INSTITUT DES ACTUAIRES

Lyon 1

ISFA GRANDE ÉCOLE D'ACTUARIAT ET DE GESTION DES RISQUES
INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

Mémoire présenté devant l'ISFA pour l'admission à l'Institut des Actuaires

le 15 Avril 2020

Par : Armand BONNAC
Titre : Tarification du cyber-risque pour les collectivités locales : une modélisation inspirée par le modèle pandémique.

Confidentialité : ☐ Non ☑ Oui (Durée : ☐ 1 an ☑ 2 ans)

Les signataires s'engagent à respecter la confidentialité ci-dessus

Membres présents du jury de l'Institut des Actuaires :

Directeur des Ressources Humaines

A. COULOUMY
F. PICARD
B. BALTESAR

Membres présents du Jury du Master Actuariat de l'ISFA :

E. MASIELLO

Entreprise :
Nom : SMACL Assurances
Signature :

Directeur de Mémoire en entreprise :
Nom : Hervé Fraysse
Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)

Secrétariat :

Bibliothèque :

Signature du responsable entreprise

Signature du candidat

Résumé

Le cyber-risque est le risque numéro 1 des risques émergents. La fréquence et le coût moyen des cyber-attaques ne cessent de croître depuis une décennie. Plusieurs acteurs agissent afin d'accompagner les entreprises dans la protection de leurs données à caractère personnel. Parmi ces acteurs, les assureurs proposent un panel de garanties pour couvrir les pertes matérielles et immatérielles. Seule difficulté : comment déterminer un tarif spécifique à ce risque ? La quantité et la granularité des données actuellement disponibles ne permettent pas l'application de méthodes statistiques standards. Cet obstacle a incité des chercheurs à développer d'autres outils mathématiques spécifiques pour modéliser et tarifer le cyber-risque. Dans ce mémoire, nous allons utiliser une de ces méthodes qui consiste à modéliser le cyber-risque en s'inspirant de la modélisation pandémique. Cette nouvelle approche sera appliquée sur le portefeuille de SMACL Assurances pour tarifer un contrat d'assurance cyber.

Après avoir présenté l'environnement du cyber-risque et les particularités du portefeuille de SMACL Assurances, on définira la méthodologie de cette nouvelle approche. Le cyber-risque et le risque pandémique présentent en effet des similitudes :

- un réseau informatique peut s'assimiler à une population, où un ordinateur infecté peut alors contaminer l'ensemble du réseau comme n'importe quelle maladie infectieuse.
- les différents états liés à une cyber-attaques peuvent s'assimiler à ceux d'une maladie infectieuse, à savoir sain/en état de marche, infecté/hors service, guéri/réparé.

Parmi les modèles pandémiques, celui qui est couramment utilisé pour modéliser les réseaux informatiques est le modèle compartimental SIS. C'est un modèle simple avec seulement deux états : sain et infecté. Pour passer d'un état à un autre, des taux de passages sont ainsi définis : un taux d'infection pour le passage d'un état sain à infecté et un taux de guérison pour le passage d'un état infecté à sain. Dans le cadre du cyber-risque, le modèle SIS a été généralisé en intégrant un troisième taux de passage qui vient compléter le taux d'infection. Ainsi ces deux taux peuvent s'interpréter de la manière suivante :

- l'un correspond à une menace provenant de l'extérieur, comme par exemple un cyber-criminel qui souhaiterait introduire un virus dans un ordinateur ou bien une personne qui irait consulter un site internet malicieux ;
- l'autre correspond à une menace provenant de l'intérieur. Une fois le virus introduit dans le réseau, celui-ci irait se propager dans le réseau et contaminerait ainsi d'autres ordinateurs.

Le modèle prendra en compte également les pertes liées aux cyber-attaques. On en distinguera deux types :

- celles liées à l'infection elle-même, qui correspondrait à la perte d'information ;
- celles liées à la restauration de l'ordinateur infecté à un état sain, qui correspondrait à la perte de revenu ou à l'interruption de l'activité.

Une fois la méthodologie définie, on analysera les différents paramètres qui la composent, à savoir le type et la taille du réseau, les taux de passages puis les types de pertes.

Enfin une dernière partie sera consacrée à l'application numérique de cette méthode sur le portefeuille de SMACL Assurances. Pour cela, nous estimerons les différents paramètres : ceux liés à la composition

du réseau et aux taux de passages seront estimés à partir de données externes provenant d'études réalisées par des spécialistes dans le domaine de la cyber-criminalité. Pour les paramètres liés aux coûts, ils sont estimés dans un premier temps de manière à obtenir un tarif qui soit cohérent à la fois, par rapport aux contraintes techniques de SMACL Assurances et à ses contraintes commerciales. L'objectif est de proposer une structure tarifaire qui prendra en compte des critères qui ne faisaient pas partie de la tarification jusqu'à présent. Ainsi les premiers résultats obtenus de la modélisation seront utilisés pour différencier nos sociétaires et non pour calculer une prime pure. A l'avenir, les paramètres de ce modèle pourront s'ajuster en fonction des nouvelles informations que l'on obtiendra.

Mots-clés : *cyber-risque* ; *personnes morales de droit public* ; *modélisation pandémique* ; *modèle SIS* ; *tarification.*

Abstract

Cyber-risk is the number one risk of the emergent risks. The frequency and average cost of cyber-attacks kept growing over the last decade. Many actors are involved in the protection of personal and private data to support companies. Among them, insurers offer a large number of guaranties to cover material and immaterial losses. The only issue: how to determine the specific price to this risk? Data is not available in the required amount or in the desired granularity to apply standard statistical methods. This main issue encourages researchers to develop other specific mathematical tools to model and price the cyber-risk. In this report, we will use one of these methods that aims to modeling cyber-risk based on the pandemic model.

After presenting the cyber-risk environment and the particularities of the SMACL Assurances portfolio, we will define the methodology of this new approach. Cyber risk and pandemic risk present similarities:

- a computer network can be assimilated to a population, where a computer infected can then contaminate the entire network like any infectious disease.
- the various states related to a cyber attack can be assimilated to those of an infectious disease, such as healthy / in working order, infected / out of service, cured / repaired.

Among the pandemic models, the one commonly used to model computer networks is the SIS model. It is a simple model with only two states: healthy and infected. To pass from one state to another, two rates are defined: an infection rate and a cured rate. As part of cyber risk, the SIS model has been generalized by including a third rate to complete the infection rate. So these two infection rates could be interpreted as follows:

- one would be related to a threat from outside, such as a cyber-criminal who wants to introduce a virus into a computer or someone who goes to a malicious website;
- the other would be related to a threat from inside. Once the virus goes into the network, it would spread over the network and thus infects other computers.

The model will also take into account losses related to cyber attacks. There are two types of losses:

- those related to the infection itself, such as information loss;
- those related to the time spent to restore the computer infected, such as revenue loss or business disruption.

Once the methodology has been defined, we will analyze the various parameters, namely the type and size of the network, the rates and the types of losses.

Finally, a last part will be dedicated to the numerical application of this method on the SMACL Assurances portfolio. We will estimate the different parameters: those related to the network and the rates will be estimated from external data based on studies carried out by specialists in the field of cyber-crime. For the parameters related to the losses, they are estimated in order to get a price which will be consistent with the technical constraint of SMACL Assurances and his commercial constraint. Its aim is to propose a pricing framework that will take into account new criteria that was not included in the pricing before. Thus, the model will be used more to distinguish our customers based on new criteria

than to calculate a technical premium. In the future, the parameters of this model may be adjusted according to the new information available.

Keywords: *cyber-risk*; *public legal entity*; *pandemic model*; *SIS model*; *pricing*.

Remerciements

Je tiens à remercier Hervé Fraysse, responsable du service actuariat, pour ses conseils avisés et ses recommandations.

Je remercie également Monsieur Christian Robert, mon maître de stage, pour ses précieux conseils et toute l'attention qu'il a porté à mon mémoire.

Merci à Maochao Xu pour avoir pris le temps de répondre à mes questions sur l'un des articles qu'il a publié.

Enfin je désire remercier Marie-Élise Lorin, responsable du service gestion des risques et conformité, Olivier Daroux, responsable sécurité des systèmes d'information, Thomas Cambarrot, responsable adjoint du service actuariat, et toutes les personnes qui ont bien voulu m'éclairer sur certains aspects de mon mémoire.

Table des matières

Introduction

« *Une cyber-attaque perpétrée par un État ou un groupe extrémiste violent pourrait être aussi destructrice que l'attentat terroriste du 11 septembre 2001* » Léon E. Panetta, Secrétaire d'État Américain à la Défense (2006). Cette phrase, bien qu'alarmiste, révèle l'importance des cyber-attaques aujourd'hui. Leur nombre et leur coût n'ont cessé d'augmenter depuis une décennie et certaines d'entre-elles ont même impacté plusieurs pays à la fois avec des pertes se chiffrant en millions de dollars. Les cibles de ces attaques sont diverses, elles concernent toutes celles et ceux utilisant les technologies numériques, et ce quels que soient leurs secteurs d'activité. Il peut donc s'agir soit d'un particulier, soit d'un organisme privé ou public. Louis Gautier, secrétaire général de la défense et de la sécurité a introduit dans le rapport d'activité annuel de l'ANSSI (Agence Nationale de la Sécurité des Systèmes d'Information) que « *les technologies numériques procurent des gains de productivité et sont donc source de richesse et de compétitivité pour notre pays, mais elles induisent également des vulnérabilités nouvelles* ». Face à ce risque émergent, plusieurs organismes ont été créés au niveau national ou européen, afin d'aider et d'accompagner les entreprises à se protéger. De nouvelles lois ont été promulguées afin de pénaliser et/ou sanctionner celles dont le niveau de sécurité est insuffisant. Cette exigence de protection est due au caractère personnel des données détenues par ces entreprises, qui sont une source d'enrichissement frauduleux pour les cyber-criminels. La sensibilité de ces données varie selon leur domaine.

Un autre acteur accompagne les entreprises dans la protection de leurs données : il s'agit des assureurs qui proposent des garanties spécifiques pour couvrir les pertes matérielles et immatérielles dans le cadre d'une cyber-attaque. Seule difficulté : comment déterminer un tarif propre à ce risque ? Du fait du cycle de production inversé, les méthodes de tarification en assurance Non-Vie sont, pour la plupart, basées sur les données historiques, mais leur quantité et leur granularité ne permettent pas, actuellement, l'application de méthodes statistiques standard pour le cyber-risque.

Cet obstacle a incité des chercheurs à développer des méthodes mathématiques spécifiques pour modéliser et tarifer ce risque. Plusieurs mémoires d'actuariat traitent également de ce sujet. Parmi eux, principalement deux méthodes ont été utilisées : une qui s'appuie sur un modèle de scoring de risque par assuré [14] et une autre sur un modèle prédictif qui se base sur un jeu

de données* [16]. L'objectif de ce mémoire est d'utiliser une troisième méthode qui consiste à modéliser le cyber-risque en s'inspirant de la modélisation pandémique. Cette dernière était jusqu'à présent utilisée en assurance de personnes pour mesurer l'impact d'une pandémie sur une garantie santé, mais les similitudes avec le cyber-risque permettent d'appliquer cette méthode dans ce nouveau domaine et ainsi pallier le manque de données par le biais de simulations.

La première partie a pour objet la description de l'environnement du cyber-risque. Après l'avoir défini, on présentera les différents acteurs en jeu, leur rôle dans la protection des données à caractère personnel et les moyens de prévention mis en oeuvre. On terminera sur les actions à mettre en place via un processus ERM (*Entreprise Risk Managment*) lequel peut ainsi permettre aux entreprises de ne pas compromettre leur situation financière et leur réputation.

