris sur l'ensemble des adversaires réalistes.

Ainsi, les objectifs principaux de cette thèse sont :

1. la conception d'un modèle enrichi de représentation d'une infrastructure à partir de la théorie des graphes,
2. la conception d'un modèle d'adversaire réaliste,
3. l'implémentation des modèles précédents sous la forme d'un démonstrateur de recherche,
4. la validation du modèle proposé par l'évaluation d'une infrastructure existante.

Le programme de recherche est structuré en cinq étapes. Les deux premières étapes visent à définir les modèles et les objets représentant les infrastructures ainsi que les attaquants auxquels elles sont confrontées. La difficulté majeure rencontrée dans l'élaboration d'un modèle d'infrastructure pertinent est sa capacité de description. En effet, plus le modèle sera riche et plus il pourra décrire l'infrastructure et les adversaires qui l'attaquent. La contrepartie de la richesse attendue d'un modèle est son caractère exponentiel. Dans ces modèles de sécurité, nous nous attendons donc à la réduction du problème de recherche des vulnérabilités d'une infrastructure de sécurité à des problèmes difficiles, soit NP-hard voire NP-complet. Les verrous à lever consisteront donc en la conception d'heuristiques pour répondre à ces problèmes en temps fini avec une réponse "acceptable".
La troisième étape consiste en la définition d'une méthodologie générique pour évaluer la sécurité d'une infrastructure. Cette étape doit aboutir à la conception d'heuristiques de recherche de vulnérabilités. Cette étape n'est pas discutée ici car elle a été exécutée par d'autres personnes.
Afin de valider les modèles et la méthodologie proposés, le programme de thèse prévoit le développement d'un démonstrateur de recherche sous la forme d'une plate-forme d'évaluation. Enfin, la dernière étape consistera à l'évaluation d'un système existant en mettant en oeuvre la méthodologie proposée. L'objectif de cette dernière étape est de valider les modèles et la méthodologie et d'en proposer une amélioration si nécessaire.

Cette thèse fait partie d'un projet industriel appelé InfraSec qui vise à aider les infrastructures à faire face aux menaces. Concrètement, InfraSec est un outil d'audit de sécurité conçu pour permettre aux entreprises de mesurer leur exposition aux risques et d'anticiper les attaques en identifiant des structures

d'attaque. InfraSec est la combinaison d'une méthodologie d'audit et d'un outil de modélisation et d'analyse. La méthodologie recueille des renseignements sur l'écosystème de l'infrastructure auditée. Ceux-ci sont ensuite injectés dans l'outil éponyme pour modéliser l'infrastructure et tous ses composants (humains, techniques, organisationnels, etc.). Le résultat est une cartographie claire et pertinente qui, combinée à des calculs algorithmiques complexes, permet d'identifier les composants vitaux de l'infrastructure ainsi que les structures d'attaque les plus efficaces (en termes de difficultés, de coûts et de temps).
Le projet InfraSec a été bien accueilli par la communauté, lors de forums internationaux (FIC 2015) ou devant des panels d'experts (présentation à la Direction Générale de l'Armement MI, un centre d'expertise technique de l'armée française).

Dans ce résumé, le chapitre 1 présente les grandes lignes de notre étude sur les définitions d'une infrastructure critique, étude réalisée afin de comprendre ce qu'est réellement une infrastructure et ce à quoi elle fait face. A la fin de ce chapitre nous faisons la proposition d'une nouvelle définition de cette notion. Le chapitre 2 présente le modèle d'infrastructure que nous avons retenu pour le projet InfraSec. Et enfin, le chapitre 3 présente les structures d'attaque, c'est-à-dire les structures mathématiques qui peuvent être utilisées pour construire des scénarios d'attaque dont les caractéristiques permettent d'évaluer la sécurité d'une infrastructure.

B.2 Qu'est-ce qu'une infrastructure ?

Avant de pouvoir modéliser une infrastructure, il est important de savoir ce qu'est une infrastructure. Pour cela, nous nous sommes intéressés aux définitions officielles d'une infrastructure critique.

Il existe de nombreuses définitions de la notion d'infrastructure critique et leur nombre ne fait que croître avec les ans. De plus, ces définitions ont connu de nombreuses modifications et en connaîtront certainement d'autres car de plus en plus d'acteurs s'intéressent à cette notion. Même certaines entreprises ont leur propre définition d'une infrastructure critique [41].

Vingt-cinq définitions ont été étudiées, dont deux de pays africains, une de pays arabes, cinq de pays et d'organisations d'Asie et du Pacifique, cinq de pays et d'organisations américains et douze de pays et d'organisations européens. Les définitions retenues sont les plus récentes qu'il a été possible de trouver. Malgré tout, du fait de la grande quantité de documents sur les infrastructures critiques (en particulier sur la protection de celles-ci) qu'il est possible de trouver, il est assez difficile de s'assurer que les définitions retenues sont réellement les plus récentes.

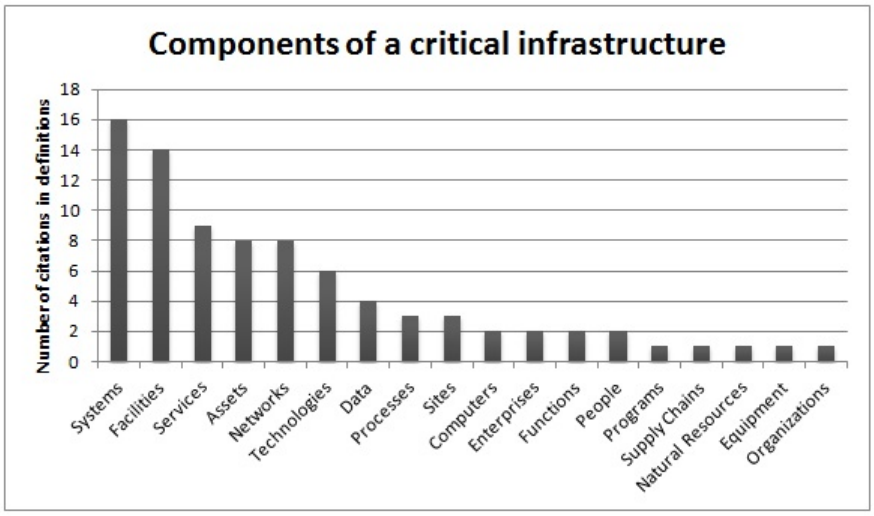

Figure B.1: Histogramme des composants identifiés dans les définitions

B.2.1 Les éléments identifiés dans les définitions

Les définitions d'une infrastructure critique comprennent généralement une liste de ses composants et les conséquences de sa perturbation, de son endommagement ou de sa destruction. La liste des composants est la partie qui diffère le plus d'un pays à un autre.
De nombreux composants essentiels à une infrastructure ont été identifiées dans les différentes définitions étudiées, dont les biens, les systèmes ou les réseaux (voir figure B.1, page 137). Nous avons remarqué que cette liste de composants tend à se réduire avec le temps. Par exemple, le livre vert sur un programme européen de protection des infrastructures critiques [115] donne une définition plus complète en 2005 que celle de la directive 2008/114/CE du conseil européen [116].

B.2.2 Les éléments non identifiés dans les définitions

Parmi tous les composants cités dans les différentes définitions étudiées, l'absence des composants humains est celle qui se remarque le plus. Presque aucune des définitions ne mentionne les humains comme faisant partie d'une infrastructure critique, bien que les humains soient essentiels au fonctionnement de toute infrastructure existante, critique ou non. Nous définissons les composants humains comme le personnel et les facteurs humains définis selon le Clinical Human Factors Group (CHFG) comme les facteurs environnementaux, organisationnels et professionnels, ainsi que les caractéristiques individuelles qui influencent le comportement au travail [140].

Le Royaume-Uni et l'Afrique du Sud sont les seuls pays qui incluent clairement le composant humain en tant que composant d'une infrastructure critique dans leur définition (voir figure B.1, page 137). Notons que la définition du Royaume-Uni ne l'incluait pas jusqu'à récemment selon une étude précédente publiée en 2014 [45] [48].

Certains pourraient objecter qu'un système, composant cité dans de nombreuses définitions, pourrait être défini comme étant composé de personnes, de processus et de technologies. Toutefois, les composants humains ne sont alors pas clairement énoncés et, par conséquent, les définitions peuvent induire en erreur les responsables en charge de la sécurité quant à l'importance de ces composants.

Il convient également de noter qu'aucune définition ne prend en compte les interdépendances avec des composants externes, ce qui n'offre qu'une vision très étriquée d'une infrastructure, alors considérée seulement comme une structure totalement isolée. En effet, même si certaines définitions mentionnent le concept d'interdépendance, comme celles du Canada, de la Hongrie et de la Pologne, ces interdépendances ne sont que celles inhérentes à l'infrastructure critique elle-même ou celles avec d'autres infrastructures critiques, mais jamais celles avec des infrastructures de base, tels que les sous-traitants, les fournisseurs, les centres de données ou autres.
Il serait également préférable de prendre en compte l'environnement des infrastructures critiques, notamment les environnements politique et culturel. Les attaquants pourraient utiliser ces environnements pour nuire à l'infrastructure, comme par exemple déclencher une grève, ce qui pourrait perturber le transport des ressources ou des produits nécessaires.

Ces omissions ont déjà été précédemment mentionnées [146].

B.2.3 Quelles peuvent être les conséquences de ces omissions ?

Comme indiqué précédemment, le composant humain est absent de la quasi-totalité des définitions étudiées, alors que les humains sont essentiels au fonctionnement des infrastructures critiques. Par ailleurs, Mitnick et Simon considèrent les humains comme le maillon faible de la sécurité [86] et leurs travaux démontrent bien qu'en dépit de l'utilisation des meilleurs éléments de protection de sécurité possibles, il est possible pour un attaquant d'accéder à des informations ou à des composants critiques simplement en utilisant des techniques d'ingénierie sociale.
Par exemple, Mitnick et Simon montrent comment un attaquant, ou un manipulateur dans ce cas, peut obtenir un nom d'utilisateur et le mot de passe correspondant en les demandant simplement à son propriétaire après avoir prétendu

faire partie du service informatique de l'entreprise. Et avec ces informations, le manipulateur dispose de tout ce dont il a besoin pour pénétrer dans le réseau de l'entreprise et localiser les éléments qu'il recherche.
Un événement récent illustre parfaitement ce cas. En 2016, un hacker a réussi à pénétrer dans les serveurs du FBI et a pu accéder à un téraoctet de données, dont il a extrait les coordonnées de près de 20 000 employés du FBI et 9 000 employés de la sécurité interne [30]. Pour réussir cela, il n'a pas eu à utiliser ses compétences informatiques, mais a plutôt abusé de la confiance de certains employés du gouvernement américain. Les affaires Snowden [61] et Wikileaks [147] montrent également bien en quoi les humains peuvent représenter un défaut majeur pour la sécurité des infrastructures. Dans ces cas cependant, peu de mesures peuvent être prises car il est difficile d'empêcher les employés de donner des renseignements confidentiels de leur plein gré, contrairement aux cas présentés par Mitnick et Simon qui peuvent être évités.

Les composants externes et l'environnement de l'infrastructure sont également absents des définitions étudiées. Pour démontrer l'erreur que représente cette omission, Filiol et Raynal ont planifié une attaque qui utilise ces composants pour retarder le départ d'un navire militaire [49]. Plutôt que d'attaquer directement le navire et ses systèmes informatisés, ils ont préféré cibler et utiliser les composants manquants dans les définitions : l'environnement politique et social a été utilisé pour déclencher une grève parmi les employés d'un fournisseur ainsi qu'une émeute qui a arrêté la livraison de pièces d'hélicoptère, tandis que le composant humain a été exploité en incriminant faussement le capitaine de l'unité militaire qui devait embarquer sur le navire.

Pour ces exemples, ce sont les éléments manquants des définitions qui ont permis aux attaques de réussir. Le danger est que, si ces composants n'apparaissent pas dans les définitions officielles, ils ne soient pas pris en compte dans les politiques de sécurité, et qu'ils ne soient pas protégés comme ils devraient l'être. Surtout que les attaquants tendent à utiliser le maillon le plus faible pour atteindre leurs objectifs.

B.2.4 Discussion autour d'une nouvelle définition d'une infrastructure critique

Au vu des conclusions de l'étude, les définitions officielles d'une infrastructure critique apparaissent restrictives, statiques et locales car elles sont dictées principalement par le point de vue du défenseur. Ainsi, afin d'avoir une définition plus complète et réaliste, la définition suivante a été dictée par le point de vue de l'attaquant.