La deuxième partie est consacrée à SMACL Assurances qui souhaite définir une structure tarifaire pour la commercialisation d'un produit d'assurance Cyber. On présentera la mutuelle et les caractéristiques de son marché principal, les collectivités territoriales, son coeur de cible. On y abordera enfin les enjeux liés à la tarification du cyber-risque.

La troisième partie définit la méthodologie de cette nouvelle approche tarifaire qui s'inspire de la modélisation pandémique. Après avoir expliqué l'intérêt de cette méthode et les similitudes entre les deux risques, pandémique et cyber, on présentera les principaux modèles pandémiques. On terminera cette partie en définissant les paramètres et le processus d'une infection dans le cadre du cyber-risque.

Après avoir défini la méthodologie, on analysera dans une quatrième partie, les différents paramètres impactant la probabilité d'une infection obtenus dans la partie précédente. On analysera également sa sévérité calculée sur la base de simulation de type Monte Carlo.

Enfin la dernière partie est consacrée à l'application numérique de cette méthode sur le portefeuille de SMACL Assurances. D'une part, on estimera les paramètres du modèle puis, d'autre part, on définira une structure tarifaire propre aux spécificités de la mutuelle et à son coeur de cible, les collectivités. Pour finir on proposera une grille tarifaire selon différents critères.

∗. Le mémoire soutenu par Anaïs MARTINEZ sur la « Modélisation assurantielle du risque cyber » est confidentiel jusqu'au 03/07/2020.

Chapitre 1

L'environnement du cyber-risque

1.1 Définition du cyber-risque

Une cyber-attaque est une atteinte à des systèmes informatiques réalisée dans un but malveillant. Les cibles de ces attaques sont les équipements informatiques ; à savoir, les ordinateurs ou serveurs, mais aussi les périphériques tels que les imprimantes, les téléphones mobiles ou les tablettes.

Selon le Livre blanc du Courtier AON (septembre 2013), le cyber-risque est caractérisé par six « I » :

- Invisible : la cyber-attaque est soit impossible à détecter soit la détection est trop tardive. Les conséquences de l'attaque deviennent alors difficilement contrôlable ;
- Intensif et incompréhensible : il est difficile de déterminer l'origine de l'attaque ainsi que les dommages matériels et immatériels qu'elle peut engendrer. Seuls des spécialistes seront en mesure de détecter et d'évaluer les pertes ;
- International : l'attaque peut se produire en un lieu différent de celui depuis lequel elle est lancé. La cyber-criminalité ne se préoccupe pas des frontières. Plusieurs pays peuvent être impactés ;
- Incertain : une cyber-attaque peut atteindre n'importe quel organisme ayant un système d'information. C'est cette incertitude qui rend nécessaire l'assurance contre ce risque,
- Intentionnel : une cyber-attaque n'est pas un accident, c'est un acte criminel passible de la loi pénale.

La cause de ces attaques est due principalement à trois facteurs : la défaillance technique (panne), l'erreur humaine (négligence interne à l'entreprise) et la mauvaise sécurisation du système d'information.

1.1.1 Histoire du cyber-risque jusqu'à nos jours

L'histoire du cyber-risque débute avec l'apparition des premiers réseaux de télécommunication, au début des années 1980. Les attaques y étaient encore rares et parfois sans conséquences. C'est à partir des années 2000 que les attaques en ligne ont augmenté de façon exponentielle avec, par exemple, les attaques de Yahoo (2013) ou de Ebay (2014) qui causèrent des pertes astronomiques se chiffrant en milliards d'euros. Plus récemment encore, l'attaque connue sous le nom de WannaCry qui, en 2017, attaqua des milliers d'ordinateurs dans plus de 150 pays. Il s'agissait là du plus grand piratage informatique jamais enregistré dans l'histoire, touchant des entreprises privées comme Renault mais aussi le ministère de l'intérieur russe ou encore le centre hospitalier de Liège. Quelques jours plus tard, une autre cyber attaque d'une toute aussi grande ampleur, connue sous le nom de NoPetya, utilisa les mêmes failles de sécurité que son

prédécesseur WannaCry pour se propager. Elles sont passées d'un simple virus informatique à des attaques plus sophistiquées telles que les *ransomware*.

D'après l'étude de PwC (2016), la fréquence et la gravité des attaques ne cessent d'augmenter depuis ces deux dernières décennies. Il est encore très difficile d'évaluer son coût mais on l'estimerait à plusieurs centaines de milliards de dollars annuels ! L'institut Ponemon * a recensé les coûts liés à ces attaques aux États-Unis ; celles-ci s'élèvent, en moyenne, à 13 millions de dollars en 2018, soit une augmentation de 80% par rapport à 2013 (figure 1.1).

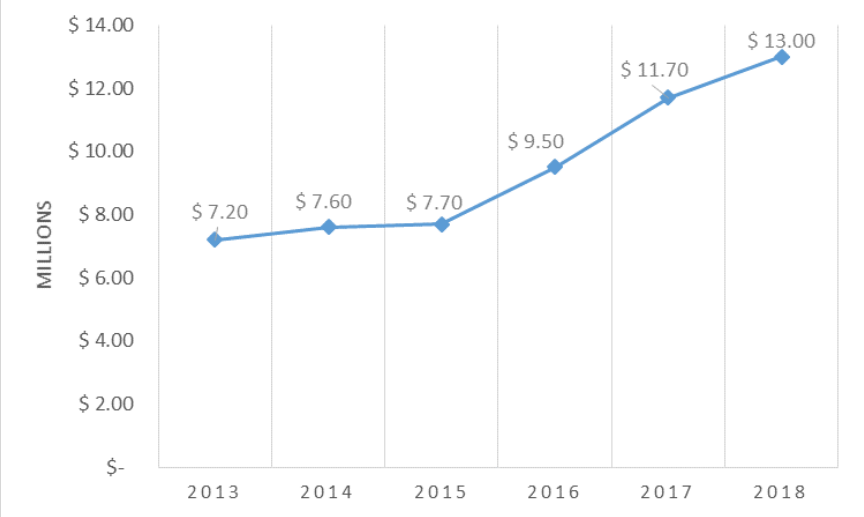

FIGURE 1.1 – Le coût moyen des cyber-attaques ces 6 dernières années

L'institut Ponemon et Accenture Security † ont publié en 2019 une étude sur les différents coûts que peuvent engendrer une cyber-attaque [15]. Le tableau 1.1 donne le coûts moyen annuel 2018 selon le type de perte et le type d'attaque :

Coût / Attaque	Interruption de l'activité	Perte d'information	Perte de revenue	Dommage équipement	Total
Malware	0,5$	1,4$	0,6$	0,1$	2,6$
Web-based attacks	0,3$	1,4$	0,6$	-$	2,3$
Denial of service	1,1$	0,2$	0,4$	0,1$	1,7$
Malicious insiders	0,6$	0,6$	0,3$	0,1$	1,6$
Phishing	0,4$	0,7$	0,3$	-$	1,4$
Malicious code	0,2$	0,9$	0,2$	-$	1,4$
Stolen devices	0,4$	0,4$	0,1$	0,1$	1,0$
Ransomware	0,2$	0,3$	0,1$	0,1$	0,7$
Botnets	0,1$	0,2$	0,1$	-$	0,4$
Total	4,0$	5,9$	2,6$	0,5$	13,0$

TABLE 1.1 – Coût moyen annuel par type d'attaque en 2018 (en million de dollars)

∗. L'institut Ponemon, créé en 2002, est un centre de recherche américain dédié à la protection des données et à la sécurité de l'information

†. Accenture Security, entité du groupe Accenture, conseille les entreprises et administrations dans le domaine de la cyber-sécurité

Ces attaques se sont développées au fil du temps. Elles sont passées d'un simple virus informatique à des attaques plus sophistiquées telles que les Rançongiciel ou les saturations de réseaux. L'objectif de ces attaques, que l'on détaillera dans la partie suivante, peut être multiple : exploitation ou revente des données, atteinte à l'image d'une entreprise, rançons etc.

1.1.2 Les risques liés aux cyber-attaques

Les cyber-risques sont les conséquences de cyber-attaques, qui peuvent être de 2 types : atteinte aux données et/ou atteinte du système d'information.

Atteinte aux données

L'atteinte aux données est une cyber-attaque qui vise à voler les informations personnelles. D'après la CNIL, sont considérées comme données personnelles, « toutes informations identifiant directement ou indirectement une personne physique (ex. nom, numéro d'immatriculation, numéro de téléphone, photographie, date de naissance, commune de résidence, empreinte digitale...) ». L'objectif de ces attaques peut être de deux types :

- soit exploiter ou revendre ces données (hameçonnage ou *phishing*) ;
- soit demander une rançon en échange de leur restitution (*ransomware*).

Atteinte aux systèmes d'information

L'atteinte aux systèmes d'information à pour but d'infiltrer un système informatique, dont l'objectif est de porter atteinte au fonctionnement de l'entreprise ainsi qu'à son image. Ces attaques peuvent-être menées de plusieurs façons :

- par déni de service ou par saturation (DDoS pour *Distributed Denial of Service*) : il s'agit de mettre hors ligne un serveur en le saturant et ainsi rendre inexploitable un site internet (un site marchand par exemple) ;
- par défiguration : il s'agit de modifier l'apparence ou le contenu d'un site internet à des fins politiques ou idéologiques ;
- par pollution de la source : il s'agit de piéger le site internet d'une entreprise ou d'une organisation pour ensuite infecter l'ordinateur de la personne ayant consulté ce site.

1.2 Les organismes de contrôle et de surveillance

1.2.1 L'agence européenne chargée de la sécurité des réseaux et de l'information

L'agence européenne chargée de la sécurité des réseaux et de l'information (ENISA) a été créée le 10 mars 2004. Son siège est à Heraklion en Grèce. Ses missions sont d'assurer la sécurité des réseaux et de l'information. L'agence opère de différentes façons :

- en intervenant en tant qu'experte auprès des autorités nationales et des institutions européennes ;
- en favorisant l'échange de meilleures pratiques ;
- en facilitant les contacts entre les institutions (nationales et européennes) et les entreprises.

L'ENISA, en collaboration avec les instances nationales et les institutions européennes, s'emploie à développer une culture de la sécurité des réseaux d'information dans toute l'Union Européenne.

1.2.2 L'Agence Nationale de la Sécurité des Systèmes d'Information

L'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI), créée le 7 juillet 2009, est un service à compétence nationale rattaché au Secrétaire Général de la Défense et de la Sécurité Nationale (SGDSN). L'Agence est chargée :

- de détecter et de réagir le plus tôt possible en cas d'attaque informatique. Un centre de détection est chargé de surveiller les réseaux et de mettre en œuvre les moyens de défense adaptés aux attaques ;
- de prévenir les menaces en développant des logiciels de très haute sécurité ainsi que des services adaptés pour les administrations et les acteurs économiques ;
- de jouer un rôle de conseil et de soutien aux administrations et aux Opérateurs d'Importance Vitale (OIV) ;
- d'informer régulièrement le public sur les menaces.

1.2.3 La Commission Nationale de l'Informatique et des Libertés

La Commission Nationale de l'Informatique et des Libertés (CNIL), est une autorité administrative indépendante créée le 6 janvier 1978. Elle est amenée à exercer plusieurs fonctions :

- veiller à ce que l'informatique ne porte atteinte ni au droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques ;
- réguler les données personnelles. Elle accompagne les organismes publics et privés dans leur mise en conformité avec la directive n^{o}2016/679 de l'Union européenne, dite Règlement Général sur la Protection des Données (RGPD), entré en vigueur le 25 mai 2018 ;
- aider les particuliers à maîtriser leurs données personnelles et exercer leurs droits ;
- proposer au gouvernement des mesures législatives ou réglementaires pour adapter la protection des données aux évolutions techniques ;
- contrôler et sanctionner les organismes. En cas de manquements constatés, elle peut les mettre en demeure d'y remédier.

1.3 La réglementation

1.3.1 Loi de Programmation Militaire

La Loi de Programmation Militaire (LPM) est une loi visant à planifier les dépenses militaires sur un horizon de 5 ans. Un livre blanc formalise les grandes orientations stratégiques définies par l'État, dont celles sur les dispositions relatives à la protection des infrastructures vitales contre la cyber-menace (chapitre IV). Elle précise les obligations en matière de cyber-sécurité à la charge des Opérateurs d'Importance Vitale (OIV) publics ou privés.

Les OIV sont regroupés par secteur d'activité d'importance vitale (SAIV). Il existe 12 secteurs d'activités qui sont sous la coordination du Ministre associé :

Catégories	Secteurs
- les secteurs étatiques	activités civiles de l'État ; activités militaires de l'État ; activités judiciaires ; espaces et recherches ;
- les secteurs de la protection des citoyens	santé ; gestion de l'eau et alimentation ;
- les secteurs de la vie économique et sociale de la nation	énergie ; communication ; électronique ; audiovisuel et information ; transports ; finances et industrie ;

TABLE 1.2 – Les secteurs d'activité d'importance vitale

La liste des OIV est définie par arrêté ministériel et n'est pas publique. Elle compterait plus de 230 sociétés, de la PME aux grands groupes du CAC 40. D'après l'article R1331-1 du code de la défense, les OIV gèrent des systèmes d'information dont « *[...]le dommage, l'indisponibilité ou la destruction, par suite d'un acte de malveillance, de sabotage ou de terrorisme, risquerait directement ou indirectement d'obérer gravement le potentiel de guerre ou économique, la sécurité ou la capacité de survie de la nation ou la vie de la population* ».

Les systèmes d'information des OIV sont appelés SIIV (Systèmes d'Information d'Importance Vitale). Ces SIIV sont situés sur des PIV (Points d'Importance Vitale) qui sont répartis sur le territoire national dans des ZIV (Zones d'Importances Vitales).

1.3.2 La directive NIS (Network and Information Security)

La directive NIS a pour but d'assurer un niveau de sécurité commun pour les réseaux et les systèmes d'information de l'Union européenne. Elle a été adoptée par les institutions européennes le 6 juillet 2016 et chaque État membre avait jusqu'au 9 mai 2018 pour la transposer dans son droit national. La directive est structurée autour de quatre axes :

- renforcer les capacités nationales en terme de cyber sécurité. Les États membres doivent se doter d'autorités nationales compétentes en matière de cyber-sécurité, de services pour répondre aux incidents informatiques (CSIRT [*]) et pour développer la stratégie de cyber sécurité,

- mettre en place une coopération volontaire entre États membres de l'Union européenne portant sur les aspects politiques et opérationnels. L'objectif de cette coopération est de faciliter le partage d'informations techniques sur les risques et les vulnérabilités ;

- renforcer la cyber sécurité des Opérateurs de Services (OSE [†]) qui sont essentiels au fonctionnement de l'économie et de la société. Les OSE ont l'obligation de notifier les incidents ayant un impact sur la continuité de leurs services. Des règles ont été définies au niveau national et les OSE devront s'y conformer ;

- instaurer des règles européennes communes en matière de cyber sécurité pour les prestataires de services numériques, en particulier dans les domaines de l'informatique par le *Cloud*, des moteurs de recherche et des places de marché en ligne [‡].

∗. Computer Security Incident Response Team.

†. les OSE correspondent essentiellement à de grandes entreprises ou des opérateurs du secteur public. La directive NIS est complémentaire au dispositif de cyber sécurité destinée aux Opérateurs d'Importance Vital (OIV).

‡. *Marketplaces* en anglais. Il s'agit des sites marchands comme par exemple Amazon, La Redoute ou la Fnac.

1.3.3 Règlement Général Européen sur la Protection des Données (RGPD)

Le Règlement Général sur la Protection des Données (RGPD) de l'Union européenne vise à harmoniser les règles relatives à la protection des données entre tous les pays de l'Union. Il a été adopté par le Parlement européen le 14 Avril 2016 puis a été mis en application le 25 Mai 2018. Il oblige les entreprises à mettre en conformité leurs systèmes d'information. En cas de manquement à leurs responsabilités, des sanctions pécuniaires peuvent s'élever jusqu'à 4% du chiffre d'affaires annuel mondial ou à 20 millions d'euros (le montant le plus élevé étant retenu). Ces sanctions peuvent être rendues publiques. Depuis son entrée en vigueur, la CNIL a reçu plus de onze mille plaintes en 2018, soit une augmentation de 32% par rapport à l'année précédente. Onze sanctions ont été prononcées, dont dix étaient pécuniaires et une était un avertissement non public. 2019 semble confirmer cette tendance à la hausse.

Chaque entreprise devra désigner un délégué à la protection des données qui aura pour mission de contrôler le respect du règlement, de conseiller le responsable de traitement et enfin d'avoir le rôle de point de contact.

1.4 Les garanties d'un contrat d'assurance Cyber

1.4.1 Les garanties de responsabilité civile

Dans le cadre du cyber-risque, les garanties de responsabilité civile couvrent les dommages causés à un tiers en cas d'atteinte aux données ou aux systèmes d'informations de l'assuré. Ces dommages peuvent être corporels, matériels ou immatériels. D'autres garanties interviennent dans le cadre de la responsabilité civile, à savoir :

- la garantie Enquête administrative et sanction : il s'agit de couvrir les frais de défense engagés par l'assuré dans le cadre d'une enquête exercée à son encontre par une autorité administrative compétente en matière de protection des données (la CNIL par exemple).
- la garantie Défense pénale et recours : il s'agit de prendre en charge les frais des actions judiciaires en vue :
  - de pourvoir à la défense de l'assuré devant la justice,
  - d'obtenir la réparation des dommages subis par l'assuré s'il en avait été tenu responsable.

1.4.2 Les garanties de dommages aux biens

Les garanties de dommages aux biens couvrent les biens matériels et immatériels en cas d'atteinte aux données personnelles et aux systèmes d'information. Bien que les biens immatériels

soient plus exposés au cyber-risque, la garantie va permettre l'indemnisation des frais suivants :

- frais d'investigation pour identifier la cause, l'origine et l'étendue du dommage ;
- frais de décontamination du système d'exploitation pour décontaminer le système d'information suite à l'introduction d'un logiciel malveillant ;
- frais de reconstitution des données volées dont l'assuré est propriétaire. Pour y parvenir, des dépenses supplémentaires doivent être engagées comme :
  - l'acquisition de programmes informatiques équivalents au jour du sinistre ;
  - la recherche et le rassemblement des informations disponibles à partir des sauvegardes internes ou sur tout périphérique de sortie ;
  - la saisie manuelle des données à partir des supports de données ou des programmes d'origine.

  Si les données ont été achetées, l'indemnisation est limitée à la valeur d'achat initiale.
- pertes financières résultant de fraudes, abus de confiance, détournements, escroqueries ou vols commis par l'intermédiaire du système d'information de l'assuré ;
- frais de notification. Il s'agit des frais occasionnés :
  - pour identifier les personnes concernées par l'atteinte à leurs données personnelles ;
  - pour le leur notifier ;
  - pour en avertir les autorités administratives compétentes.

1.4.3 Les garanties gestion de crise / assistance

Les garanties gestions de crise et assistance viennent compléter les garanties de responsabilité civile et de dommages aux biens. Ces garanties comprennent les services suivants :

- assistance informatique. Interventions d'experts informatiques pour procéder à la remise en route du système ;
- communication de crise. En cas d'atteinte à l'image ou à la réputation d'une personne morale, des services sont mis en place :
  - aide à la mise en place d'une stratégie de relations publiques et de communication avec les médias ;
  - Media Training pour préparer et aider à la communication directe avec les médias (journaux, télévision...) ;
  - rédaction d'argumentaires et de communiqués de presse ;

  - une ligne téléphonique est dédiée à l'appel des médias.

- gestion de l'image : en cas de crise « réputationnelle », un audit est réalisé pour évaluer l'impact et le retentissement de la situation actuelle sur les réseaux sociaux. Le ciblage est réalisé sur les principaux réseaux sociaux à fort impact (Twitter, Facebook, Linked'in...).

1.4.4 Les garanties optionnelles

D'autres garanties sont proposées dans le cadre d'un produit d'assurance Cyber, à savoir :

- la garantie Cyber extorsion : il s'agit des frais inhérents au paiement de la rançon, tels que les intérêts d'emprunt, les frais d'accompagnement au paiement de la rançon (les honoraires de consultants/négociateurs spécialisés ou de traducteurs/interprètes) ;

- la garantie Pertes d'exploitation : il s'agit de garantir la perte de marge brute résultant de la baisse du chiffre d'affaires causée par une interruption d'activité.

1.5 La réassurance

Le marché de la réassurance est en constante évolution pour le cyber-risque. Il existe aujourd'hui entre 60 et 70 acteurs. Seule une minorité d'entre eux couvrent les risques les plus complexes, les autres se focalisent principalement des risques mineurs et de moyenne importance. Leur rôle consiste également à accompagner les assureurs dans leur stratégie et leur politique de souscription tant les polices d'assurances sont diverses et variées : déterminer les garanties à inclure, rédiger les conditions générales, définir les exclusions et les cibles... Cet accompagnement est nécessaire sous peine de fortes restrictions dans le cadre des traités de réassurance.

Les capacités limites proposées par les réassureurs varient selon leur taille. Pour les plus petits d'entre eux, la limite sera de l'ordre de 3 à 5 millions de dollars, alors que pour les plus grands, elle peut aller jusqu'à 25 millions de dollars. Ces capacités sont déterminées selon la composition du portefeuille de la cédante. Si un portefeuille est risqué, la capacité sera soit faible soit chère. Par exemple, un portefeuille composé majoritairement d'institutions financières ou d'infrastructures hospitalières sera moins recherché puisque ce sont des cibles privilégiées par les cyber-criminels compte tenu des données précieuses qu'elles détiennent. Les cyber-criminels ciblent non seulement l'activité des banques mais aussi les données de leurs clients pour des usurpations d'identité ou pour les revendre sur le darkweb.

Les traités de réassurances se développent pour ce risque afin de proposer une offre adaptée aux assureurs déjà présents sur ce marché et à ceux qui souhaitent se lancer. On distingue deux types de réassurances, celle dite traditionnelle et celle dite alternative.

1.5.1 La réassurance traditionnelle

La réassurance traditionnelle comprend des traités classiques, de types *Quote-part*, *Stop-Loss* ou Excédent de Sinistre* (XS), dans lesquels certaines conditions vont être rajoutées. Dans le cadre du *Quote-part*, certains réassureurs vont exiger une limite par événement ou une limite annuelle. Par contre, la commission de réassurance doit être élevée pour financer les coûts d'acquisition et de développement des contrats cyber. Ce type de traité permet de se protéger contre une sinistralité de fréquence, de sévérité ou systémique. La maîtrise des services et la gestion des sinistres sont du ressort de l'assureur. Un traité de type *Stop-Loss* peut compléter le traité *Quote-part* afin de protéger en intensité et également diminuer la fréquence de sinistres y compris pour des sinistres systémiques. Ces traités spécifiques pour le cyber-risque vont nécessiter un investissement plus important de la part des cédantes car la mise en place d'une offre de ce type représente un coût significatif.

En réassurance traditionnelle, il existe aussi les traités de réassurance combinés, le risque cyber est inclus dans les autres traités destinés aux risques Responsabilité Civile ou Dommages aux Biens. Cela va également nécessiter un investissement de mise en place de l'offre plus important.