Les infrastructures critiques peuvent être une entreprise, une institution, une organisation, des installations, des services et des équipements, qu'ils soient régionaux, nationaux ou internationaux, qui, s'ils étaient perturbés, endom-

magés ou détruits, auraient de graves répercussions sur la santé, la sûreté, la sécurité ou le bien-être économique des citoyens ou sur le bon fonctionnement des gouvernements et d'autres infrastructures qui en dépendent.
Elles comprennent tout composant qui pourrait avoir un impact grave sur la santé, la sûreté, la sécurité ou le bien-être d'une population (y compris les employés) ou qui pourrait entraîner la perturbation, l'endommagement ou la destruction de l'infrastructure critique et avoir une incidence grave sur son bon fonctionnement.

Plus précisément, une infrastructure critique comprend les personnes qui, si elles sont cooptées, corrompues ou éliminées, pourraient perturber, endommager ou détruire l'infrastructure critique.
Elle comprend également (liste non exhaustive):

1. les installations (accès, bâtiments, sites),
2. les équipements (ordinateurs, imprimantes, disques durs),
3. les structures,
4. les propriétés,
5. les fonds,
6. les ressources physiques ou naturelles,
7. le matériel,
8. les réseaux, qu'ils soient physiques (comme l'électricité ou l'eau) ou virtuels (comme l'Intranet ou Internet),
9. les informations/données, qu'elles soient physiques ou virtuelles (données confidentielles, telles que des mots de passe ou codes d'accès, procédures, organigrammes, contrats),
10. les technologies de l'information et de la communication,
11. les services,
12. les processus,
13. l'image de marque,
14. les systèmes ou une partie de ceux-ci,
15. d'autres infrastructures avec lesquelles il existe de fortes dépendances (fournisseurs de services ou de produits, par exemple).

Ces composants se retrouvent également dans l'environnement politique et culturel de l'infrastructure.

Certains composants d'une infrastructure peuvent être particulièrement critiques dû à leur importance. Nous les appelons les composants critiques de l'infrastructure.

Malheureusement, tous les composants de l'infrastructure qui peuvent entraîner sa perturbation, son endommagement ou sa destruction ne peuvent pas être identifiés et énumérés avec certitude. Cette tâche est impossible puisque la sécurité doit prendre en compte la capacité des attaquants à être novateurs, créatifs et, par essence, imprévisibles. Ils peuvent transformer un composant considéré comme inoffensif en composant critique. Cela peut expliquer pourquoi certaines définitions, comme celles de la Suisse [112] et de la Hollande [35], n'abordent pas vraiment les composants d'une infrastructure critique.

Une première version de cette définition a été publiée en 2014 [45].

Maintenant que les infrastructures ont été définies le plus clairement possible à travers l'étude et la proposition d'une nouvelle définition d'une infrastructure critique, il est temps de voir comment nous pouvons les modéliser.

B.3 Modèle d'infrastructure

Les menaces qui pèsent sur les infrastructures sont de plus en plus nombreuses et peuvent prendre des formes très diverses. Dans ce contexte, il est important de représenter les données qui affectent la sécurité d'une infrastructure. Or, il est impossible d'analyser cette masse d'information sans un outil de représentation. Et pour créer cet outil, un modèle d'infrastructure doit être défini.

Les graphes sont couramment utilisés pour modéliser les infrastructures et répondent à une liste impressionnante de problématiques : extraction de l'information [73] [65], décodage des codes de parité de faible densité [62], modélisation des réseaux de régulation génétique, recherche de gènes et diagnostic de maladies [78], test d'une application, support à la gestion des systèmes d'information hospitaliers [152], comprendre les mécanismes par lesquels les pannes, les idées et les maladies se propagent à l'intérieur de réseaux [149] [92] [82] [8], ou modélisation de la structure topologique des réseaux et étude des problèmes allant du routage à la réservation de ressources [153] pour n'en citer que quelques uns.

Il existe de nombreux champs d'application pour les modèles basés sur les graphes et par conséquent, il existe de nombreux modèles différents. C'est également vrai pour le domaine d'application qu'est la sécurité des infrastructures. Afin de donner un bref aperçu des modèles de sécurité existants, nous les divisons en deux catégories: les modèles qui modélisent les attaques contre une infrastructure et les modèles qui modélisent les infrastructures.

Les premiers sont principalement représentés par les arbres d'attaque, introduits par Bruce Schneier en 1999 [135].

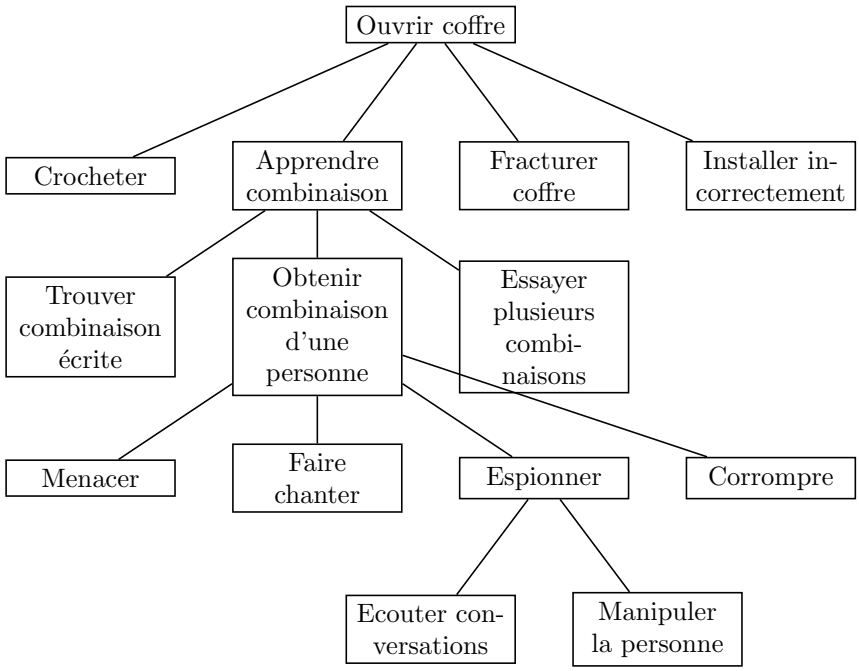

Figure B.2: Un exemple d'arbre d'attaque (Bruce Schneier)

Les deuxièmes n'ont pas de nom spécifique pour les désigner, probablement parce que, contrairement aux précédents modèles, ceux-ci ne se limitent pas au domaine de la sécurité.

Dans un modèle qui représente une infrastructure, une infrastructure est représentée par un graphe dont les noeuds modélisent les composants de l'infrastructure et dont les arcs modélisent les relations entre ces composants. Le modèle utilisé par l'outil Maltego en est un exemple [126]. La figure B.3 (page 143) montre l'exemple précédent (figure B.2, page 142), qui illustrait les arbres d'attaque, modélisé avec cette autre approche.

Malgré le fait que les modèles représentant les attaques aient de nombreux avantages : ils peuvent être réutilisés, que ce soit dans un autre arbre d'attaque ou pour une autre infrastructure, et ils décrivent précisément les attaques auxquelles font face les infrastructures ; nous leur avons préféré les modèles représentant les infrastructures. Les raisons en sont les suivantes. En premier lieu, l'objectif de cette thèse a toujours été de décrire le plus précisément possible

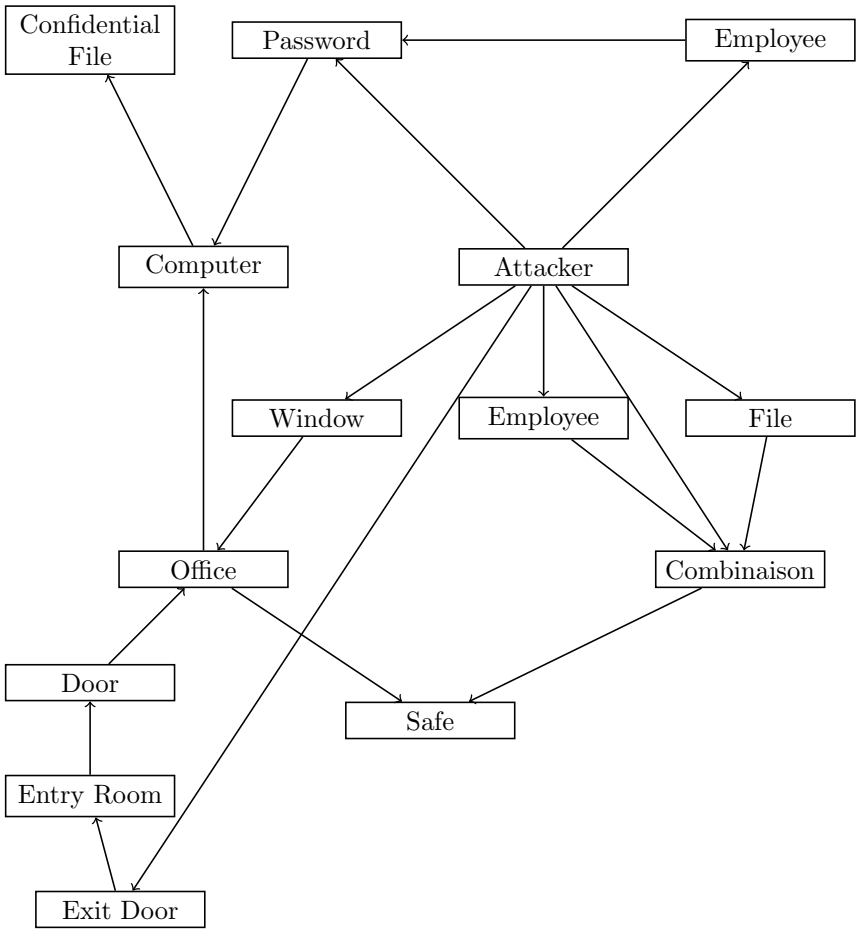

Figure B.3: Représentation d'une infrastructure simple

une infrastructure afin d'en évaluer la sécurité. C'est pourquoi les définitions des infrastructures critiques ont été étudiées. Deuxièmement, ces modèles sont plus polyvalents. Pour évaluer la sécurité des infrastructures, les modèles qui représentent les attaques sont contraints par le fait de ne pouvoir effectuer que des recherches sur les chemins les plus courts ou simplement des recherches sur les chemins. Alors que les modèles qui représentent les infrastructures possèdent plus de possibilités puisqu'ils permettent l'étude d'autres structures mathématiques. Enfin, les modèles à base d'arbres ont été rejetés car ils peuvent être considérés comme une forme particulière de graphes, une forme qui a été jugée trop pauvre pour représenter avec précision une infrastructure. En effet, cette dernière peut avoir plusieurs composants critiques à protéger, ce qui ne peut être décrit dans un modèle basé sur les arbres en raison de l'existence d'une

racine unique. Par conséquent, le modèle InfraSec présenté dans ce chapitre est un modèle basé sur les graphes qui décrit l'infrastructure ciblée.

B.3.1 Comment un noeud modélise un composant d'une infrastructure ?

Dans le modèle proposé, un noeud représente un composant de l'infrastructure modélisée, que ce soit un employé, un bâtiment, un réseau, une matière première ou une machine de production. Il est caractérisé par :

1. un libellé,
2. une catégorie afin d'identifier les noeuds partageant des caractéristiques communes,
3. un ensemble de propriétés communes à tous les noeuds :
   (a) Caché (oui / non) permet de retirer le noeud du graphe sans le supprimer complètement (principe de faux positif). Cela permet de garder en mémoire ce noeud sans qu'il ne soit pris en compte dans le graphe.
   (b) Entrée (oui / non) permet de définir si ce noeud peut servir de point départ pour une attaque.
   (c) Cible (Prioritaire / Secondaire / Non) permet d'identifier si le noeud est une cible pour l'attaquant ainsi que l'importance qu'il peut avoir pour l'attaquant.
4. un ensemble de propriétés spécifiques à la catégorie du noeud (voir les chapitres ci-dessous).
5. un ensemble de notes permettant d'évaluer la vulnérabilité du noeud au regard des compétences de l'attaquant :
   (a) vulnérabilité humaine (0 à 10) : vulnérabilité du noeud aux attaques visant les humains,
   (b) vulnérabilité physique (0 à 10) : vulnérabilité du noeud aux attaques de type crochetage,
   (c) vulnérabilité informatique (0 à 10) : vulnérabilité du noeud aux attaques informatiques.
6. un ensemble de notes permettant d'évaluer les compétences d'attaques qu'un attaquant pourrait exploiter s'il compromet ce noeud :
   (a) compétence humaine (0 à 10) capacité à réaliser une attaque sur un humain,
   (b) compétence physique (0 à 10) capacité à réaliser une attaque de type crochetage,

   (c) compétence informatique (0 à 10) capacité à réaliser une attaque informatique.

7. un texte ouvert permettant à l'utilisateur de décrire librement le noeud.

Par ailleurs, toutes les propriétés du noeud sont utilisées afin de définir les notes de vulnérabilités et de compétences des noeuds.

La catégorie représente l'appartenance d'un noeud à un ensemble de noeuds partageant des caractéristiques communes.
Afin de faciliter la mise en oeuvre d'un modèle de données précis et cohérent, il est admis qu'une catégorie peut appartenir à une autre catégorie (notion de sous catégorie).

La liste des catégories reconnues pour le moment est la suivante :

1. l'attaquant,

2. la société, qui représente la structure de plus haut niveau modélisé dans InfraSec. Tous les composants appartiennent à une société, que cette société soit la cible ou non,

3. les composants logiques, qui représentent tous les éléments permettant de décrire logiquement la société. Un composant logique doit obligatoirement appartenir à une société.

   (a) l'organisation, qui représente les entités organisationnelles, qu'elles soient internes ou externes (société, service, prestataire, client, fournisseur, etc.),

   (b) l'implantation, qui représente la localisation et la répartition géographique des éléments (site géographique, bâtiment, pièce, etc.),

   (c) le réseau, qui représente les différents réseaux et sous-réseaux du système d'information.

4. les composants réels, qui représentent tous les composants réels et donc attaquables d'une société. Un composant réel doit obligatoirement appartenir à un composant logique.

   (a) l'information, qu'elle soit numérique, papier ou immatériel (l'information ne doit pas être un produit),

   (b) la personne, qui représente le personnel de l'infrastructure mais aussi toutes les personnes qui peuvent avoir un impact sur son écosystème,

   (c) la voie d'accès, qui représente l'entrée et la sortie d'une entité physique (porte, fenêtre, portail, etc.),

   (d) l'informatique, qui représente les équipements informatique, réseau et téléphonique,

(e) le matériel, qui représente le mobilier (armoire, coffre, etc.) ainsi que le matériel spécifique à la société (machine de fabrication, de pesage etc.) ou à la sécurité (clé, caméra, alarme, etc.),

(f) le produit, qui représente ce qui est vendu ou acheté par l'infrastructure, que ce soit une information, un savoir, une matière première ou un produit manufacturé.

De plus, les catégories de noeuds peuvent être détaillées en sous catégories sur deux échelons. La création de plusieurs échelons de catégorie permet d'affiner la description des noeuds via la mise en place de profils. Un profil correspond à un type de noeud plus précis que la catégorie : la sous-catégorie secrétaire donne une indication plus précise que la catégorie personne par exemple. Ainsi, le modèle peut proposer à l'auditeur de renseigner automatiquement les propriétés du noeud avec des valeurs calculées selon un auto-apprentissage réalisé en prenant en compte tous les noeuds de la même sous-catégorie.

Dans le graphe, seules certaines catégories de noeuds représentent les éléments de l'écosystème pris en compte par un attaquant pour mener une attaque. Les autres catégories sont indispensables pour comprendre et exploiter les différentes strates organisationnelles de l'infrastructure (services, ateliers, filiale, etc.). Cela permet, entre autres, de mieux comprendre les flux internes. Néanmoins si l'attaquant veut nuire à un service, il va véritablement cibler une personne, des données, voire du matériel de ce service, mais pas le service en lui-même qui représente une entité concrètement indéterminée.

Il est donc nécessaire qu'InfraSec manipule les noeuds et les arcs selon deux perspectives ou niveaux différents :

1. le niveau Analyse offre une vue incomplète du graphe, un sous-graphe, limitée aux noeuds et aux arcs facilitant la compréhension de l'écosystème,

2. le niveau Attaque offre une vue incomplète du graphe, un sous-graphe, limitée aux noeuds et aux arcs permettant la définition d'une attaque réaliste.

Les vues du niveau Analyse souhaitées sont :

1. affichage d'un sous-graphe dont la disposition est relative à l'organisation de l'infrastructure,

2. affichage d'un sous-graphe dont la disposition est relative à la répartition géographique de l'infrastructure,

3. affichage d'un sous-graphe dont la disposition est relative au système d'information de l'infrastructure,

4. affichage d'un sous-graphe selon une disposition définie par l'auditeur.

Les vues du niveau Attaque souhaitées sont :

1. affichage d'un sous-graphe contenant tous les noeuds et les arcs pouvant intervenir dans l'attaque d'un noeud précis,

2. affichage d'un sous-graphe contenant tous les noeuds et les arcs pouvant intervenir dans l'attaque d'un arc précis,

3. affichage d'un sous-graphe selon une disposition définie par l'auditeur.

B.3.2 Évaluation des vulnérabilités et des compétences

Un des objectifs de cette thèse est d'identifier les chemins d'attaque les plus efficients. Pour ce faire, le modèle retenu prend en compte la valeur des propriétés des arcs et des noeuds. Il est donc indispensable que ces valeurs soient définies selon des principes arrêtés afin de préserver la cohérence générale du graphe.

Les notes de vulnérabilité d'un noeud sont automatiquement calculées selon la valeur des propriétés du noeud. Pour ce faire, les principes suivants sont mis en oeuvre :

1. chaque vulnérabilité de noeud a par défaut une note de 5, la note est ensuite modifiée selon les propriétés du noeud,

2. chaque propriété d'un noeud est associée à trois coefficients numériques représentant respectivement l'impact de la propriété sur les vulnérabilités humaines, l'impact de la propriété sur les vulnérabilités informatiques et l'impact de la propriété sur les vulnérabilités physiques,

3. chaque propriété de noeud, ayant une influence sur au moins une note de vulnérabilité du noeud, est associée à une liste de choix de valeurs arrêtées,

4. chaque élément de la liste de choix, pour une propriété de noeud, est associé à une valeur numérique. Si la valeur est négative, alors la note de vulnérabilité est atténuée. La sécurité du noeud est donc plus forte.

Soit p_i la valeur de la propriété i et $coef_i$ le coefficient de la propriété i. La formule de calcul est la suivante :

$$\text{Vulnérabilité} = 5 + \sum_{i=0}^{n} p_i \times coef_i$$

Enfin, pour des raisons pratiques, l'auditeur peut lui-même fixer les notes de vulnérabilités. Dès lors, il doit se baser sur l'échelle de valeur présentée dans la figure B.4 afin de respecter la logique du logiciel.

Vulnérabilité du noeud		Compétence de l'attaquant ou du noeud exploité	
Note	Détails	Note	Détails
0	Il n'est pas crédible d'attaquer le noeud.		
1	Les vulnérabilités ne peuvent être exploitées que par un attaquant d'un niveau exceptionnel qui dispose de moyens illimités.	10	L'attaquant a les compétences pour réaliser n'importe quelle attaque et il dispose de moyens illimités.
2	Les vulnérabilités ne peuvent être exploitées que par un attaquant d'un niveau exceptionnel.	9	L'attaquant a les compétences pour réaliser n'importe quelle attaque mais il ne peut pas tout se permettre.
3	L'exploitation des vulnérabilités du noeud est complexe et nécessite des moyens importants.	8	L'attaquant est capable de mener des attaques complexes mais il ne peut pas tout faire.
4	L'exploitation des vulnérabilités du noeud est complexe mais ne nécessite pas des moyens importants.	7	L'attaquant est capable de mener des attaques complexes mais il ne peut pas tout faire et dispose de moyens financiers limités.
5	La protection a été construite autour du noeud par des professionnels de la sécurité.	6	L'attaquant a un niveau professionnel, il dispose de moyens financiers importants mais il lui manque encore de l'expérience pour mener des attaques complexes.
6	La protection mise en oeuvre correspond aux préconisations professionnelles minimales qui ont été adaptées au contexte du noeud.	5	L'attaquant a un niveau professionnel mais ne peut pas encore mener des attaques complexes par manque d'expérience et de moyens financiers.
7	La protection mise en oeuvre correspond aux préconisations professionnelles minimales mais n'est pas réellement adaptée au noeud.	4	Lattaquant commence à avoir les connaissances nécessaires pour exploiter les vulnérabilités de base mais il ne sait pas encore être discret.
8	La protection mise en oeuvre n'est pas suffisante.	3	L'attaquant est un débrouillard qui peut exploiter les vulnérabilités les plus simples.
9	Le noeud possède une protection de façade qui peut être contournée facilement.	2	L'attaquant n'a pas de connaissance lui permettant d'exploiter la vulnérabilité mais, avec un minimum de temps, il est capable de s'adapter.
10	Le noeud n'est pas du tout protégé.	1	L'attaquant n'a pas de connaissance mais il est capable d'exploiter des vulnérabilités évidentes.
		0	Une attaque n'est pas possible.

Figure B.4: Echelle de valeur des vulnérabilités et des compétences

Les compétences d'attaque représentent les capacités d'une personne à porter une attaque humaine, informatique ou physique. Idéalement les compétences doivent être calculées automatiquement. Pour l'instant, l'auditeur définit lui-même ces valeurs en respectant l'échelle présentée dans la figure B.4.

B.3.3 Comment un noeud modélise un attaquant

Nous ne nous attendons pas à être en mesure de modéliser le comportement humain à ce stade, mais nous pensons que le modèle doit intégrer les capacités de l'attaquant, car cela peut avoir un impact significatif sur les décisions de sécurité.
La catégorie Attaquant ne contient pas de sous catégorie et n'est représentée que par un seul et unique noeud : l'attaquant.

Un attaquant est un individu ou un groupe d'individus avec des capacités diverses leur permettant d'atteindre leurs buts. Nous divisons ces compétences en trois catégories : les capacités physiques (métier de serrurier, force physique, techniques de vol, etc.), les capacités d'ingénierie sociale (prétextes, détournement, hameçonnage, etc.) et les capacités techniques (malware, hameçonnage, déni de service, etc.).
Ces capacités permettent à un attaquant d'atteindre une ou plusieurs cibles de l'infrastructure avec un coût et un temps d'exécution minimaux, une efficacité maximale et cela sans se faire prendre.

Le noeud Attaquant est unique dans le graphe. Il est lié obligatoirement à tous les noeuds dont la propriété "Point d'entrée" est fixée à Oui. Il permet de quantifier l'effort initial que doit fournir l'attaquant pour attaquer une infrastructure.

Le noeud Attaquant n'a pas de propriétés additionnelles. Le modèle ne prend en compte que les trois notes de compétences. Les caractéristiques du noeud Attaquant sont :

1. une note de compétence physique,

2. une note de compétence en ingénierie sociale,

3. une note de compétence informatique.

Par défaut, l'attaquant détient des compétences aux valeurs maximales, ainsi le logiciel considère que l'attaquant peut réaliser toutes les attaques. Néanmoins, il est possible de limiter le potentiel de l'attaquant afin d'affiner les recherches sur un profil d'attaquant préalablement défini.

B.3.4 Comment un arc modélise une relation de dépendance entre deux composants

Les composants d'une infrastructure dépendent plus ou moins les uns des autres. Ces dépendances, qu'elles soient matérielles, sociales, logistiques, environnementales ou logicielles, représentent des liens entre ces composants.

Un composant c_1 dépend d'un composant c_2 s'il est possible avec c_2 :

1. d'avoir un accès physique à c_1 si c_1 est un endroit,
2. d'obtenir, de modifier ou de supprimer c_1 si c_1 est un objet physique ou une information,
3. de corrompre, d'exploiter à l'insu ou de blesser c_1 si c_1 est une personne.

Les composants c_1 et c_2 sont interdépendants si c_1 dépend de c_2 et si c_2 dépend de c_1.

Un arc représente un lien de dépendance entre deux composants. Si un composant c_2 dépend d'un composant c_1, alors il y a un arc a du noeud représentant c_1 au noeud représentant c_2. L'arc est désigné par (c_1, c_2).
Si les composants c_1 et c_2 sont interdépendants, alors il y a un arc a_1 du noeud représentant c_1 au noeud représentant c_2 et un arc a_2 du noeud représentant c_2 au noeud représentant c_1 (voir figure B.5 page 150).

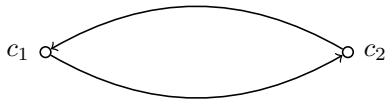

Figure B.5: Deux composants interdépendants

Un arc est la représentation d'une relation entre deux noeuds. Il est caractérisé par :

1. son noeud d'origine,
2. son noeud de destination,
3. son type.

Il ne peut pas y avoir deux arcs équivalents, c'est-à-dire ayant les mêmes noeuds d'origine et de destination ainsi que le même type.

Le nombre de types d'arc est relatif à la précision de la description de l'existant. Les types d'arcs sont regroupés en trois groupes :

1. les arcs d'impact représentent une relation logique entre deux noeuds impliquant une modification de la vulnérabilité du noeud de destination. Si l'attaquant accède à au noeud d'origine alors au moins une des propriétés du noeud de destination est modifiée.

2. les arcs hiérarchiques correspondent à une relation père/fils entre deux noeuds. Le noeud de destination appartient au noeud de d'origine. Aucune valeur n'est liée à ces arcs.

3. les arcs opérationnels représentent les vecteurs d'attaque possibles entre deux noeuds.