1.5.2 La réassurance alternative

La réassurance alternative consiste à proposer un traité de réassurance de type classique mais en « Marque blanche » où le réassureur fournit à la fois un support global assurance et réassurance. C'est lui qui apporte le produit et les services associés afin de faciliter les assurances désireuses d'entrer sur ce marché. Ceci permet aux assureurs de se lancer rapidement sur ce nouveau risque, avec peu d'investissements et avec un produit complet. Les traités classiques de type *Quote-part* ou XS seront proposés mais avec des conditions particulières. Pour le traité *Quote-part* par exemple, le taux de cession sera plus élevé, entre 75% et 100%. Il n'existera pas de limite annuelle contrairement à la réassurance traditionnelle. Par contre, la commission de réassurance sera, elle, plus faible puisque c'est le réassureur qui apporte le produit et les services associés. Un partenariat sur 3 ans minimum est souvent exigé. Pour le traité XS, une franchise annuelle sera proposée, connue sous le nom anglais de *Annual Aggregate Limite* (AAL). Ce type de traité sera proposé en « Marque blanche » à l'assureur qui a un petit portefeuille en croissance. Le réassureur aura donc la maîtrise des services et la gestion des sinistres, contrairement à la réassurance traditionnelle.

∗. *Excess of Loss* en anglais (XL)

1.6 Processus Entreprise Risk Management (ERM)

Dans cette dernière partie, on va se placer du point de vue de l'entreprise et on va définir la manière dont elle peut gérer le cyber-risque afin de ne pas compromettre sa situation financière ou sa réputation. La gestion des risques consiste à identifier, évaluer et prioriser les risques liés à une activité d'entreprise. Ces risques doivent être traités de manière méthodique afin de réduire à la fois la probabilité d'occurrence et l'impact de ces événements. Le processus ERM est le processus permettant de traiter ces risques méthodiquement et ainsi l'entreprise sera en mesure de mieux les gérer. Il se compose de trois étapes :

- Étape 1 : Identifier des risques
- Étape 2 : Les quantifier/mesurer
- Étape 3 : Les gérer/contrôler

Une entreprise va ainsi pouvoir identifier, mesurer puis contrôler l'ensemble des risques qui portent atteinte à sa stratégie. Le processus ERM est propre à chaque entreprise : il dépend de son activité, de sa situation financière ou économique, de sa stratégie de développement... On va par la suite définir ces trois étapes dans le cadre du cyber-risque de manière globale. Cette analyse nécessite d'être ajustée selon les caractéristiques spécifiques de chaque entreprise.

1.6.1 Identifier les risques

Les cyber-risques étudiés sont ceux présentés dans la partie 1.1.2. Il s'agit des risques liés à l'atteinte aux données (*phishing* et *ransomware*) et ceux liés à l'atteinte aux systèmes d'information (DDoS, attaques par défigurations, par pollution de la source). Les conséquences de ces attaques sont différentes selon leur nature :

- Les conséquences d'une atteinte aux données :
  - *phishing* : augmentation du risque de fraude et d'atteinte à la réputation, sanctions judiciaires et financières en cas de non respect de la réglementation RGPD ;
  - *ransomware* : destruction possible des données stratégiques, perte financière en cas de paiement de la rançon, dégradation de l'image si le vol est accompagné d'une divulgation publique.
- Les conséquences d'une atteinte aux systèmes d'information : perte d'exploitation, destruction possible de données stratégiques.

1.6.2 Mesurer les risques

La mesure des risques peut être aussi bien quantitative que qualitative. Dans notre cas, on va mesurer les cyber-risques de manière qualitative. On va étudier leur fréquence et leur impact, en brut de toute mesure de traitement.

Étude de la fréquence : pour l'étude de la fréquence, on distingue 4 niveaux de survenance : une survenance faible (1/4), modérée (2/4), élevée (3/4) et très élevée (4/4). La survenance peut être envisagée suivant plusieurs horizons (à un an ou à 5 ans par exemple).

Attaques	Explication	Fréquence
Atteinte aux données personnelles	Le vol de données personnelles est un risque de plus en plus fréquent. La question n'est plus de savoir « si » l'entreprise va se faire voler des données mais plutôt « quand ».	3/4
Atteinte aux systèmes d'information	L'atteinte aux systèmes d'information est moins fréquente que l'atteinte aux données. Il s'agit d'attaques ciblées sur un organisme en particulier avec un secteur d'activité bien précis, comme par exemple les sites internet marchand ou politique.	2/4

TABLE 1.3 – Étude de la fréquence des attaques

Étude de l'impact : pour l'étude sur l'impact, on distingue trois types d'impact : l'impact financier (F), juridique (J) et de l'image (I). Comme pour l'étude de la fréquence, on classifie ces impacts selon 4 niveaux : faible, modéré, élevé et très élevé. Pour l'atteinte aux données personnelles, on différencie les attaques *phishing* et ransomware. L'impact final retenu est souvent le maximum des 3 impacts suscités.

Attaques	Explication	Impact
phishing	**Financier** : coûts liés à la perte des données, à leur restauration, à la communication aux victimes et à tous les frais annexes (frais d'investigation, de décontamination, de notifications...).	Très élevé
	Juridique : suite à la nouvelle réglementation RGPD, des sanctions financières peuvent être appliquées jusqu'à 4% du chiffre d'affaire mondial ou à 20 millions d'euros.	Modéré
	Image : ces attaques sont souvent divulguées publiquement et nuisent à la réputation de l'entreprise.	Modéré
ransomware	**Financier** : les coûts correspondent au montant de la rançon qui est généralement élevé, ainsi qu'aux coûts de restauration des serveurs.	Élevé
	Juridique : Comme pour le *phishing*, le ransomware est concerné par la nouvelle réglementation RGPD.	Modéré
	Image : La demande de rançon peut rester en interne mais elle peut aussi être divulguée publiquement.	Modéré
Atteinte aux systèmes d'information	**Financier** : coût liés à la perte d'exploitation, au transfert d'activité vers un site de repli, au rachat de matériel, aux restaurations de serveurs...	Modéré
	Juridique : Comme pour les autres attaques, l'atteinte aux systèmes d'information est concerné par la nouvelle réglementation RGPD et requiert un niveau de sécurité suffisant sous peine de sanctions juridiques.	Modéré
	Image : ce type d'attaque peut être divulgué publiquement mais cela dépend du secteur d'activité de l'entreprise.	Faible

TABLE 1.4 – Étude de l'impact des attaques

Ces tableaux peuvent être synthétisés à partir d'une matrice, appelé « matrice des risques » ou plus communément « cartographie des risques » (figure 1.2).

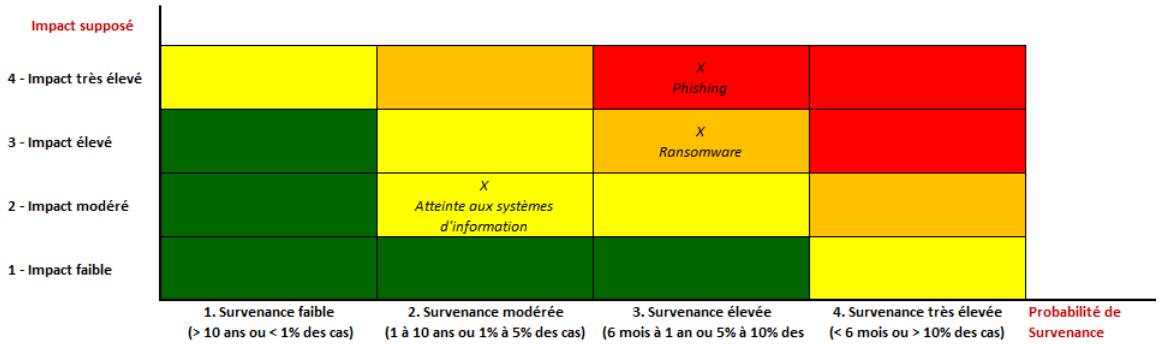

FIGURE 1.2 – Cartographie des cyber-risques

1.6.3 Gérer et contrôler les risques

La gestion d'un risque au sein d'une entreprise dépend de plusieurs facteurs : son degré de priorité, ses avantages en termes de coût/bénéfices, du budget prévu... Il existe principalement 4 options pour gérer un risque :

- soit le réduire,
- soit l'accepter tel qu'il a été identifié,
- soit le contourner,
- soit le transférer vers un tiers ou le partager (un sous-traitant spécialisé ou un assureur).

Dans le cadre du cyber-risque, deux options sont généralement utilisées : le réduire ou le partager.

Concernant le transfert de risque, de plus en plus d'assureurs proposent des contrats cyber avec des limites de garanties plus ou moins importantes selon le secteur d'activité de l'entreprise. Il existe également des sous-traitants informatiques qui gèrent entièrement la sécurité du réseau informatique.

Concernant la réduction du risque, plusieurs mesures de sécurité peuvent être mises en place. On les regroupe en trois grandes catégories : les mesures techniques, opérationnelles et humaines.

Catégories	Mesures
Mesures techniques	- Protection des postes de travail (antivirus, habilitation restreintes) - Protection des serveurs (antivirus) - Protection des réseaux (pare-feu internet, anti-spam intelligent) - Séparation des réseaux avec la mise en place de plusieurs systèmes d'informations sur des sites distincts. - Disposer de plusieurs fournisseurs Internet - Test périodique d'intrusion / audit de vulnérabilité
Mesures organisationnelles	- Gestion régulière des sauvegardes et restauration - Gestion des identités, des accès et habilitations - Se préparer à la gestion de crise avec des sites de repli prédéfinis
Mesures humaines	Sensibilisation des salariés à partir de conférences sur la cyber-sécurité, de Vidéoquizz, d'ateliers ou d'alertes mails.

TABLE 1.5 – Les mesures de sécurité pour réduire le risque

Des mesures spécifiques à chaque cyber-risque peuvent être mises en place. Une nouvelle matrice est cette fois-ci, établie sur la gravité de ces risques, contrairement à la matrice précédente qui était sur l'impact et la fréquence. La gravité mesure le produit de l'impact et de la fréquence. Dans notre exemple, une gravité très élevée correspond aux cases en rouge dans la matrice des risques, et une gravité faible aux cases en vert. Cette matrice prend en compte les mesures qui ont été mises en œuvre.

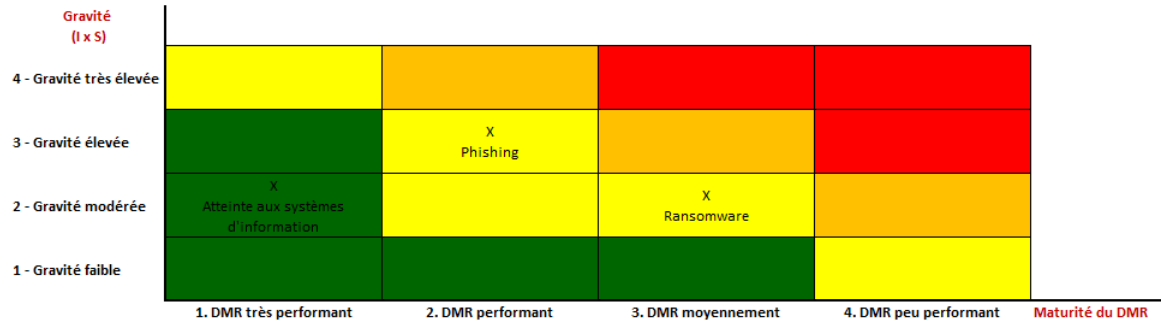

FIGURE 1.3 – Matrice sur la gravité des cyber-risques

Dans notre cas, cette matrice montre que les mesures mises en œuvre diminuent la gravité des attaques. Cette matrice va évoluer au fil du temps en fonction de la réglementation et de la stratégie de l'entreprise.