Les arcs hiérarchiques représentent une relation fils /père du noeud de destination au noeud d'origine. Le noeud de destination appartient (est le fils) au noeud d'origine.
Ces arcs permettent de structurer et d'organiser le modèle, et donc de mieux le comprendre. Mais ils ne sont pas pris en compte pour trouver des modèles d'attaque contre l'infrastructure modélisée.

Les arcs hiérarchiques sont divisés en 3 types :

1. les arcs d'appartenances représentent la hiérarchie fonctionnelle. Ex : ce service appartient à cette société ou cet équipement appartient à ce réseau.

2. les arcs de localisation représentent la position géographique. Ex : cet équipement ou cette personne est positionnée dans cette pièce,

3. les arcs de détention représentent la possession ou l'utilisation. Ex : cet équipement appartient ou est utilisé par cette personne.

Les arcs hiérarchiques ne sont pas associés à des propriétés.

Les arcs d'impact représentent une relation de cause à effet entre deux noeuds impliquant une modification de la vulnérabilité du noeud de destination. Si l'attaquant accède à au noeud d'origine alors au moins une des propriétés du noeud de destination est modifiée.

Il n'existe qu'un seul type d'arc dans ce groupe : l'arc d'impact.

Les propriétés de ce type d'arc ne sont pas encore clairement définies. L'idée est de décrire comment un noeud peut avoir un impact sur les valeurs des propriétés d'un autre noeud. Par exemple, dans le cas d'une clé et d'une porte, si l'agresseur a accès à la clé, la vulnérabilité de la porte devrait être maximale car la porte n'est plus protégée.

Notre meilleure piste est d'associer un pourcentage à un arc d'impact et de recalculer les valeurs de chaque arc opérationnel dont le point d'origine est le noeud de destination de l'impact d'arc selon les formules de calcul suivantes :

Effort = Effort - Effort × Pourcentage
Coût = Coût - Coût × Pourcentage
Temps = Temps - Temps × Pourcentage.

Les arc opérationnels représentent un lien entre deux composants d'une infrastructure qui peut être utilisé lors d'une attaque. Un arc opérationnel aura toujours des composants opérationnels à ses deux extrémités.

On distingue trois types d'attaque :

1. les attaques physiques,
2. les attaques informatiques,
3. les attaques liées à l'ingénierie sociale.

Les caractéristiques d'un arc opérationnel sont :

1. trois notes d'effort (0 à 10), une pour chaque type d'attaque,
2. trois coûts (en €), un pour chaque type d'attaque,
3. trois temps d'exécution (en minutes), un pour chaque type d'attaque.

Dans ce chapitre, nous avons donné un aperçu de la façon dont une infrastructure peut être modélisée pour évaluer sa sécurité. Nous avons distingué deux grandes catégories de modèles basés sur les graphes : les modèles qui représentent les attaques et les modèles qui décrivent l'infrastructure. Par la suite, nous avons expliqué pourquoi nous avons privilégié ces derniers.
Ensuite, nous présentons le modèle qui a été retenu pour cette thèse. Nous avons expliqué comment le modèle proposé représente une infrastructure : qu'est-ce qu'un modèle de noeud ? Qu'est-ce qu'un modèle d'arc ? Quelles valeurs, booléennes et continues, leur sont associées ? Etc.
Maintenant que le modèle retenu a été présenté, il est temps d'expliquer comment nous comptons l'utiliser pour évaluer la sécurité d'une infrastructure.

B.4 Structures d'attaque

Le modèle d'infrastructure retenu ayant été présenté dans le chapitre précédent, il est temps désormais de savoir comment ce modèle peut servir à évaluer la sécurité d'une infrastructure. Ce modèle sert en fait d'aide à la construction de scénarios d'attaque et ce sont les caractéristiques de ces scénarios d'attaque qui nous aident à évaluer la sécurité d'une infrastructure.

Pour construire ces scénarios d'attaque, nous recherchons des structures d'attaque dans les infrastructures modélisées. Nous appelons **structure d'attaque** toute structure mathématique liée à la théorie des graphes qui nous donne suffisamment d'informations pour évaluer la sécurité d'une infrastructure.

Dans ce chapitre, nous mettons l'accent sur les structures d'attaque qui ont donné des résultats prometteurs : le plus court chemin et la couverture des sommets. Le plus court chemin permet d'évaluer une infrastructure selon la philosophie suivante : plus l'attaque est peu chère, rapide et facile, moins l'infrastructure est sécurisée. La couverture des sommets permet d'évaluer la résilience de l'infrastructure en identifiant les composants qui peuvent mener à la paralysie de l'infrastructure.

B.4.1 Plus court chemin

La première structure d'attaque étudiée est le plus court chemin dans un graphe. L'objectif ici est la recherche de chemins optimaux et la question à laquelle nous voulons répondre est la suivante : Quel est le chemin le moins cher, le plus rapide et le plus facile entre l'attaquant et sa (ou ses) cible(s) ? Nous admettons alors qu'une infrastructure n'est pas sécurisée s'il est peu coûteux, rapide et facile de l'attaquer. Ceci permet d'avoir une approche réaliste de l'évaluation de la sécurité d'une infrastructure.

Pour rappel, une infrastructure est modélisée par un graphe dirigé pondéré $G = (V, A, c, t, e)$ dont les noeuds $v \in V$ représentent les composants de l'infrastructure ciblée et dont les arcs $a \in A$ représentent les liens de dépendance entre deux composants. À cela s'ajoutent une application de coût $c : A \to [0, \infty)$, une application de temps $t : A \to [0, \infty)$ et trois applications d'effort $eI : A \to [0, 10]$ avec $I = H, T, P$.
De plus, un attaquant est modélisé par un noeud et est considéré comme un composant de l'infrastructure. Afin de caractériser l'attaquant, trois types de capacité sont définis : une capacité physique, une capacité en ingénierie sociale et une capacité technique. Nous cherchons des chemins entre l'attaquant et le ou les composants critiques pour construire des scénarios d'attaque dont les caractéristiques (coût, temps et effort) seront utiles pour évaluer la sécurité de l'infrastructure. Ces chemins entre un attaquant et les composants critiques d'une infrastructure sont appelés chemins d'attaque.

Nous définissons un **chemin d'attaque** aP de longueur n comme un sextuplet (P, c_P, t_P, eP_P, eH_P, eT_P) où P est un chemin, soit une séquence finie d'arcs de la forme $(a_0 = (v_0, v_0, v_1), a_1 = (v_1, v_2), ...a_{n-1} = (v_{n-1}, v_n))$ tel que tous les arcs et les sommets soient distincts, c_P est le coût de P, t_P est le temps d'exécution de P, eP_P est l'effort physique de P, eH_P est l'effort humain de P et eT_P est l'effort technique. L'origine de P est l'attaquant et l'extrémité est un composant critique. Il doit y avoir un lien opérationnel entre le sommet $i-1$ et

le sommet i.

Le coût c_P est la somme des coûts de chaque arc du chemin P, nous avons donc :

$$c_P = \sum_{i=0}^{n-1} c_{a_i}.$$

Le temps d'exécution t_P est la somme des temps d'exécution de chaque arc du chemin P, nous avons donc:

$$t_P = \sum_{i=0}^{n-1} t_{a_i}.$$

L'effort physique eP_P est la valeur maximale des efforts physiques de chaque arc du chemin P, l'effort humain eH_P est la valeur maximale des efforts humains de chaque arc du chemin P et l'effort technique eT_P est la valeur maximale des efforts techniques de chaque arc du chemin P, nous avons donc :

$$\begin{aligned} eP_P &= \max_{0 \leq i \leq n-1} eP_{a_i}, \\ eH_P &= \max_{0 \leq i \leq n-1} eH_{a_i}, \\ eT_P &= \max_{0 \leq i \leq n-1} eT_{a_i}. \end{aligned}$$

Un chemin d'attaque nous permet d'identifier tous les composants et les liens opérationnels associés qui doivent être compromis pour qu'une attaque réussisse. Il donne également l'ordre des composants qui doivent être compromis pour atteindre un composant critique de l'infrastructure. Mais comment trouver des chemins d'attaque peut nous aider à évaluer une infrastructure exactement ?

La sécurité d'un objet est souvent définie comme la sécurité de son maillon le plus faible. Nous généralisons différemment cette approche aux infrastructures. Dans notre contexte, la sécurité d'une infrastructure n'est pas définie par la sécurité de son composant le plus faible mais par les valeurs des caractéristiques d'un chemin d'attaque.

Tout d'abord, nous définissons une **menace**, soit un sextuplet (cible, coût, délai, effort physique, effort humain, effort technique). L'infrastructure analysée est alors considérée comme vulnérable s'il existe un chemin d'attaque qui permet à l'attaquant d'atteindre la cible tout en respectant les limites de coût, de temps et d'effort définies par la menace.
Il est à noter que ces limites de coûts, de temps et d'efforts seront différentes d'une infrastructure à l'autre. Ces limites ne peuvent pas être les mêmes pour une start-up et pour une multinationale.

Plus précisément, la sécurité d'une infrastructure est considérée compromise si :

1. le coût du chemin d'attaque le moins coûteux est inférieur au coût de la menace,
2. le temps du chemin d'attaque le plus rapide est inférieur au temps de la menace,
3. l'effort physique du chemin d'attaque le plus facile est inférieur à l'effort physique de la menace,
4. l'effort humain du chemin d'attaque le plus facile est inférieur à l'effort humain de la menace,
5. l'effort technique du chemin d'attaque le plus facile est inférieur à l'effort technique de la menace,
6. les cinq conditions précédentes sont respectées,
7. la note du "plus court" chemin, calculée à partir d'une formule qui prend en compte les cinq caractéristiques, est inférieur à la note de la menace.

Pour le dernier élément, nous supposons qu'un algorithme systématique est utilisé pour les chemins entre l'attaquant et la cible. Disons que m chemins soient trouvés entre les deux composants. Pour déterminer le "plus court" chemin, la formule qui prend en compte les cinq caractéristiques est la suivante :

$$Grade_P = coef_e \times \frac{\frac{1}{3}\times(eP_P+eH_P+eT_P)}{\sum_{i=0}^{m}\frac{1}{3}\times(eP_{Pi}+eH_{Pi}+eT_{Pi})} + coef_c \times \frac{cP_P}{\sum_{i=0} cP_{Pi}} + coef_t \times \frac{tP_P}{\sum_{i=0} tP_{Pi}}$$

où les coefficients $coef_e$, $coef_c$ and $coef_t$ sont choisis par l'auditeur en fonction du type d'attaque qu'il veut privilégier (physique, humain ou technique). Notez que la somme de ces trois coefficients doit être égale à 1.
Par défaut, les valeurs des coefficients sont $\frac{1}{3}$ et nous supposons alors que l'attaquant est compétent dans les trois domaines.
Cette formule est partiellement basée sur la moyenne pondérée.

Une fois un chemin d'attaque calculé, il appartient à l'auditeur de construire un scénario d'attaque basé sur ce chemin d'attaque, les informations collectées, son expérience et son imagination.

Afin de trouver ces chemins d'attaque, nous avons tout d'abord opté pour une approche locale du problème en utilisant l'algorithme de Dijkstra. A cette époque, seul le coût d'un chemin d'attaque était pris en compte pour construire des scénarios d'attaque et évaluer la sécurité d'une infrastructure. Par conséquent, l'algorithme de Dijkstra a été utilisé pour trouver la valeur du chemin d'attaque le moins cher entre un attaquant et un composant critique.

Cet algorithme a été choisi parce qu'il trouve toujours le chemin optimal dans un temps polynomial. Mais il nous a fallu renoncer à cette approche lorsque

nous avons opté pour une approche plus dynamique du modèle d'infrastructure, un modèle qui prend en compte les liens d'impact entre un composant et sa ou ses clés (clé physique, mot de passe, combinaison, indices sur le mot de passe et la combinaison, etc.).
A ce stade, le graphe représentant une infrastructure peut être vu comme un ensemble de graphes interagissant ensemble, le graphe des composants et le graphe des clefs. A notre connaissance, aucun algorithme existant ne fonctionne sur une telle structure. Nous avons donc décidé de développer notre propre algorithme pour trouver les chemins d'attaque entre un attaquant et un composant critique.

L'algorithme proposé est récursif et se propage noeud par noeud. Il commence à partir d'un composant critique ciblé, ce qui permet naturellement de construire une arborescence et de ne considérer que les composants ayant un lien avec la cible de l'attaque. Des limites de coût, temps et efforts sont également fixées en fonction du profil d'attaquant que l'on souhaite avoir (aisé ou non, compétent ou non dans certains domaines, etc.). Ces limites permettent d'arrêter l'exploration de certaines branches de l'arborescence sur le chemin en cours si ses caractéristiques sont supérieures à celles fixées par les limites.

Pour chaque noeud, l'algorithme vérifie dans un premier temps si des clefs existent pour ce dernier et si celles-ci sont valables à utiliser. Puis il décide quel type d'attaque est le plus favorable pour passer au prochain noeud. En fonction de ces informations, l'algorithme met à jour le chemin en cours.

Nous avons testé l'algorithme sur un exemple fictif d'une centaine de noeuds, les calculs se font en moins de 10 secondes.

L'étude de cette structure d'attaque a été publiée en 2015 [46].

B.4.2 Couverture des sommets

L'étude de la couverture des sommets en tant que structure d'attaque potentielle a été motivée par les attaques récentes contre les lignes électriques en Crimée, qui ont laissé les trois quarts de sa population sans électricité pendant plusieurs jours - et jusqu'à trois semaines dans certaines zones. Ces attaques ont nécessité la destruction de quatre pylônes seulement pour laisser la plupart des 1,8 millions d'habitants de la péninsule sans électricité [81] [93].
De plus, le 25 mai 2005, entre 1,5 et 2 millions d'habitants ont été privés d'électricité pendant plusieurs heures à Moscou et dans les régions voisines en raison d'un incendie et d'une explosion dans une sous-station. La défaillance de cette sous-station a entraîné une panne d'électricité dans plusieurs zones grâce à un effet de cascade [66].
Afin d'éviter ou du moins de minimiser les effets de ce type d'attaque et de

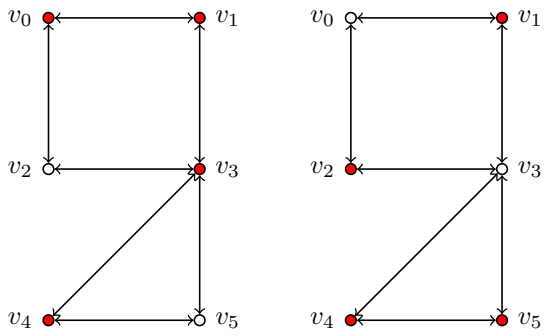

Figure B.6: Deux exemples de couverture des sommets

défaillance, les composants d'une infrastructure dont la perturbation, les dommages ou la destruction peuvent conduire à sa paralysie doivent être sécurisés. Mais avant tout, ils doivent être identifiés.

La piste qui est étudiée pour identifier les composants critiques d'une infrastructure est la couverture des sommets.

Soit G un graphe non orienté. G est défini par deux ensembles (V, A) où V est un ensemble de sommets et A est un ensemble d'arcs. Une **couverture des sommets** de G est un sous-ensemble de V, appelé V', tel que chaque arc $(v_1, v_2) \in A$ contient au moins un sommet de V'. Cela signifie que $\forall (v_i, v_j) \in A$, avec $i, j \in \mathbb{N}$, on a soit $v_i \in V'$, soit $v_j \in V'$, soit v_i et $v_j \in V'$. La figure B.6 (page 157) montre des exemples de couverture des sommets, l'ensemble V' de chaque graphe étant en rouge.
On dit que l'ensemble V' "couvre" tous les sommets de G.

Comme illustré dans la figure B.6 (page 157), le même graphe peut avoir plusieurs couvertures des sommets, par conséquent une couverture des sommets n'est pas unique et le problème de couverture des sommets peut avoir plusieurs solutions.

Une **couverture minimale des sommets** est une couverture des sommets de taille minimale. La figure B.7 (page 158) montre un exemple d'une couverture des sommets minimale pour le même graphe que la figure B.6 (page 157).

Le **problème de couverture des sommets minimale** est l'optimisation du problème de couverture des sommets. C'est un problème algorithmique qui consiste à trouver un ensemble de sommets de taille minimale pour couvrir tous les sommets d'un graphe donné. De plus, le problème de la couverture des sommets minimale peut avoir plusieurs solutions pour un même graphe.

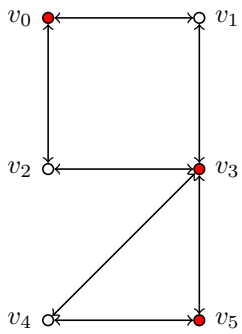

Figure B.7: Une couverture des sommets minimale

Soit $G = (V, A)$ un graphe, G représente une infrastructure. Si un attaquant, ou un groupe d'attaquants, corrompt, infiltre, prend les commandes ou vole tous les composants qui sont compris dans une couverture des sommets de G, alors il a un accès direct à tous les autres sommets de G.
Si un attaquant, ou un groupe d'attaquants, détruit, endommage ou arrête tous les composants qui sont compris dans une couverture des sommets de G, alors tous les autres composants de l'infrastructure finissent par être complètement isolés les uns des autres, comme le montre la figure B.8 (page 159).
Par conséquent, la résolution du problème de couverture des sommets dans un graphe représentant une infrastructure permet d'identifier ses composants critiques - ceux dont les noeuds correspondants sont dans une couverture des sommets - dont la perturbation, la corruption, l'endommagement, le vol ou la destruction conduisent à la paralysie de l'ensemble de l'infrastructure.
Ces composants peuvent ne pas sembler critiques lorsqu'ils sont considérés individuellement. En effet, un composant peut ne pas être critique seul si l'attaquant ne cible que ce composant, mais il peut être critique s'il est ciblé avec un ensemble de composants bien choisis.

L'identification des composants critiques n'est pas la seule chose que la couverture des sommets peut apporter d'un point de vue opérationnel. Une couverture de sommet d'un graphe n'est pas unique et il est alors possible d'identifier plusieurs ensembles de composants critiques et d'associer une équipe d'attaquants à chacun d'eux. Pour mener à bien leur mission, les différentes équipes n'ont pas à connaître l'existence des autres attaquants et de leurs cibles. De plus, même à l'intérieur de chaque équipe, les différents membres ne doivent pas nécessairement connaître l'existence des autres et de leur cible particulière (un composant dans la couverture des sommets peut n'être assigné qu'à un seul attaquant). Ces dispositions permettent de maximiser la probabilité de succès (redondance opérationnelle) tout en minimisant les risques opérationnels car

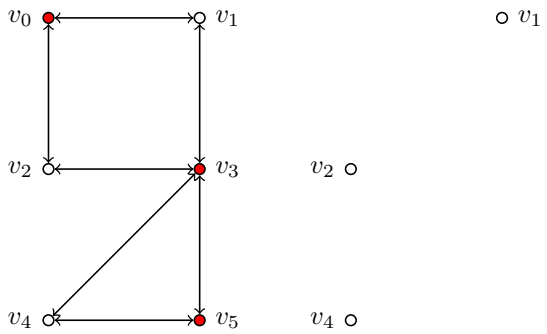

Figure B.8: Conséquence de la destruction de tous les composants dans une couverture des sommets

même si un ou plusieurs des attaquants, ou une équipe entière, sont pris, ils ne peuvent pas compromettre le reste de l'opération.

Comment la couverture des sommets peut-elle permettre d'évaluer la sécurité d'une l'infrastructure ? Cette évaluation est basée sur les données suivantes :

1. la taille des couvertures des sommets minimales du graphe G représentant l'infrastructure,
2. la faisabilité, le coût et le temps d'exécution des attaques contre les composants présents dans les couvertures des sommets.

On peut dire qu'une infrastructure est résiliente à une attaque basée sur l'approche de couverture des sommets si la somme des coûts et des temps, ainsi que la difficulté des chemins d'attaque ciblant tous les composants d'une couverture des sommets minimale sont trop importantes en fonction du coût, du temps et de la difficulté de la menace définie. Notez que plus la taille de la couverture minimale du sommet est grande, plus l'infrastructure est susceptible d'être résiliente.

Malheureusement, la définition de la couverture des sommets nécessite un graphe non dirigé alors que le modèle retenu utilise un graphe dirigé pour représenter une infrastructure. Par conséquent, comme il est hautement improbable que tous les composants d'une infrastructure soient interdépendants, il n'est pas possible, ou rarement possible au mieux, de chercher une couverture des sommets dans l'ensemble du graphe représentant l'infrastructure.
Une solution à cette impasse est de ne considérer qu'une partie du graphe, un sous-graphe dont les noeuds sont tous interdépendants.

Avec cette action, le graphe initial peut perdre quelques noeuds qui sont importants pour la pertinence de l'évaluation de sa sécurité, mais il pourrait être

tout de même intéressant d'essayer de résoudre le problème de la couverture des sommets minimale sur ce graphe réduit.

Il peut également être intéressant de considérer l'ensemble du graphe en supposant que tous les liens de dépendance existants entre deux composants sont bidirectionnels, même s'ils ne le sont pas en réalité. Certains des résultats peuvent être erronés, mais pourraient peut-être être exploités.
Seule l'expérience nous dira si ces possibilités peuvent être maintenues pour évaluer la sécurité de certaines infrastructures.
D'autre part, les réseaux comme le système électrique, Internet ou le "réseau humain" peuvent être facilement représentés par un graphe non dirigé.

Le problème de la couverture des sommets minimale est un problème NP-complet, l'un des 21 problèmes de Karp, ce qui signifie qu'il n'est pas toujours possible de trouver des solutions en un temps raisonnable car les algorithmes utilisés pour les résoudre sont exponentiels dans le pire des cas. Les algorithmes de base sont donc inutilisables en pratique.
Beaucoup de travaux ont été réalisés pour améliorer la complexité des algorithmes de couverture des sommets minimale. L'algorithme retenu est celui de Dharwadker, un algorithme de complexité polynomiale pour certaines catégories de graphe.

Pour résumer cette section, les récentes attaques en Crimée et à Moscou montrent l'importance de l'identification des composants critiques des infrastructures. Il a été montré comment la couverture des sommets, une structure liée à la théorie des graphes, permet cette identification et comment les scénarios d'attaque contre ces composants critiques permettent d'évaluer la sécurité de l'infrastructure ciblée.
Grâce aux scénarios d'attaque, les vulnérabilités de sécurité peuvent être déterminées et la résilience peut être évaluée ; et grâce aux caractéristiques de ces scénarios, il est possible de savoir si les vulnérabilités sont exploitables sur le plan opérationnel ou non, et donc si elles peuvent représenter un danger réel pour l'infrastructure.
La couverture des sommets permet également de maximiser la probabilité de succès (redondance opérationnelle) tout en minimisant les risques opérationnels en répartissant les composants critiques entre les différents attaquants sans qu'ils ne connaissent l'existence des autres cibles ainsi que des autres attaquants.

D'autres structures liées à la couverture des sommets ont été étudiées avec plus ou moins de succès. Seule la couverture de sommets "dégradée" pourrait avoir des résultats exploitables.

L'étude de cette structure d'attaque a été publiée une première fois en 2016 [47] et une deuxième fois en 2017 [54].

B.5 Conclusion

Partant du constat que les infrastructures sont toujours loin d'être sécurisées, l'objectif principal de cette thèse est de trouver de nouvelles façons d'évaluer leur sécurité. Naturellement, comprendre ce qu'est une infrastructure, quelles en sont les composants et comment ils interagissent les uns avec les autres a été la première tâche que nous avons effectuée. Cela nous a mené à l'étude des diverses et nombreuses définitions d'une infrastructure critique : plus d'une vingtaine de définitions ont été étudiées. Nous avons observé que la grande majorité de ces définitions ne mentionnent pas les composants humains, ni les composants externes et l'environnement proche, alors que les composants informatiques et technologiques sont parfois trop mis en valeur. Nous ne voulons pas minimiser les vulnérabilités que peuvent apporter les composants informatiques mais il est important de ne pas les mettre en avant au détriment d'autres composants tout aussi primordiaux. Nous avons observé le même problème avec les définitions d'une attaque, qui souffre de l'importance prise par le terme "cyber". Une fois ces observations faîtes, nous avons expliqué comment l'absence de ces composants dans les protocoles de sécurité peut avoir des conséquences désastreuses. Notez que nous ne sommes pas les premiers à soulever ce problème. Beaucoup ont déjà mentionné ces omissions. Mais comme elles ne sont toujours pas prises en compte dans les définitions officielles, nous avons pensé qu'il pourrait être utile de soulever à nouveau ce problème. Nous avons fini par proposer notre propre définition d'une infrastructure critique puisque les définitions officielles nous paraissent trop restrictives, statiques et locales. Nous avons essayé d'avoir la définition la plus exhaustive et la plus réaliste possible, mais cette définition reste discutable car il est certain que nous avons aussi oublié certains composants. Il y a tellement de types d'infrastructure différents qu'il est facile d'oublier certains éléments qui leur sont très spécifiques. Ce travail a été publié une première fois dans le cadre de la neuvième conférence internationale sur la guerre informatique et la sécurité (9^{th} *International Conference of Cyber Warfare and Security*) [45] et une deuxième fois dans le *Journal of Information Warfare* [48].