1.7 Conclusion

Une cyber-attaque est une atteinte aux données personnelles et/ou aux systèmes informatiques. Les risques sont considérables : pertes financières, interruption d'activité ou même atteinte à la réputation de l'entreprise. D'après la Fédération Française de l'Assurance (FFA), ce risque est classé en première position parmi les risques émergents, et cela sur les 5 prochaines années. Ces coûts n'ont cessé de croître depuis une décennie, atteignant un coût moyen de 13 millions de dollars en 2018, soit une augmentation de 80% par rapport à 2013. Les cibles de ces attaques sont multiples, il s'agit de tous les organismes, qu'ils soient privés ou publics, disposant d'un système informatique. Ces attaques peuvent impacter les organismes de plusieurs pays à la fois, et ce fut le cas pour l'attaque Wannacry, la plus importante de l'histoire, qui avait infecté les ordinateurs dans plus de 150 pays en 2017.

Plusieurs acteurs sont impliqués dans la mise en place de protection des victimes contre ce risque. L'union européenne et les gouvernements ont mis en place des mesures pour légiférer :

- sur le niveau de sécurité que doivent adopter les entreprises,
- sur les sanctions à appliquer si elle ne sont pas respectées.

Des organismes nationaux et européens existent pour contrôler, surveiller et accompagner les entreprises. C'est le cas de la CNIL, l'ENISA et l'ANSSI qui ont été créées respectivement en 1978, 2004 et 2009.

D'autres acteurs, comme les assureurs et les réassureurs, permettent aux entreprises de se couvrir contre ce risque qui évolue très vite. Les assureurs proposent des garanties adaptées, que ce soit au niveau des dommages matériels ou immatériels mais aussi au niveau de l'assistance et de la gestion de crise. Les réassureurs, quant à eux, proposent différents traités de réassurance

selon l'ancienneté et l'expérience de l'assureur sur ce marché. Des traités dits en « Marque blanche » permettent d'accompagner ceux qui souhaitent se lancer sur ce risque. D'après Munich-Re, le marché de l'assurance pour le cyber-risque pourrait atteindre près de 10 milliards de dollars en 2020 et jusqu'à 20 milliards en 2030.

Enfin, le dernier acteur, l'entreprise elle-même, peut prendre des mesures en interne pour diminuer ce risque. Elle peut mettre en place un processus ERM qui va permettre de traiter et gérer ce risque de façon méthodique. Il se compose de trois étapes : identifier, mesurer et contrôler le risque. Cela demande à l'entreprise de bien connaître ce risque et les impacts qu'il peut avoir sur ses activités.

Parmi tous ces acteurs, les assureurs sont aujourd'hui confrontés à la problématique liée à la tarification d'une assurance cyber. Comment tarifer ce risque alors que les données actuellement disponibles ne permettent pas d'appliquer les méthodes statistiques standard ? C'est la question que se pose SMACL Assurances, spécialiste du marché des collectivités. Dans la prochaine partie on présentera la mutuelle et ses enjeux liés à l'assurance cyber.

Chapitre 2

L'assurance Cyber : un enjeu pour SMACL Assurances

2.1 Brève présentation de SMACL Assurances

SMACL Assurances, Société Mutuelle d'Assurance des Collectivités Locales, a été créée en 1974 par et pour les élus locaux. Elle couvre les personnes morales de droit public, son cœur de cible, de manière spécifique à leurs besoins. Depuis, au fil du temps, elle a développé ses activités en y englobant également les associations, les entreprises et les particuliers.

Les différents types de sociétaires sont :

- les personnes morales de droit public : les collectivités territoriales, les établissements publics communaux, départementaux et régionaux. SMACL Assurances leur propose des garanties IARD, une assurance construction et le contrat risque statutaire ;
- les personnes morales de droit privé : les entreprises, les associations, les sociétés d'économie mixte et sociétés coopératives. SMACL Assurances leur propose uniquement l'assurance IARD ;
- les particuliers peuvent s'assurer pour les risques IARD, la protection juridique et la complémentaire santé. Elles sont concernées par :
  - les élus et anciens élus des assemblées communales, départementales et régionales,
  - les personnels des collectivités territoriales et des autres personnes morales de Droit Public,
  - les administrateurs et les personnels des personnes morales de Droit Privé.

SMACL Assurances a un chiffre d'affaires d'environ 412 millions d'euros en 2019 et se compose d'environ 17 000 collectivités, 60 000 associations, 7 000 entreprises et 50 000 particuliers.

Depuis le 1er janvier 2019, elle a intégré le groupe VYV né en septembre 2017, qui est une alliance de plusieurs mutuelles santé et de prévoyance.

2.2 Les caractéristiques du marché public

2.2.1 Le statut de la fonction publique

La fonction publique, qui représente un emploi sur cinq en France, est constituée des agents travaillant au sein des administrations publiques, telles que les communes, les conseils généraux, les conseils régionaux, les hôpitaux publics,.... En France, on distingue trois grandes fonctions publiques :

- la fonction publique d'Etat (2,4 millions d'agents) : services centraux des ministères, services régionaux ou départementaux de l'Etat ;

- la fonction publique territoriale (1,9 millions d'agents) : régions, départements, communes et établissements publics rattachés ;
- et la fonction publique hospitalière (1,2 millions d'agents) : établissements d'hospitalisation publics, maisons de retraite publiques, services départementaux de l'aide sociale à l'enfance, établissements publics pour mineurs ou adultes handicapes ou inadaptés, centres d'hébergement et de réadaptation sociale publics ou à caractère public.

Chaque fonction publique est régie par des dispositions spécifiques à caractère national. Parmi elles, sont définis les droits et obligations des agents :

- le droit à la protection (Loi n°83-634 du 13 juillet 1983, article 11) : ≪ les fonctionnaires et les agents non titulaires ont droit à une protection et le cas échéant à une réparation lorsque, dans l'exercice de leurs fonctions, ils ont fait l'objet de menaces, d'outrage, de voies de fait, d'injures ou de diffamation. Ils ont droit à une protection, dans certaines circonstances, en cas de poursuites pénales et civiles par un tiers pour faute de service ≫.
- l'obligation d'effectuer les tâches confiées (Loi n°83-634 du 13 juillet 1983, article 28) : ≪ tout fonctionnaire, quel que soit son rang dans la hiérarchie, est responsable de l'exécution des tâches qui lui sont confiées. Il n'est dégagé d'aucune des responsabilités propres de ses subordonnés ≫.

C'est en raison de ces dispositions particulières que des assurances spécifiques sont proposées aux agents et aux collectivités.

Parmi les agents publics, on distingue :

- les agents titulaires : il s'agit des agents ayant passé un concours de la fonction publique et qui occupent un emploi permanent. Ils sont soumis au statut spécifique des fonctionnaires.
- les agents non-titulaires : il s'agit des agents qui exercent une activité professionnelle pour une durée déterminée et qui ne bénéficient pas du statut de fonctionnaire.

2.2.2 Le contrat d'assurance : marché public de services

Le marché des collectivités est un marché très spécifique. D'après l'ordonnance n°2015-899 du 23 juillet 2015, les marchés publics sont ≪ des contrats conclus à titre onéreux par un ou plusieurs acheteurs publics avec un ou plusieurs opérateurs économiques publics ou privés, pour répondre à leurs besoins en matière de travaux, de fournitures ou de services ≫.

En 1998, le droit français de la directive européenne 92/50/CE impose aux collectivités de se conformer aux règles des marchés publics pour l'ensemble de leurs services. Lors de la passation

de ces marchés (marchés de travaux, marchés de services ou marchés de fournitures) et pour leur gestion, elles sont donc soumises au code des marchés publics et les contrats d'assurances y sont également soumis. Ce décret a incité les assureurs à modifier leurs contrats qui fonctionnaient jusqu'alors sur le mode de la reconduction en tacite. Dorénavant, les contrats d'assurances doivent avoir un terme et être conclus pour une ou plusieurs années. A l'issue de ce terme, l'assureur doit se présenter à nouveau à la procédure d'appel d'offre et une renégociation du contrat est alors possible. La procédure d'appel d'offre concerne les marchés dont la valeur est supérieure au seuil de 25 000€. Au-delà de cette somme, l'organisme public doit respecter une procédure formalisée pour passer son marché *.

2.2.3 Les différents formes de collectivités

On distingue différentes formes de collectivités territoriales. Elles sont répertoriées selon leur taille (en nombre d'habitants) et selon leur type d'administration. Le tableau 2.1 détaille cette classification † :

Type de collectivité	Nombre de collectivités
Communes de moins de 1000 habitants	25 580
Communes de 1000 à 1999 habitants	4 525
Communes de 2000 à 3999 habitants	2 199
Communes de 3000 à 4999 habitants	929
Communes de 5000 à 9999 habitants	1 138
Communes de 10000 à 49999 habitants	801
Communes de 50000 à 100000 habitants	75
Communes de plus de 100000 habitants	40
Communautés urbaines / métropoles	29
Communautés d'agglomération	219
Communautés de communes	1 018
Départements	98
Régions	14

TABLE 2.1 – Les différents types de collectivité

Une collectivité est défini par trois critères :

- elle est dotée de la personnalité morale ce qui lui permet d'agir en justice ;
- elle a une liberté d'administration : un budget et un effectif qui lui sont propres ;
- elle a une assemblée délibérante élue au suffrage universel direct (Conseil municipal, Conseil départemental, Conseil régional).

∗. Suite à ce décret, SMACL Assurances avait ouvert une salle des marchés afin répondre aux appels d'offre des collectivités.

†. Données disponible sur https ://www.collectivites-locales.gouv.fr

2.3 Les données exploitées par les collectivités

Les collectivités territoriales gèrent de plus en plus de données, dont certaines s'avèrent particulièrement sensibles. La loi « informatique et liberté » prévoit la possibilité de sanctionner pénalement les maires, les présidents de conseils généraux et de conseils régionaux en cas de manquement grave. Une étude sur les données personnels que traitent les collectivités avait été réalisé par la Gendarmerie Nationale en septembre 2017, commandité par le CREOGN *. Cette étude avait classé l'ensemble de ces données selon leur type [5] :

Types de données	Données
Individuelles	nom, prénom, date de naissance, lieu de naissance, nationalité, adresse postale, téléphone, mail ;
État civil	naissance, mariage (témoins, professions, ...), PACS, filiation, décès ;
Familiales	aides sociales, CAF, relations familiales et liens de famille, données de patrimoine ;
Biométriques	empreintes digitales, photo ;
Médicales	numéro de sécurité sociale, fiche médicale fournie par la famille, certificat médical, régime alimentaire, handicap ;
Ressources Humaines	CV, position, ancienneté, statut, absences, maladies, arrêts maladies, accidents de travail, sanctions, type de véhicule pour remboursement, situation de santé des conjoints ou enfants en vue d'ouverture de droits ;
Financières	RIB, dettes, non-valeur ;
Imposition	revenus fiscaux, quotient familial, données de redevance des ordures ménagères ;
Urbanisme	propriété des parcelles, location ;
Concession	lieu de la concession funéraire au cimetière, places disponibles ;
Police municipale	suivi de délinquance, infractions, verbalisation, pièces d'identité, numéro d'immatriculation véhicule, contrat d'assurance.

TABLE 2.2 – Classification des données personnelles que traitent les collectivités, réalisée par la Gendarmerie Nationale.

∗. Centre de Recherche de l'École des Officiers de la Gendarmerie Nationale.

2.4 Le principe de tarification des collectivités

Il existe deux modes de tarification qui dépendent de la taille de la collectivité :

- pour les plus importantes d'entre elles (agglomération, départements, régions), il s'agit d'une offre sur mesure où la tarification se base sur leur sinistralité historique et leur exposition aux risques.
- pour les plus petites d'entre elles (communes jusqu'à 7000 habitants), il s'agit d'une offre standard où la tarification se base sur des méthodes statistiques classiques.

Concernant l'offre standard, la tarification se fait à partir d'un taux ou d'un prix unitaire qu'on applique à une assiette. Cette assiette est différente selon le produit :

- pour les produits d'assurances dommages aux biens, l'assiette sera le nombre de m^2 ;
- pour les produits d'assurance responsabilité civile, l'assiette sera soit :
  - la masse salariale de la collectivité,
  - son nombre d'habitants,
  - ou son budget de fonctionnement.
- pour les produits d'assurance automobile, l'assiette sera le nombre de véhicule en parc de la collectivité.

Dans le cadre du produit d'assurance cyber, le tarif sera dans un premier temps forfaitaire. Il pourra à l'avenir dépendre d'une assiette que l'on détaillera dans les chapitre 4 et 5.

2.5 Pourquoi c'est un enjeu pour SMACL Assurances ?

Le cyber-risque touche de plus en plus souvent les collectivités. La dernière attaque en date a touché l'hôpital de Rouen : tout le système informatique de l'établissement avait été piraté. Les collectivités réalisent la gravité de ces attaques et leurs conséquences, d'autant plus qu'avec l'arrivée du règlement RGPD, les élus peuvent être sanctionnés en cas de manquement grave.

La demande pour se protéger contre de tels risques ne cesse de croître, et ce, quelle que soit la taille de la collectivité. Les collectivités les plus importantes émettent un cahier des charges qui liste l'ensemble des risques pour lesquels elles souhaitent se couvrir et le cyber-risque y prend une part de plus en plus importante. Pour répondre à cette demande, SMACL Assurances a créé un premier produit « Cyber Solution », qu'elle propose depuis 2018. Le tarif dépend de deux critères : du montant de franchise et du niveau de garantie, allant de 50 000€ à 600 000€.

Compte tenu de l'enjeu financier que comporte ce risque, SMACL Assurances souhaite revoir la structure tarifaire de ce produit.

2.6 Conclusion

SMACL Assurances spécialiste sur le marché public depuis 1974 propose notamment des couvertures d'assurances spécifiques aux personnes morales de droit public. Il s'agit de couvrir aussi bien les agents de la fonction publique que les bâtiments ou les véhicules appartenant aux collectivités territoriales (les communes, les départements, les régions, les hôpitaux, les maisons de retraites...).

Le statut de la fonction public est régi par des dispositions spécifiques à caractère national qui définissent les droits et obligations des agents. Parmi elles, les fonctionnaires et les agents non titulaires ont droit à une protection contre les risques qu'ils peuvent rencontrer lors de l'exercice de leur fonction. Ils sont également responsables de l'exécution des tâches qui leur sont confiées. SMACL Assurances propose donc des produits d'assurance spécifiques pour répondre à ces dispositions particulières.

Les collectivités sont soumises à la directive européenne 92/50/CE qui leur impose de se conformer aux règles des marchés publics pour l'ensemble de leurs services, et les contrats d'assurances en font partis. Ce décret oblige que les contrats aient une durée limité (de 1 à 4/5 ans en moyenne selon la taille de la collectivité) avec une date butoir déterminée lors de leur souscription. Les contrats ne peuvent plus être en « tacite reconduction ». Enfin, à l'échéance du contrat, l'assureur doit se soumettre à la procédure d'appel d'offre.

SMACL Assurances tarifie ces risques différemment selon la taille de la collectivité. Pour les plus importantes d'entre elles (agglomérations, départements ou régions), il s'agit d'une tarification sur mesure, basée sur leur sinistralité historique. A l'inverse, pour les plus petites d'entre elles (communes de moins de 7000 habitants), il s'agit d'une tarification basée sur des méthodes statistiques classiques.

Dans le cadre du cyber risque, SMACL Assurances propose un produit d'assurance depuis 2018. Ce produit faisait suite à une demande de plus en plus accrue de la part des collectivités pour ce couvrir contre ce risque. Un premier tarif avait été défini, dépendant de deux critères : le montant de garantie et le niveau de franchise. Compte tenu de l'enjeu financier que comporte ce risque, SMACL Assurances souhaite revoir la structure tarifaire de ce produit. Au vu du manque de données actuellement disponibles, que ce soit au niveau du portefeuille ou au niveau marché, il n'est pas possible d'appliquer des méthodes statistiques standards. Pour cette raison, plusieurs recherches ont été initiées par des scientifiques pour contourner ce problème de données. Parmi elles, une a attiré mon attention : il s'agit de modéliser le cyber-risque en s'inspirant de la modélisation pandémique. Dans la prochaine partie, on présentera cette nouvelle approche.

Chapitre 3

Modélisation du cyber-risque inspirée par la modélisation du risque pandémique

3.1 Pourquoi s'inspirer de la modélisation du risque pandémique ?

3.1.1 Les obstacles à contourner pour modéliser le cyber-risque

Obstacle 1 - Les données : le principal obstacle pour modéliser le cyber-risque concerne les données. Aujourd'hui leur quantité mais aussi leur granularité rendent difficile l'application de méthodes statistiques standard. Les données actuellement disponibles concernent principalement les cyber-attaques qui ont eu lieu aux États-Unis. Ces données, accessibles publiquement *, peuvent être utilisées pour modéliser ce risque et c'était le sujet du mémoire d'actuariat de Florian PONS en 2013 [16]. Il en avait déduit une « évaluation prudente de la prime pure » mais comme il l'avait précisé dans son mémoire, ces données présentent certaines limites :

- elles ne correspondent pas à la population assurée qui est propre à chaque portefeuille d'assurance. Les personnes assurées peuvent avoir en effet, des comportements ou des caractéristiques spécifiques, ce qui est le cas pour le portefeuille de SMACL Assurances composé principalement de personnes morales de droit public ;

- elles concernent les cyber-attaques qui ont eu lieu aux États-Unis uniquement. Or d'après l'institut Ponemon, une donnée volée aux États-Unis s'évalue à 225 dollars en moyenne alors qu'en France, elle s'évalue à 146 dollars, soit quasiment 35% moins cher.

L'entrée en vigueur du RGPD en mars 2018 oblige dorénavant les entreprises à signaler auprès de la CNIL une cyber-attaque dans leur système informatique. À terme, ce nouveau règlement permettra d'avoir accès à un ensemble d'informations † sur ce risque propre à la France, voire propre à l'Union Européenne, et ainsi le modéliser de manière plus pertinente.

Obstacle 2 - Le risque de cumul : un autre obstacle concerne le risque de cumul. La vitesse de propagation d'une cyber-attaque peut-être telle qu'il est difficile de l'appréhender, et elle peut ainsi infecter l'ensemble d'un réseau informatique, ce qui peut avoir comme conséquence l'accumulation des dommages/pertes. Le cyber-risque est considéré comme un risque non-stationnaire, qui évolue dans le temps en infectant tous les équipements informatiques connectés entre eux.

Obstacle 3 - Les coûts liés aux cyber-attaques : un dernier obstacle concerne les coûts liés aux cyber-attaques à prendre en compte dans le calcul de la prime pure. Ces coûts devraient inclure non seulement les pertes ou le vol des données mais aussi les dommages potentiels comme la réputation et les coûts associés à l'interruption du business.

∗. quelques sites internet permettent d'avoir accès à ces données dont celui-ci https ://www.privacyrights.org/data-breaches

†. J'avais contacté la CNIL au début de ce mémoire et ils m'avaient confirmé que les informations obtenues par le nouveau règlement RGPD seraient rendues publiques. Il n'existe pas encore de communiqué officiel

3.1.2 Le modèle épidémiologique : un modèle adapté au cyber-risque

Les différents obstacles, précisés en 3.1.1, incitent à développer des modèles mathématiques spécifiques pour le cyber-risque. Plusieurs méthodes ont été étudiées mais celle qu'on propose d'étudier assimile le cyber-risque au risque pandémique [18, 11, 8, 9, 19, 10, 2, 13, 12, 1]. Cette approche va permettre de modéliser et tarifer à un niveau micro. Un tarif sera définit pour chaque équipement informatique alors qu'au niveau macro, le tarif est définit sur l'ensemble d'un réseau informatique. Ce modèle aura aussi l'avantage de s'adapter facilement aux différents scénarios propres à chaque assuré.

Les similitudes avec le risque pandémique

Bien que la modélisation épidémiologique soit vieille de plus de trois cent ans, c'est en 1991 que deux chercheurs, Kephart et White, ont appliqué pour la première fois ce modèle à la propagation d'un virus informatique. Depuis, les recherches se sont intensifiées dans ce sens de par l'ampleur des cyber-attaques. Les raisons qui conduisent à assimiler le cyber-risque au risque pandémique sont les suivantes :

- un réseau informatique est un ensemble d'équipement reliés entre eux. Si un de ces équipements est infecté alors il pourra contaminer l'ensemble du réseau informatique, comme pour n'importe quelle maladie infectieuse.
- les différentes phases dues à une cyber-attaques peuvent s'assimiler à celles d'une maladie infectieuse, à savoir sain/en état de marche, infecté/hors service, guéri/réparé...

Les particularités du modèle pandémique et ses avantages

Le modèle pandémique présente l'avantage de pouvoir contourner, en grande partie, les obstacles cités précédemment. Ses particularités sont les suivantes :

- une modélisation basée sur des simulations : il existe principalement deux types de modélisation pour évaluer l'impact d'une pandémie : les modèles basés sur les données historiques et ceux basés sur des simulations. Comme il a été vu dans la partie précédente, le modèle se basant sur les données historiques n'est pas le plus approprié au vu des données disponibles. Le modèle basé sur des simulations a, quant à lui, l'intérêt de pouvoir contourner ce problème de données ;
- un processus stochastique : une partie des modèles pandémiques se basent sur un processus de Markov, un processus stochastique qui permet de prendre en compte la propagation dynamique d'une cyber-infection dans le temps. Les chaînes de Markov sont couramment utilisées pour modéliser les virus informatiques [18, 11, 8, 9, 10, 2, 13, 12] ;

- un modèle simple et flexible : certains modèles pandémiques sont des modèles simples et faciles d'utilisation (c'est le cas pour le modèle SIS, un modèle compartimental, que l'on détaillera par la suite). Ils permettent, en effet, de s'adapter aux différents scénarios propres à chaque assuré, notamment en prenant en compte les différents coûts liés aux cyber-attaques.

3.2 Les modèles pandémiques compartimentaux

3.2.1 Présentation

Le premier modèle mathématique d'épidémiologie a été introduit par Bernouilli en 1760. Depuis, de nombreux modèles existent et s'adaptent différemment selon les maladies contagieuses. On distingue deux types de modèles, ceux compartimentaux et ceux non compartimentaux [17, 7, 6, 3]. Par la suite on s'intéressera plus particulièrement aux modèles compartimentaux. Ce type de modèle, comme son nom l'indique, divisent la population en différents compartiments reflétant l'état d'un individu à une certaine période. Ces états peuvent être de différents types :

- sain mais susceptible de contracter la maladie
- infecté mais pas infectieux (assimilable à la période d'incubation)
- infecté et infectieux (possibilité de contaminer les individu sains)
- guéri ou décédé après avoir été infecté
- immunisé (à partir d'un vaccin par exemple)

Un individu peut passer d'un état à un autre avec des probabilités de passage. Les différents modèles compartimentaux peuvent être schématisés par la figure 3.1, où « S » représente les individus sains, « E » ceux infectés mais non infectieux, « I » ceux infectés et « R » ceux guéris.