Une fois les définitions liées à la sécurité d'une infrastructure présentée, un état de l'art des modèles d'infrastructure existants a été réalisé. Nous avons distingué deux grandes catégories : les modèles qui décrivent les attaques et les modèles qui décrivent une infrastructure. Nous avons opté pour la dernière option car, contrairement à la première, elle n'exige pas que tous les composants critiques soient identifiés et nous permet donc de trouver des solutions pour l'identification des composants critiques. Ce qui peut être nécessaire car les composants critiques ne sont pas nécessairement les plus évidents.
En résumé, une infrastructure est modélisée par un graphe dirigé dont les noeuds représentent les composantes de l'infrastructure et dont les arcs représentent

les liens de dépendance entre deux composantes. Un attaquant fait partie du modèle d'infrastructure. Plusieurs types de noeuds et d'arcs sont définis. La distinction la plus importante est faite entre les éléments utilisés pour comprendre le fonctionnement de l'infrastructure et ceux utilisés pour évaluer sa sécurité. Ces derniers sont appelés éléments opérationnels.

Comme nous privilégions le point de vue de l'attaquant, nous évaluons la sécurité de l'infrastructure à travers la construction de scénarios d'attaque. Pour construire ces scénarios d'attaque, nous recherchons des structures d'attaque, soient des structure mathématiques liées à la théorie des graphes permettant d'évaluer la sécurité d'une infrastructure. Seuls les éléments opérationnels sont utilisés pour trouver des structures d'attaque. Plusieurs structures mathématiques ont été étudiées tout au long de cette thèse. Le chemin le plus court et la couverture des sommets sont ceux qui ont donné les résultats les plus prometteurs.

L'étude du problème du plus court chemin nous a mené à définir les chemins d'attaque. Un chemin d'attaque est un sextuplet (P, c_P, t_P, eP_P, eH_P, eT_P) où P est un chemin, c_P est le coût de P, t_P est le temps d'exécution de P, eP_P est l'effort physique de P, eH_P est l'effort humain de P, eT_P est l'effort technique de P. Nous cherchons des chemins entre l'attaquant et un composant critique de l'infrastructure ciblée. Nous utilisons les caractéristiques des chemins d'attaque identifiés pour évaluer la sécurité de l'infrastructure. Pour cela, nous définissons une menace comme un sextuplet (cible, coût, délai, effort physique, effort humain, effort technique). L'infrastructure analysée est alors considérée comme vulnérable s'il existe un chemin d'attaque qui permet à l'attaquant d'atteindre la cible tout en respectant les limites de coût, de temps et d'efforts.
Plusieurs algorithmes ont été implémentés et intégrés dans l'outil InfraSec afin de trouver des chemins d'attaque dans un graphe représentant une infrastructure. Pour avoir un modèle d'infrastructure plus dynamique (avec la gestion des clés), nous avons dû opter pour un algorithme systématique, ce qui peut être problématique dans le cas d'une infrastructure avec un très grand nombre de composantes.

L'étude de la couverture des sommets permet, en plus d'évaluer la sécurité d'une infrastructure, d'identifier des composants critiques qui n'étaient pas nécessairement évidents. Une fois ces composants identifiés comme des cibles, des chemins d'attaque sont recherchés entre l'attaquant et ces derniers. Des algorithmes pour trouver la couverture des sommets doivent encore être implémentés dans l'outil InfraSec.

Certaines des structures mathématiques étudiées n'ont pas donné de résultats prometteurs. La couverture des sommets du dual du graphe a été considérée comme décevante puisqu'il n'était bien souvent pas possible d'appliquer un algorithme de couverture verticale sur le dual du graphe. L'étude du problème de la coloration d'un graphe a été rapidement stoppée, elle n'était manifestement pas

adaptée au problème que nous voulons résoudre : éviter d'avoir des scénarios d'attaque qui utilisent une attaque humaine juste après une autre. Et l'étude de la propriété de connectivité d'un graphe a été suspendue jusqu'à ce qu'une solution raisonnable en termes de temps de calcul ait été trouvée.

Parmi les quatre objectifs principaux de cette thèse, trois d'entre eux ont été réalisés : la conception d'un modèle réaliste d'attaquant et d'infrastructure, la conception d'une méthodologie générale pour l'évaluation de la sécurité, et la mise en oeuvre des modèles sous forme d'un outil de démonstration. Malheureusement, nous n'avons pas pu valider les modèles et algorithmes proposés sur une infrastructure existante. Le meilleur que nous ayons eu est un exemple réaliste.

Il y a encore beaucoup de choses à faire sur le sujet. Dans la continuité de ce que nous avons fait, il reste à :

1. modifier le modèle existant afin qu'il soit possible d'utiliser des algorithmes existants et efficaces pour les recherches de structures d'attaque,
2. étudier de nouvelles structures d'attaque permettant d'évaluer la sécurité d'une infrastructure (hypergraphe ou structures liées à la théorie de la percolation par exemple),
3. automatiser l'évaluation des caractéristiques d'un noeud ou d'un arc.

Dans le même temps, il ne faut pas oublier la partie renseignement. Nous voyons avec l'étude de la propriété de connectivité d'un graphe que la recherche de liens plus ou moins évidents entre les sous-graphes connectés de l'infrastructure permettrait d'avoir des résultats plus riches.

Et enfin, pour contourner le problème du temps de calcul qui se pose pour de nombreuses structures d'attaque identifiées, nous nous intéresserons à l'isomorphisme des graphes. L'idée est d'avoir une base de données de graphes anonymisés et de comparer le graphe d'une infrastructure avec les graphes de la base de données pour avoir une première idée de ses vulnérabilités très rapidement.

Index

Bibliography

[1] Isabelle Abele-Wigert and Myriam Dunn. *International CIIP Handbook 2006 Vol.1, An inventory of 20 national and 6 international critical information infrastructure protection policies.* CSS ETH Zurich, 2006.

[2] AIIC. *Association of Critical Infrastructures' Experts*, 2011. http://www.infrastrutturecritiche.it/aiic-en/[Accessed: 2016-12-05].

[3] Edward Amoroso. *Fundamentals of Computer Security.* Upper Saddle River: Prentice Hall, 1994.

[4] E. Angel, R. Campigotto, and C. Laforest. *Algorithms for the Vertex Cover Problem on Large Graphs.* IBISC, 2010.

[5] Erin E. Arvedlund. *Blackout Disrupts Moscow After Fire in Old Power Station.* http://www.nytimes.com/2005/05/26/world/blackout-disrupts-moscow-after-fire-in-old-power-station.html?_r=0[Accessed: 2016-03-02].

[6] NATO Parliamentary Assembly. *162 CDS 07 E rev 1 - The Protection of Critical Infrastructures*, 2007. http://www.nato-pa.int/default.asp?CAT2=1159&CAT1=16&CAT0=2&COM=1165&MOD=0&SMD=0&SSMD=&STA=0&ID=0&PAR=0&LNG=0[Accessed: 2016-12-04].

[7] American Psychological Association. *The road to resilience.* http://www.apa.org/helpcenter/road-resilience.aspx[Accessed: 2017-03-02].

[8] A. Asztalos, S. Sreenivasan, B. K. Szymanski, and G. Korniss. Cascading failures in spatially-embedded random networks. *PLoS One 9, e84563*, 2014.

[9] Peter Beaumont. Stuxnet worm heralds new era of global cyberwar. *The guardian*, September 2010.

[10] Peter Beaumont. Cyberwar on iran more widespread than first thought, say researchers. *The guardian*, September 2012.

[11] Claude Berge. *Graphs and Hypergraphs.* North Holland Publishing Company, 1976.

[12] Rzadowe Centrum Bezpieczeństwa. *The National Critical Infrastructure Protection Programme*, 2015. http://rcb.gov.pl/wp-content/uploads/NPOIK-2015_eng-1.pdf[Accessed: 2016-12-05].

[13] Bosnia and Herzegovina. *Bosnia and Herzegovina Strategy for Prevention and Fight Against Terrorism*, 2010. http://www.msb.gov.ba/dokumenti/BiH%20Strategy%20for%20Prevention%20anf%20Fight%20against%20Terrorism.doc[Accessed: 2017-02-15].

[14] Stefan Brem. *The CIP Report: The Swiss Programme on Critical Infrastructure Protection*. The Center for Infrastructure Protection and Homeland Security, 2011.

[15] Lino Briguglio, Gordon Cordina, Nadia Farrugia, and Stephanie Vella. Economic vulnerability and resilience. 2008(22), May 2008. http://www.wider.unu.edu/publications/working-papers/research-papers/2008/en_GB/rp2008-55/_files/79432653132595540/default/rp2008-55.pdf[Accessed: 2017-03-05].

[16] Elgin M. Brunner and Manuel Suter. *International CIIP Handbook 2008/2009, An inventory of 25 national and 7 international critical information infrastructure protection policies*. CSS ETH Zurich, 2009.

[17] Canada. *National Strategy for Critical Infrastructure*, 2009. http://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/srtg-crtcl-nfrstrctr/srtg-crtcl-nfrstrctr-eng.pdf[Accessed: 2016-12-05].

[18] Yang Chang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3166–3173, 2013.

[19] J. Chen, I. A. Kanj, and G. Xia. Improved upper bounds for vertex cover. *Theoretical Computer Science, Volume 411*, 2010.

[20] William J. Clinton. *Executive Order EO 13010 Critical Infrastructure Protection*, 1996. https://fas.org/irp/offdocs/eo13010.htm[Accessed: 2017-02-20].

[21] U.S. Nuclear Regulatory Commission. *Fault Tree Handbook*. U.S. Government Printing Office, 1981.

[22] United States Congress. *USA Patriot Act*, 2001. https://www.gpo.gov/fdsys/pkg/BILLS-107hr3162enr/pdf/BILLS-107hr3162enr.pdf[Accessed: 2017-02-20].

[23] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms (1st edition)*. MIT Press and McGraw-Hill, 1990.

[24] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms (2nd edition)*. MIT Press and McGraw-Hill, 2001.

[25] Federal Financial Institutions Examination Council. *Glossary*. `http://ithandbook.ffiec.gov/glossary.aspx`[Accessed: 2017-02-12].

[26] National Infrastructure Advisory Council. *Critical Infrastructure Resilience Final Report and Recommandations*. Department of Homeland Security, 2009.

[27] National Research Council. *Terrorism and the Electric Power Delivery System*. Washington DC: The National Academies Press, 2012.

[28] The Information Security Policy Council. *Action Plan on Information Security Measures for Critical Infrastructures*. [online], http://www.nisc.go.jp/eng/pdf/actionplan_ci_eng.pdf, 2005.

[29] The Information Security Policy Council. *The Second Action Plan on Information Security Measures for Critical Infrastructures*, 2009. `http://www.nisc.go.jp/eng/pdf/actionplan_ci_eng_v2.pdfv`[Accessed: 2016-12-05].

[30] Joseph Cox. *Hacker Plans to Dump Alleged Details of 20,000 FBI, 9,000 DHS Employees*, February 2016. `https://motherboard.vice.com/en_us/article/9a3y4e/hacker-plans-to-dump-alleged-details-of-20000-fbi-9000-dhs-`employees [Accessed: 2017-09-15].

[31] Paolo Crucitti, Vito Latora, and Massimo Marchiori. Model for cascading failures in complex networks. *Phys. Rev. E*, 69:045104, April 2004.

[32] Agence Nationale de la Sécurité des Systèmes d'Information. *Information Systems Defence and Security - France's Strategy*, 2011. `http://www.ssi.gouv.fr/uploads/IMG/pdf/2011-02-15_Information_system_defence_and_security_-_France_s_strategy.pdf` [Accessed: 2017-02-05].

[33] Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI). *EBIOS Expression des Besoins et Identification des Objectifs de Sécurité*, 2010. `https://www.ssi.gouv.fr/guide/ebios-2010-expression-des-besoins-et-identification-des-obj`ectifs-de-securite/ [Accessed: 2017-08-04].

[34] Jefatura del Estado. *Ley 8/2011, de 28 de abril, por la que se establecen medidas para la protectión de las infraestructuras críticas*, 2011. `http://www.boe.es/boe/dias/2011/04/29/pdfs/BOE-A-2011-7630.pdf`[Accessed: 2016-12-05].

[35] The Hague Security Delta. *Securing Critical Infrastructures in the Netherlands*, 2015. https://www.thehaguesecuritydelta.com/media/com_hsd/report/53/document/Securing-Critical-Infrastructures-in-the-Netherlands.pdf[Accessed: 2016-12-13].

[36] Jochen W. Deuerlein. Decomposition model of a general water supply network graph. *Journal of Hydraulic Engineering*, 134(6):822–832, 2008.

[37] A. Dharwadker. *The Vertex Cover Algorithm*. 2011.

[38] Cambridge Dictionary. *English definition of resilience*. http://dictionary.cambridge.org/fr/dictionnaire/anglais/resilience[Accessed: 2017-03-02].

[39] Edsger W. Dijkstra. *A short introduction to the art of programming*. 1971.

[40] Myriam Dunn and Isabelle Wigert. *International CIIP Handbook 2004, An inventory and analysis of protection policies in fourteen countries*. CSS ETH Zurich, 2004.

[41] Elia. *Critical Infrastructure*. http://www.elia.be/en/safety-and-environment/Security/critical-infrastructure[Accessed: 2016-12-04].

[42] Estonia. *Emergency Preparedness Act*, 2000. http://www.ifrc.org/docs/idrl/233EN.pdf[Accessed: 2016-12-12].

[43] Austrian federal chancellor. *Anfrage*, 2006. http://www.parlament.gv.at/PAKT/VHG/XXII/J/J_04641/imfname_067709.pdf[Accessed: 2016-12-08].

[44] Eric Filiol. Comment paralyser un pays à l'aide du cyber ? *Les Cahiers de la Défense Nationale*, 2014.

[45] Eric Filiol and Cécilia Gallais. Critical infrastructure: Where do we stand today? In *Proceedings of the 9th International Conference on Cyber Warfare and Security*, ICCWS-2014, pages 47–57. ACPI, 2014.

[46] Eric Filiol and Cécilia Gallais. How can internal and external dependencies affect infrastructures security? In *Proceedings of the 14th European Conference on Cyber Warfare and Security*, ECCWS-2015, pages 129–138. ACPI, 2015.

[47] Eric Filiol and Cécilia Gallais. Combinatorial optimization of operational (cyber) attacks against large-scale critical infrastructures: The vertex cover approach. In *Proceedings of the 11th International Conference on Cyber Warfare and Security*, ICCWS-2016, pages 129–138. ACPI, 2016.

[48] Eric Filiol and Cécilia Gallais. Critical infrastructure: Where do we stand today? a comprehensive and comparative study. *Journal of Information Warfare*, 16:64–87, 2017.

[49] Eric Filiol and Frédéric Raynal. Cyberguere : de l'attaque du bunker à l'attaque dans la profondeur. *Revue Défense Nationale n.3*, March 2009.

[50] Jim Finkle and Caroline Humer. Community health says data stolen in cyber attack from china. *Reuters*, August 2014.

[51] Organisation for Economic Co-operation, Development, Kathryn Gordon, and Maeve Dion. *Protection of 'Critical Infrastructure' and the role of investment policies relating to national security*, 2008. `http://www.oecd.org/daf/inv/investment-policy/40700392.pdf`[Accessed: 2016-12-05].

[52] Centre for the Protection of National Infrastructure. *Critical National Infrastructure*. `http://www.cpni.gov.uk/about/cni/`[Accessed: 2016-12-05].

[53] German Federal Office for the Security of Information Technologies. *Critical Infrastructure Protection : Survey of World-Wide Activities*, 2004. `https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Kritis/paper_studie_en_pdf.pdf?__blob=publicationFile`[Accessed: 2016-12-12].

[54] Cécilia Gallais and Eric Filiol. Optimization of operational large-scale (cyber) attacks by a combinatorial approach. *International Journal of Cyber Warfare and Terrorism*, 7, 2017.

[55] Damien Gayle, Alexandra Topping, Ian Sample, Sarah Marsh, and Vikram Dodd. Nhs seeks to recover from global cyber-attack as security concerns resurface. *The guardian*, May 2017.

[56] Bénédicte Gouttebroze. Cyberattaques: 77 *Proceedings of the 9th International Conference on Cyber Warfare and Security*, 2016.

[57] Australian Government. *Critical Infrastructure Resilience Strategy*, 2010. `http://ccpic.mai.gov.ro/docs/Australian+Government+s+Critical+Infrastructure+Resilience+Strategy.pdf`[Accessed: 2016-12-05].

[58] French Government. *Arrêté du 2 juin 2006 fixant la liste des secteurs dactivités dimportance vitale et désignant les ministres coordonnateurs desdits secteurs*, 2006. `http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000423259&dateTexte=&categorieLien=id`[Accessed: 2016-12-05].

[59] New Zealand Government. *New Zealand's Cyber Security Strategy*, 2011. http://www.dpmc.govt.nz/sites/all/files/publications/nz-cyber-security-strategy-june-2011_0.pdf[Accessed: 2017-02-04].

[60] New Zealand Government. *New Zealand's Cyber Security Strategy*, 2015. http://www.dpmc.govt.nz/sites/all/files/publications/nz-cyber-security-strategy-december-2015.pdf[Accessed: 2017-02-15].

[61] Glenn Greenwald, Ewen MacAskill, and Laura Poitras. *Edward Snowden: the whistleblower behind the NSA surveillance revelations*, June 2013. https://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance[Accessed: 2017-09-15].

[62] Venkatesan Guruswami. Iterative decoding of low-density parity check codes (an introductery survey. 90, October 2006.

[63] Christopher Hadnagy. *Social Engineering: The Art of Human Hacking*. John Wiley & Sons, 2010.

[64] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *NIPS*, 2006.

[65] Hany Hassan, Ahmed Hassan, and Sarah Noeman. Graph based semi-supervised approach for information extraction. pages 9–16, June 2006.

[66] Kevin Hechtkopf. *Moscow Stricken By Power Outages*, 2005. http://www.cbsnews.com/news/moscow-stricken-by-power-outages/[Accessed: 2016-03-02].

[67] Alex Horn and Samuel Gibbs. What is wannacry ransomware and why is it attacking global computers? *The guardian*, May 2017.

[68] The White House. *Fact Sheet - Protecting America's Critical Infrastructures: PDD 63*, 1998. https://fas.org/irp/offdocs/pdd-63.htm[Accessed: 2017-09-15].

[69] T. Hughes and J. Guynn. *FBI investigating 11 attacks on San Francisco-area Internet lines*, 2015. http://www.usatoday.com/story/tech/2015/06/30/california-internet-outage/29521335/[Accessed: 2017-03-02].

[70] Associated Press in Minneapolis. Target says data breach possibly affected millions of credit cards. *The guardian*, December 2013.

[71] Spanish Cyber Security Institute. *National Cyber Security, a commitment for everybody*, 2012. http://www.ismsforum.es/ficheros/descargas/a-national-cyber-security-strategy-.pdf[Accessed: 2017-02-15].

[72] Service Public Fédéral Intrieur. *1er Juillet 2011 - Loi relative à la sécurit et la protection des infrastructures critiques. Moniteur Belge N.205, Vendredi 15 juillet 2011, Deuxime dition*, 2011.

[73] Ludovic Jean-Louis, Romaric Besanon, and Olivier Ferret. A graph-based information extraction method for template filling. 54:139–170, 10 2011.

[74] R. M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, 2009.

[75] Ryan Kinney, Paolo Crucitti, Réka Albert, and Vito Latora. Modeling cascading failures in the north american power grid. *The European Physical Journal B - Condensed Matter and Complex Systems*, (46 Issue 1):101–107, July 2005.

[76] P. Lacomme, C. Prins, and M. Sevaux. *Algorithmes de graphes*. Eyrolles, 2007.

[77] Latvia. *Cyber Security Strategy of Latvia*. `https://ccdcoe.org/sites/default/files/strategy/LVA_CSS_2014-2018.pdf`[Accessed: 2017-02-15].

[78] Jörg Linde, Sylvie Schulze, Sebastian G. Henkel, and Reinhard Guthke. *Data- and knowledge-based modeling of gene regulatory networks: an update*, 2015. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4817425/`[Accessed: 2017-08-31].

[79] Y. Lu, W. Zhang, H. Lu, and Xue X. Y. Salient object detection using concavity context. *ICCV*, 2011.

[80] Edouard Lucas. *Récréations mathématiques Volume 1*. 1882.

[81] N. MacFarqhar. *Crimea in Dark After Power Lines Are Blown Up*, 2015. `http://www.nytimes.com/2015/11/23/world/europe/power-lines-to-crimea-are-blown-up-cutting-off-electricity.html?_r=1`[Accessed: 2017-03-02].

[82] C. Magnien, M. Latapy, and J.-L. Guillaume. Impact of random failures and attacks on poisson and power-law random networks. *ACM Comput. Surv. 43*, 2011.

[83] Tim Maurer and Robert Morgus. *Compilation of Existing Cybersecurity and Information Security Related Definitions*, 2014. `http://giplatform.org/sites/default/files/Compilation%20of%20Existing%20Cybersecurity%20and%20Information%20Security%20Related%20Definition.pdf`[Accessed: 2017-02-06].

[84] Sjouke Mauw and Martijn Oostdijk. Foundations of attack trees. *Information Security and Cryptology - ICISC 2005*, pages 186–198, december 2005.

[85] Maggie McGrath. Target data breach spilled info on as many as 70 million customers. *Forbes*, January 2014.

[86] Kevin D. Mitnick and William L. Simon. *The art of deception*. Wiley & Sons, Hoboken, USA, 2003.

[87] Le Monde. *Tsahal annule une opération après une fuite sur Facebook*. http://www.lemonde.fr/proche-orient/article/2010/03/03/tsahal-annule-une-operation-apres-une-fuite-sur-facebook_1313918_3218.html[Accessed: 2017-01-08].

[88] Mariella Moon. Hackers leak sony passwords, social security numbers and salaries. *Engadget*, May 2014.

[89] John Moteff and Paul Parfomak. Crs report for congress, critical infrastructure and key assets : Definition and identification. *Congress Research Service, The Library of Congress*, october 2004.

[90] John D. Moteff. *Critical Infrastructures: Background, Policy, and Implementation*. Congressional Research Services, 2011.

[91] NATO. *AAP-06 Edition 2014 NATO Glossary of Terms and Definitions (English and French)* , 2014. http://wcnjk.wp.mil.pl/plik/file/N_20130808_AAP6EN.pdf[Accessed: 2017-02-04].

[92] M. E. Newman. Spread of epidemic disease on networks. *Phys. Rev. E 66, 016128*, 2002.

[93] BBC News. *Ukraine restores some electricity to Crimea after damage*, 2015. http://www.bbc.co.uk/news/world-europe-35039667[Accessed: 2017-03-02].

[94] Joseph Nkaissery. *The Critical Infrastructure Protection Bill*, 2015. http://www.icta.go.ke/downloads/critical-bill.pdf[Accessed: 2016-12-02].

[95] Defence Strategy Department of Belgium. *Cyber Security Strategy for Defence*, 2014. https://ccdcoe.org/sites/default/files/strategy/Belgian%20Defence%20Cyber%20Security%20Strategy.pdf[Accessed: 2017-02-12].

[96] University of Birmingham. *Information Security Glossary*, 2012. https://intranet.birmingham.ac.uk/it/documents/public/Information-Security-Glossary.pdf[Accessed: 2017-02-15].

[97] Government of Canada. *Public Safety Canada / Critical Infrastructure*. http://www.publicsafety.gc.ca/cnt/ntnl-scrt/crtcl-nfrstrctr/index-eng.aspx[Accessed: 2016-12-04].

[98] Government of Canada. *Canada's Cyber Security Strategy*, 2010. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/cbr-scrt-strtgy/cbr-scrt-strtgy-eng.pdf[Accessed: 2017-02-04].

[99] Republic of Colombia. *Policy Guidelines on Cybersecurity and Cyberdefense*, 2011. https://www.sites.oas.org/cyber/Documents/Colombia%20-%20National%20Cybersecurity%20and%20Cyberdefense%20Policy.pdf[Accessed: 2017-02-15].

[100] Ministry of Foreign Affairs of the Czech Republic. *Security Strategy of the Czech Republic*, 2015. http://www.army.cz/images/id_8001_9000/8503/15_02_Security_Strategy_2015.pdf[Accessed: 2017-02-06].

[101] Royal Norwegian Ministry of Government & others. *National Guidelines on Information Security 2007-2010*, 2007. https://www.oecd.org/norway/41671072.pdf[Accessed: 2016-12-05].

[102] Timo Härkönen Director of Government Security. *The Finnish Critical Infrastructure Protection; State Crisis Management Model and Situational Awareness, International Conference on Critical Infrastructure Protection - Towards Common Concepts and Cooperation in Disaster Reduction, October 4-5, Helsinki*, 2010. http://www.slideshare.net/Nostrad/ciphrknenppt[Accessed: 2016-12-12].

[103] U.S. Department of Homeland Security. *What Is Critical Infrastructure?* http://www.dhs.gov/what-critical-infrastructure[Accessed: 2016-12-04].

[104] U.S. Department of Homeland Security. *National Infrastructure Protection Plan.* [online], http://www.dhs.gov/xlibrary/assets/NIPP_Plan_noApps.pdf, 2006.

[105] U.S. Department of Homeland Security. *National Infrastructure Protection Plan, Partnering to enhance protection and resiliency.* [online], http://www.dhs.gov/xlibrary/assets/NIPP_Plan.pdf, 2009.

[106] Government of Jamaica. *National Cyber Security Strategy*, 2015. http://mstem.gov.jm/sites/default/files/Jamaica%20National%20Cyber%20Security%20Strategy.pdf[Accessed: 2017-02-15].