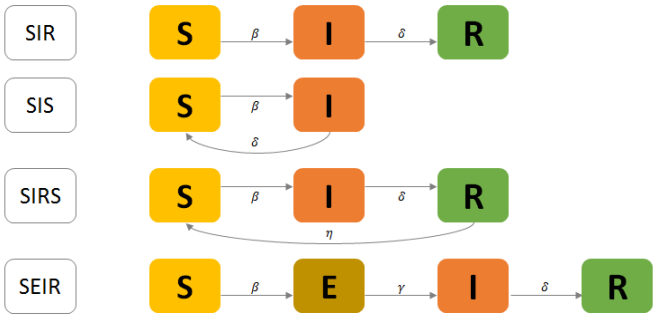

FIGURE 3.1 – Les différents modèles compartimentaux

3.2.2 Le modèle pandémique SIS

Parmi les modèles pandémiques compartimentaux, le modèle SIS est le plus couramment utilisé pour modéliser les réseaux et les services de télécommunication. Il présente la particularité d'être un modèle simple, avec seulement deux états :

- $S(t)$ les individus susceptibles d'être infectés à la période t ;
- $I(t)$ les individus infectés à la période t.

Avec $N(t) = S(t) + I(t)$ la population totale à la période t

Ce modèle ne concerne pas les maladies immunisantes, comme par exemple la varicelle, puisque les personnes ainsi infectées redeviennent susceptibles de subir une nouvelle infection.

Les taux de passage d'un état à un autre comprennent :

- un taux δ pour la probabilité de guérison ;
- un taux β pour la probabilité d'infection.

Les probabilités d'être infecté et d'être sain peuvent être représentées par les deux équations différentielles suivantes :

$$\begin{aligned}\frac{dS(t)}{dt} &= -\frac{\beta}{N}S(t)I(t) + \delta I(t)\\ \frac{dI(t)}{dt} &= \frac{\beta}{N}S(t)I(t) - \delta I(t)\end{aligned}$$

Ces deux équations sont composées de deux processus distincts qui peuvent être interprétés de la manière suivante : un individu sain avec la probabilité $S(t)$ va être infecté par un autre individu déjà infecté avec la probabilité $I(t)$. La propagation de l'infection est définie par le taux β. A l'inverse, un individu infecté avec une probabilité $I(t)$ va guérir avec un taux δ.

3.2.3 Le modèle SIS appliqué au cyber-risque

Comme il a été décrit dans la partie précédente, le cyber-risque présente des similitudes avec le modèle épidémiologique. Les différents états d'un équipement informatique peuvent, par analogie, s'assimiler à ceux d'un être humain en cas de pandémie. Dans le cadre du modèle SIS, on a donc deux états, à savoir :

- en état de marche pour « sain » ;
- hors service pour « infecté ».

Un réseau informatique, qui représente la population dans le cadre du modèle pandémique, est un ensemble d'équipements reliés entre eux. Si l'un des équipements est infecté, il pourra infecter le ou les équipements avec lequel il est relié. Comme pour le modèle SIS, la propagation de l'infection est caractérisée par le taux de passage β et le temps de réparation/guérison caractérisé par le taux δ. Une fois l'équipement informatique réparé, il est de nouveau vulnérable à une autre infection. Que ce soit le processus d'infection ou de guérison, les deux processus sont indépendants.

Dans le cadre du cyber-risque, le modèle SIS a été généralisé en ajoutant un nouvel élément. En effet, un équipement informatique pouvant créer lui-même un virus, un autre taux, le taux ϵ a été rajouté. Il existe donc deux types de menaces d'infection : celles provenant d'un autre ordinateur infecté avec un taux β et celle provenant de l'ordinateur lui-même avec le taux ϵ. L'ajout de cet élément supplémentaire d'infection s'explique par le fait que la propagation d'une épidémie est similaire à celle d'un virus dans un réseau informatique, dans lequel les équipements informatiques peuvent générer eux-mêmes des virus qui se répandent ensuite sur les autres équipements. Ce taux ϵ fait référence aux menaces extérieures comme par exemple un hacker qui tente de s'introduire dans un système informatique. Cette généralisation est appelée ici le modèle SIS-ϵ*, il correspond au modèle SIS quand $\epsilon = 0$.

3.3 Définition des paramètres pour la modélisation du cyber-risque

Comme il avait été défini dans le premier chapitre, un réseau informatique est composé de plusieurs équipements (ordinateurs, téléphones, tablettes...) reliés entre eux. Ce réseau peut être représenté par le schéma suivant (figure 3.2) :

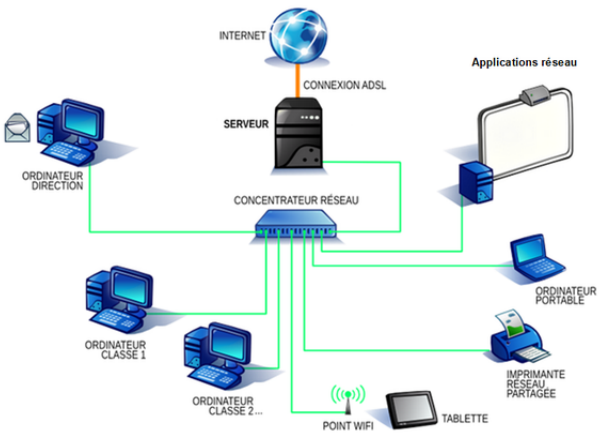

FIGURE 3.2 – Exemple d'un réseau informatique

*. Cette généralisation n'est pas nouvelle : le modèle SIS-ϵ a déjà été proposé par Hill et Al. [4], qui avaient considéré que la bonne humeur était comme une forme d'infection au sein d'un réseau social.

Quel que soit le réseau, on peut définir les paramètres suivants. Le réseau informatique lui-même peut être décrit comme $\Gamma = (V, E)$ où V correspondrait à l'ensemble des équipements informatiques que l'on appellera par la suite « nœuds » et E la connexion/liaison entre eux. $N = |V|$ correspond au nombre total de nœuds et $deg(v)$, degré de liberté du nœud v, correspond au nombre de connexion du nœud v avec les autres noeuds.

L'ensemble des connexions entre les noeuds peut être représenté par la matrice A, adjacente de Γ, où $A = (a_{vu})$ avec :

- $a_{vu} = 1$ si et seulement si $(u, v) \in E$;
- $a_{vu} = 0$ sinon ;
- et comme les noeuds ne peuvent pas s'auto-infecter, un autre cas de figure existe où $a_{vv} = 0$.

Le réseau informatique Γ peut être infecté par une cyber-attaque, où $(u, v) \in E$ correspond aux nœuds u et v qui peuvent s'attaquer entre eux. Le nœud v est donc :

- soit sain mais vulnérable aux attaques ;
- soit infecté et il peut dans ce cas, infecter d'autres nœuds à n'importe quel moment $t = 0, 1, 2, \ldots$.

L'état du réseau informatique au temps t peut être représenté comme $(I_1(t), \ldots, I_N(t))$ où $I_v(t) = 1$ représente le nœud v infecté au temps t et $I_v(t) = 0$ représente le nœud v sain au temps t.

3.4 Processus de propagation d'une infection

3.4.1 Probabilité d'infection

Le processus de propagation d'une infection suit un processus de Markov où l'on considère 2 types de menaces pouvant infecter les nœuds :

- celle provenant de l'extérieur du réseau : il peut s'agir d'un cyber-criminel qui attaque un équipement informatique ou d'un utilisateur qui visite un site dangereux. Cette menace extérieure se propage avec un taux ϵ ;
- celle provenant de l'intérieur du réseau : le nœud v est infecté alors il va pouvoir infecter les autres nœuds avec lesquels il est connecté. Cette menace intérieure se propage avec un taux β.

On peut ainsi résumer les taux de passage des différents états du noeud v de la manière suivante :

$$I_v(t) : 0 \to 1 \text{ avec un taux } q_v = \beta \sum_{j=1}^{N} a_{vj} I_j(t) + \epsilon_v$$

$$I_v(t) : 1 \to 0 \text{ avec un taux } \delta_v$$

La figure 3.3 schématise les états Sain ($I_v(t) = 0$) et Infecté ($I_v(t) = 1$) ainsi que les taux de passage :

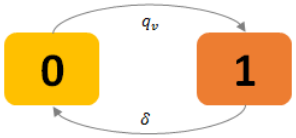

FIGURE 3.3 – Processus de Markov appliqué au cyber-risque

La probabilité d'infection est calculée à partir de l'équation différentielle suivante :

$$\frac{dI_v(t)}{dt} = (1 - I_v(t)) \left(\beta \sum_{j=1}^{N} a_{vj} I_j(t) + \epsilon \right) - \delta I_v(t)$$

Comme pour le modèle pandémique SIS, cette équation différentielle peut-être interprétée de la manière suivante :

- le noeud v est sain avec la probabilité $(1 - I_v(t))$ et il peut être infecté par les noeuds voisins infectés avec la probabilité $\sum_{j=1}^{N} a_{vj} I_j(t)$. La propagation d'infection entre les noeuds est définie par le taux β ;
- le noeud v est infecté avec la probabilité $I_v(t)$ et il va être réparé avec le taux δ.

On fait l'hypothèse que si le nœud v est infecté alors il sera réparé pour retourner à un état sain.

La probabilité d'infection pour un vecteur est formulée par :

$$p^T(t) = \Big[p_1(t), \ldots, p_N(t) \Big]$$

où $p_j(t) = P[I_j(t) = 1]$, pour $j = 0, 1, 2, \ldots, N$.

Pour calculer la probabilité d'infection, la méthode *Mean-field Approximation* (MFA) est utilisée, elle permet d'approximer une variable aléatoire par sa moyenne. Cette approximation est nécessaire dans le sens où le processus de Markov requiert un nombre réel et non une

variable aléatoire. Le caractère aléatoire provient du fait qu'il existe un nombre considérable de combinaisons possibles, puisqu'un noeud peut-être infecté par ses noeuds voisins qui peuvent-être eux-mêmes infectés par leurs propres voisins et ainsi de suite. Avec seulement deux états (Sain/infecté), le nombre total de combinaisons possibles est de 2^N, ce qui rend difficile son calcul. Pour cette raison, on présentera par la suite la méthode MFA qu'on appliquera au calcul de la probabilité d'infection $p_j(t)$.

3.4.2 Temps d'infection

Le temps d'infection est modélisé différemment si la menace provient de l'intérieur ou si elle provient de l'extérieur du réseau. Si elle provient de l'intérieur, on considère que tous les noeuds v ont D_v voisins infectés lançant des attaques via leurs connexions, avec $D_v = \sum_{j=1}^{N} a_{vj} I_j(t)$ où $I_j(t)$ est le statut du noeud j à l'instant t. Le temps d'infection sera alors modélisé par les variables aléatoires $Y_{v_1}, ..., Y_{v_{Dv}}$ qui ont la même distribution marginale F. Si la menace provient de l'extérieur, alors le temps d'infection sera modélisé par la variable aléatoire Z avec la distribution G_v. Par conséquent, le temps qu'il faudra pour que le noeud v soit infecté est :

$$T_v = min(Y_{v_1}, ..., Y_{v_{Dv}}, Z_v)$$

On assume que si le noeud v est infecté, alors il ne peut plus être attaqué. Par contre, dès qu'il sera rétabli, il sera de nouveau susceptible d'être infecté. Concernant le temps de réparation nécessaire pour un noeud infecté v, celui-ci est modélisé par R_v.

Comme pour la probabilité d'infection, on approximera le temps d'infection selon la méthode MFA.

3.5 Approximation des paramètres selon la méthode *Mean-field Approximation* (MFA)

Comme il a été dit précédemment, la probabilité d'infection et le temps d'infection sont des variables aléatoires dans le sens où un noeud peut être infecté par ses noeuds voisins qui, eux-mêmes peuvent être infectés et ainsi de suite. Or, pour appliquer le processus continue de Markov, cela requiert un nombre réel et non une variable aléatoire. Pour cette raison, une des méthodes couramment utilisée pour approximer une variable aléatoire est la méthode MFA, qu'on présente dans un premier temps pour ensuite l'appliquer à notre modèle.