[107] The Republic of Mauritius. *The National Critical Infrastructure Protection Programme*, 2014. https://www.itu.int/en/ITU-D/Cybersecurity/Documents/National_Strategies_Repository/Mauritius_2014_National%20Cyber%20Security%20Strategy%20-%202014%20-%20EN.pdf[Accessed: 2016-12-02].

[108] House of Parliament. *Resilience of UK Infrastructure*, October 2010. http://www.parliament.uk/documents/post/postpn362-resilience-of-UK-infrastructure.pdf[Accessed: 2017-06-20].

[109] Republic of Poland. *Cyberspace Protection Policy of the Republic of Poland*, 2013. http://www.cert.gov.pl/download/3/162/PolitykaOchronyCyberprzestrzeniRP148x210wersjaang.pdf[Accessed: 2017-02-15].

[110] Republic of South Africa. *Draft Cybersecurity Policy of South Africa*, 2010. https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/cyber-security-policy-of-south-africa/at_download/file[Accessed: 2017-02-15].

[111] National Institute of Standards and Technology. *Glossary of Key Information Security Terms*, 2013. http://nvlpubs.nist.gov/nistpubs/ir/2013/NIST.IR.7298r2.pdf[Accessed: 2017-02-05].

[112] The Federal Council of Switzerland. *Nationale Strategie zum Schutz kritischer Infrastrukturen*, 2012. http://www.babs.admin.ch/content/babs-internet/en/aufgabenbabs/ski/_jcr_content/contentPar/tabs/items/downloads/tabPar/downloadlist/downloadItems/67_1460980334945.download/natstratski2012de.pdf[Accessed: 2016-12-12].

[113] Attorney-General's Department of the Australian Government. *Critical Infrastructure Resilience*. http://www.ag.gov.au/NationalSecurity/InfrastructureResilience/Pages/default.aspx[Accessed: 2016-12-05].

[114] Office of the Comissioner of Electronic Communications & Postal Regulation. *Cybersecurity Strategy of the Republic of Cyprus*, 2012. https://www.enisa.europa.eu/activities/Resilience-and-CIIP/national-cyber-security-strategies-ncsss/national-cyber-security-strategy-cyprus/at_download/file[Accessed: 2017-02-15].

[115] Commission of the European Communities. *Green Paper on a european programme for critical infrastructure protection*, 2005. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52005DC0576&from=EN[Accessed: 2016-12-09].

[116] The Council of the European Union. Council directive 2008/114/ce of 8 december 2008 on the identification and designation of european critical infrastructures and the assessment of the need to improve their protection. *Official Journal of the European Union*, December 2008.

[117] Federal Ministry of the Interior and Federal Republic of Germany. *National Strategy for Critical Infrastructure Protection (CIP Strategy)*. [online], http://www.bmi.bund.de/cae/servlet/contentblob/598732/publicationFile/34423/kritis_englisch.pdf, 2009.

[118] Federal Ministry of the Interior of Germany. *Cyber Security Strategy for Germany*, 2011. http://www.cio.bund.de/SharedDocs/Publikationen/DE/Strategische-Themen/css_engl_download.pdf;jsessionid=5B6636607CB58EFBB61431566F7E5B15.2_cid334?__blob=publicationFile[Accessed: 2017-02-05].

[119] Federal Ministry of the Interior of Germany. *National Strategy for Critical Infrastructure Protection (CIP Strategy)*, 2012. http://www.bmi.bund.de/cae/servlet/contentblob/598732/publicationFile/34423/kritis_englisch.pdf[Accessed: 2016-12-05].

[120] Government of the Netherlands. *Crisis, national security and terrorism: Protecting critical infrastructure.* http://www.government.nl/issues/crisis-national-security-and-terrorism/protecting-critical-infrastructure[Accessed: 2013-07-16].

[121] Federal Chancellery of the Republic of Austria. *Austrian Cyber Security Strategy*, 2013. https://www.bka.gv.at/DocView.axd?CobId=50999[Accessed: 2017-02-05].

[122] U.S. Department of Transportation. *Effects of catastrophic events on transportation system management and operations*, 2002. http://ntl.bts.gov/lib/jpodocs/repts_te/13754_files/13754.pdf[Accessed: 2017-01-02].

[123] Cabinet Office. *Sector Resilience Plan for Critical Infrastructure 2010*, March 2010. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/271335/sector-resilience-plan-2010.pdf[Accessed: 2017-06-20].

[124] SATRC Working Group on Policy and Regulations. *SATRC Report on Critical Information Infrastructure Protection and Cyber Security.* Asia-Pacific Telecommunity (APT), 2012.

[125] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Dover Publications Inc., 2000.

[126] Paterva. *Official Maltego Documentation*, 2017. https://docs.paterva.com/en/[Accessed: 2017-08-05].

[127] Department of the Prime Minister & Cabinet of New Zealand Patrick Helm. *Critical Infrastructure Resilience : Perspective from New Zealand*, 2008. http://idrc.info/fileadmin/user_upload/idrc/former_conferences/idrc2008/presentations2008/Helm_Patrick_Owen_Critical_Infrastructure_Protection_A_Perspective_from_New_Zealand.pdf[Accessed: 2016-12-05].

[128] Cynthia Phillips and Laura Painton Swiler. A graph-based system for network-vulnerability analysis. In *Proceedings of the 1998 Workshop on New Security Paradigms*, NSPW '98, pages 71–79, New York, NY, USA, 1998. ACM.

[129] CNII Portal. *About Critical National Information Infrastructure.* `http://cnii.cybersecurity.my/main/about.html`[Accessed: 2016-12-02].

[130] Quatar. *Quatar National Cyber Security Strategy*, 2014. `http://www.motc.gov.qa/sites/default/files/national_cyber_security_strategy.pdf`[Accessed: 2017-02-15].

[131] Reuters. Massive anthem health insurance hack exposes millions of customers' details. *The guardian*, February 2015.

[132] S. Rinaldi. Modeling and simulating critical infrastructures and their interdependencies. Hawaii.

[133] S. Rinaldi, J. Peerenboom, and T. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control System Magazine*, (21):11–25, 2001.

[134] Rosslin John Robles, Min-kyu Choi, Eun-suk Cho, Seok-soo Kim, Gilcheol Park, and Jang-Hee Lee. Common threats and vulnerabilities of critical infrastructures. *International Journal of Control and Automation.*

[135] Bruce Schneier. Attack trees: Modeling security threats. *Dr. Dobb's journal*, december 1999.

[136] Bruce Schneier. *Secrets & Lies: Digital Security in a Networked World.* Wiley, 2000.

[137] Oleg Sheyner, Joshua Haines, Somesh Jha, Richard Lippmann, and Jeannette M. Wing. *Automated Generation and Analysis of Attack Graphs. Proceedings of the 2002 IEEE Symposium on Security and Privacy*, 2002.

[138] H. Sims. *The Internet is down!!! Massive AT&T Outage Takes Humbolt County Phones, Internet, Etc Offline*, 2015. `http://lostcoastoutpost.com/2015/sep/3/internet-down-massive-t-outage-takes-humboldt-coun/`[Accessed: 2017-03-02].

[139] R. Smith. *Assault on California Power Station Raises Alarm on Potential for Terrorism. The Wall Street Journal*, 2014.

[140] Eric Solano. *Methods for Assessing Vulnerability of Critical Infrastructure.* `https://sites.duke.edu/ihss/files/2011/12/IHSS_Solano.pdf`[Accessed: 2017-03-02].

[141] Spain. *The National Security Strategy*, 2012. `http://www.lamoncloa.gob.es/documents/estrategiaseguridad_baja_julio.pdf`[Accessed: 2017-02-15].

[142] Jukka Suomela. *Survey of local algorithms. ACM Computing Surveys, volume 45, issue 2, article 24*, 2013.

[143] C.-W. Ten, C.-C. Liu, and M. Govindarasu. *Vulnerability Assessment of Cybersecurity for SCADA Systems Using Attack Trees. Proceedings of the IEEE Power Engineering Society General Meeting*, pages 24–28, 2007.

[144] TISN. *Critical infrastructure.* `http://www.tisn.gov.au/Pages/Critical_infrastructure.aspx`[Accessed: 2016-12-05].

[145] Ian Traynor. Russia accused of unleashing cyberwar to disable estonia. *The guardian*, May 2007.

[146] Daniel Ventre, Colonel Fran¸cois Chauvancy, Eric Filiol, François-Bernard Huyghe, and Joseph Henrotin. *Cyberwar and Information Warfare.* Wiley-ISTE, 2011.

[147] Paul Walker. *Bradley Manning trial: what we know from the leaked WikiLeaks documents*, July 2013. `https://www.theguardian.com/world/2013/jul/30/bradley-manning-wikileaks-revelations`[Accessed: 2017-09-15].

[148] Lingyu Wang, Tania Islam, Tao Long, Anoop Singhal, and Sushil Jajodia. *An Attack Graph-Based Probabilistic Security Metric. Proceedings of the 22nd Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, pages 283–296, 2008.

[149] Duncan J. Watts. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA 99*, 2002.

[150] Andreas Wenger, Jan Metzger, and Myriam Dunn. *International CIIP Handbook 2002, An inventory of protection policies in eight countries.* CSS ETH Zurich, 2002.

[151] Robin J. Wilson. *Introduction to Graph Theory, Third Edition.* Longman Inc., 1985.

[152] A. Winter and R. Haux. A three-level graph-based model for the management of hospital information systems. *Methods of Information in Medicine*, (34 Issue 4):378–396, Septembre 1995.

[153] Ellen W. Zegura, Kenneth L. Calvert, and Michael J. Donahoo. A quantitative comparison of graph-based models for internet topology. *IEE/ACM Transactions on Networking (TON)*, (5 Issue 6):770–783, December 1997.

Formalisation et modélisation algébriques et combinatoires des concepts d'attaque et d'infrastructure critique

RESUME : Les définitions actuelles des infrastructures de sécurité sont inadaptées à la réalité des attaques observées ou potentielles. Il en est de même des attaques elles-mêmes et en conséquence la quasi-totalité des approches se réduit à identifier le champ strictement technique informatique (systèmes, réseaux) et à oublier d'autres dimensions propres au renseignement. La modélisation elle-même des infrastructures, des attaquants et des attaques est extrêmement réduite.

Il est alors nécessaire de concevoir une définition très élargie, laquelle doit être dictée par la vision de l'attaquant et non celle du défenseur. Cette thèse vise à développer de nouveaux modèles d'infrastructure de sécurité basés sur la théorie des graphes et à modéliser de manière très élargie le concept d'attaque, incluant ou non un champ cyber. Cette représentation déjà utilisée pour décrire la topologie des infrastructures critiques sera enrichie pour appréhender de manière exhaustive l'environnement avec lesquelles elles interagissent. Les interdépendances avec d'autres entités sont un élément clef dans la construction de scenarii d'attaques sophistiquées. Cette représentation enrichie doit aboutir à des nouveaux modèles d'attaquants, plus réalistes et mettant en œuvre des composants externes de l'infrastructure mais appartenant à son environnement proche. L'objectif majeur est la recherche de chemins optimaux dans un scénario d'attaque défini par l'objectif de l'adversaire. Cette approche globale, apporte une définition plus fine (et donc plus réaliste) de la sécurité comme étant le coût le plus faible du chemin d'attaque pris sur l'ensemble des adversaires réalistes.

Mots clés : infrastructure, sécurité, attaque, graphe, modélisation

Formalization and algebraic and combinatorial analysis of generalized attack scenarios

ABSTRACT : The current definitions of a critical infrastructure are not adapted to the actual attacks which are observed these days. The problem is the same for the definition of an attack and therefore, most of the approaches are reduced to identify the technical and IT domain only, and they forget the others domains specific to the intelligence. The models of infrastructure, attacker and attack is also extremely narrowed.

Therefore, it is necessary to have a new definition of a critical infrastructure, more complete and made according to the attacker point of view. This thesis aims to develop new models of infrastructure and attack accurately, models which will based on graph theory, with or without the cyber part. This graph-based representation is already used a lot to describe infrastructure, it will be enriched in order to have a more exhaustive view of an infrastructure environment. The dependencies with other entities (people, others critical infrastructures, etc.) have to be taken into account in order to obtain pertinent attack scenarios. This enriched representation must lead to new models of attackers, more realistic and implementing external components of the infrastructure which belong to its immediate environment. The main objective is the research of optimal paths or other mathematical structures which can be translated into attack scenarios. This global approach provides a finer (and therefore more realistic) definition of security as the lowest cost of the attack path.

Keywords : infrastructure, security, attack, graph, model

ARTS ET MÉTIERS ParisTech

ParisTech INSTITUT DES SCIENCES ET TECHNOLOGIES PARIS INSTITUTE OF TECHNOLOGY