3.5.1 Présentation de la méthode MFA

L'approximation MFA (approximation en champ moyen en français) est une méthode fréquemment utilisée en physique atomique. Cette approche consiste à transformer une variable aléatoire en une moyenne (voir schéma 3.4). Dans le cas d'un réseau informatique, cette approximation a été étudiée par Van Miegham et al (2009). Au lieu d'être infecté à la période t par le nombre de noeuds voisins infectés, le noeud v est maintenant infecté par le nombre **moyen** de noeuds voisins infectés. Cette approximation indépendante de premier ordre est connue sous le nom de *N-intertwined mean-field approximation (NIMFA).*

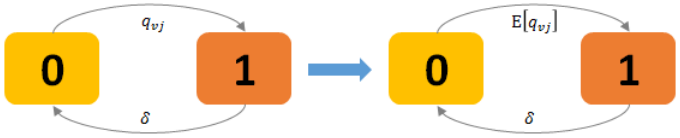

FIGURE 3.4 – La méthode MFA appliquée au cyber-risque

D'après le schéma de la figure 3.4, la méthode MFA, représentée par la flèche bleue, transforme la variable aléatoire q_{vj} en sa moyenne $\mathbb{E}[q_{vj}]$.

Précédemment, on avait fait l'hypothèse que tous les noeuds v à la période t ont deux états. Ils sont soit :

- infectés avec la probabilité $p_v(t) = P[I_v(t) = 1]$
- sains avec la probabilité $1 - p_v(t) = P[I_v(t) = 0]$

Si on applique la théorie de Markov, le générateur infinitésimal $Q_v(t)$ de ces deux états de la chaîne de Markov est :

$$Q_v(t) = \begin{pmatrix} -q_{v1} & q_{v1} \\ q_{v0} & -q_{v0} \end{pmatrix}$$

Avec

$$\begin{cases} q_{v1} &= \beta \sum_{j=1}^{N} a_{vj} 1_{\{I_j(t)=1\}} \\ q_{v0} &= \delta \end{cases}$$

Le taux q_{v1} correspond donc au taux de passage du noeud v d'un état sain à un état infecté et il est égal à la somme de tous les taux voisins infectés du noeud v *. Comme on a pu l'expliquer précédemment, pour pouvoir appliquer la théorie de Markov, le nombre total d'infections q_{v1} doit être un nombre réel et non une variable aléatoire. La nature aléatoire de q_{v1} se traduit par le nombre incalculable de combinaisons possibles : un noeud v peut être infecté par ses noeuds voisins qui peuvent être eux-mêmes infectés par d'autres noeuds et ainsi de suite. Le nombre

∗. On dit que le noeud v est entremêlé au reste du réseau puisqu'il peut être infecté par n'importe quel noeud voisin déjà infecté ($I_j(t) = 1$), d'où le nom de cette méthode en anglais *intertwined mean-field approximation* où *intertwine* signifie « entremêlé ».

total de combinaisons possibles est de 2^N. Pour cela, la méthode MFA permet d'approximer la variable aléatoire en un nombre réel en remplaçant q_{v1} par sa moyenne $\mathbb{E}[q_{v1}]$ et ainsi nous permettre d'appliquer la théorie continue de Markov. A partir de la méthode MFA, le nouveau générateur infinitésimal $\bar{Q}_v(t)$ peut être défini comme suit :

$$\bar{Q}_v(t) = \begin{pmatrix} -\mathbb{E}[q_{v1}] & \mathbb{E}[q_{v1}] \\ \mathbb{E}[q_{v0}] & -\mathbb{E}[q_{v0}] \end{pmatrix}$$

Avec

$$\begin{cases} \mathbb{E}[q_{v1}] &= \beta \sum_{j=1}^{N} a_{vj} P[I_j(t) = 1] = \beta \sum_{j=1}^{N} a_{vj} p_v(t) \\ \mathbb{E}[q_{v0}] &= q_{v0} = \delta \end{cases}$$

3.5.2 Approximation de la probabilité d'infection

Dans la partie précédente, on avait déterminé la probabilité d'infection d'après l'équation différentielle suivante :

$$\frac{dI_v(t)}{dt} = (1 - I_v(t)) \left(\beta \sum_{j=1}^{N} a_{vj} I_j(t) + \epsilon \right) - \delta I_v(t)$$

A partir de la méthode d'approximation MFA, on remplace $I_v(t)$ de l'équation précédente par sa moyenne $\mathbb{E}[I_v(t)]$. On obtient ainsi l'équation suivante :

$$\begin{aligned} \frac{d\mathbb{E}[I_v(t)]}{dt} =& \mathbb{E}\left[(1 - I_v(t)) \left(\beta \sum_{j=1}^{N} a_{ij} I_j(t) + \epsilon \right) - \delta I_v(t) \right] \\ \frac{d\mathbb{E}[I_v(t)]}{dt} =& (1 - \mathbb{E}[I_v(t)]) \left(\beta \sum_{j=1}^{N} a_{ij} \mathbb{E}[I_j(t)] + \epsilon \right) - \delta \mathbb{E}[I_v(t)] \end{aligned}$$

Soit $p_v(t) = \mathbb{E}[I_v(t)]$ et $p'_v = \frac{d\mathbb{E}[I_v(t)]}{dt}$, on obtient alors :

$$\begin{aligned} p'_v(t) =& (1 - p_v(t)) \left(\beta \sum_{j=1}^{N} a_{ij} p_j(t) + \epsilon \right) - \delta p_v(t) \\ p'_v(t) =& \beta \sum_{j=1}^{N} a_{ij} p_j(t) + \epsilon - \beta \sum_{j=1}^{N} a_{ij} p_j(t) p_v(t) - \epsilon p_v(t) - \delta p_v(t) \\ p'_v(t) =& \beta \sum_{j=1}^{N} a_{ij} p_j(t) + \epsilon - \beta \sum_{j=1}^{N} a_{ij} p_j(t) p_v(t) - (\epsilon + \delta) p_v(t) \end{aligned}$$

L'équation différentielle sous la forme matricielle s'écrit de la manière suivante :

$$\begin{aligned} P' =& \beta AP + \epsilon - \beta AP\text{diag}(p_v(t)) - \text{diag}(\epsilon + \delta)P \\ P' =& \left[\beta A - \text{diag}(\epsilon + \delta)\right] P + \epsilon - \beta AP\text{diag}(p_v(t)) \end{aligned}$$

avec $P = [p_1(t), p_2(t)..., p_N(t)]^T$.

Notons que pour tout $t \geq 0$ alors :

$$p_v(t) = \frac{\epsilon}{\delta + \epsilon}$$

On obtient donc l'équation suivante :

$$\begin{aligned} P' =& \left[\beta A - \text{diag}(\epsilon + \delta)\right] P + \epsilon - \beta AP\text{diag}(\frac{\epsilon}{\delta + \epsilon}) \\ P' =& \left[\beta\text{diag}\left(1 - \frac{\epsilon}{\delta + \epsilon}\right) A - \text{diag}(\epsilon + \delta)\right] P + \epsilon \\ P' =& \left[\text{diag}\left(\frac{\beta\delta}{\delta + \epsilon}\right) A - \text{diag}(\epsilon + \delta)\right] P + \epsilon \end{aligned}$$

Posons

$$Q = \text{diag}\left(\frac{\beta\delta}{\delta + \epsilon}\right) A - \text{diag}(\delta + \epsilon)$$

On obtient alors :

$$\begin{aligned} P' =& QP + \epsilon \\ P' - QP =& \epsilon \end{aligned}$$

Il s'agit d'une équation différentielle linéaire non homogène d'ordre 1. Elle peut être résolue explicitement de la manière suivante :

$$\begin{aligned} P(t) &= e^{Qt}P(0) + \int_0^t e^{Q(t-s)}\epsilon ds \\ &= e^{Qt}P(0) + Q^{-1}[e^{Qt} - I]\epsilon \end{aligned}$$

3.5.3 Approximation du temps d'infection

Comme pour la probabilité d'infection, le temps d'infection est approximé par la méthode MFA. Pour rappel, le nombre de voisins infectés est calculé de la manière suivante :

$$D_v = \sum_{j=1}^{N} a_{vj} I_j(t)$$

où $I_j(t)$ est le statut du noeud j à l'instant t.

A partir de la méthode MFA, on remplace $I_j(t)$ de l'équation précédente par sa moyenne $\mathbb{E}[I_j(t)]$ et on obtient ainsi

$$\mathbb{E}[I_j(t)] = P(I_j(t) = 1) = p_j(t)$$

Par conséquent, on a

$$\mathbb{E}[D_v] = \sum_{j=1}^{N} a_{vj} p_j(t)$$

3.6 Processus de renouvellement alterné

Après avoir approximé la probabilité et le temps d'infection par la méthode MFA, on va à présent déterminer le nombre de fois qu'un noeud v sera infecté puis sain. Pour cela, on utilise le processus de renouvellement qui va permettre de dénombrer le nombre d'occurrences de ces deux états. Dans un premier temps on présentera ce processus puis on l'appliquera à notre modèle.

3.6.1 Présentation

Le processus de renouvellement est un modèle stochastique, il modélise les événements qui surviennent aléatoirement dans le temps. Ce processus est appelé « renouvellement » car lorsqu'un événement survient, il se renouvelle automatiquement. On le dit « alterné », quand il alterne entre deux états. Dans notre cas, ces deux états sont : sain et infecté.

Précédemment, dans la partie 3.4.2, on avait défini le temps qu'il faut pour que le noeud v soit infecté ($T_v = min(Y_{v_1}, ..., Y_{v_{D^v}}, Z_v)$) et pour qu'il soit réparé (R_v). Le processus de renouvellement alterné pour le noeud v commence à l'état sain et reste dans cet état pendant une période T_{v1}. Puis une fois infecté, il reste dans cet état pendant une période R_{v1} et ainsi de suite. Les périodes successives du noeud v peuvent être définies de la manière suivante :

- soit $\{T_{vn} : n \geq 1\}$ avec $T_v = (T_{v1}, T_{v2}, ...)$: les périodes successives pour que le noeud v devienne infecté,

- soit $\{R_{vn} : n \geq 1\}$ avec $R_v = (R_{v1}, R_{v2}, ...)$: les périodes successives pour que le noeud v soit réparé.

$((T_{v1}, R_{v1}); (T_{v2}, R_{v2}); ...)$ est une séquence indépendante et identiquement distribuée. L'intervalle de renouvellement peut ainsi être caractérisé comme suit : $X_{vn} = R_{vn} + T_{vn}$,

La fonction de renouvellement est définie comme l'espérance de saut d'une période à une autre. Soit $\mathbb{E}[T_v]$ la durée moyenne pour que le noeud v devienne infecté et $\mathbb{E}[R_v]$ la durée moyenne pour qu'il soit réparé.

On rappelle que $I_v(t)$ est le statut du noeud v à l'instant t et que la probabilité d'infection est $p_v(t) = P[I_v(t) = 1]$. A partir du théorème de renouvellement, on obtient alors :

$$p_v = \lim_{n\to\infty} \frac{1}{t} \int_0^t I_v(s)ds = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + \mathbb{E}[T_v]}$$

Ce modèle comprend les particularités suivantes :

- Un noeud déjà infecté ne peut pas être infecté de nouveau. Les noeuds voisins infectés cessent d'attaquer quand le noeud v est infecté ;
- les cyber-attaques peuvent être indépendantes ou dépendantes entre elles.

Par la suite, on va déterminer la probabilité d'infection.

3.6.2 Indépendance entre les processus d'infection

Dans le cas où les périodes d'infection sont indépendantes entre elles, le temps moyen d'infection $\mathbb{E}[T_v]$ est défini comme suit :

$$\mathbb{E}[T_v] = \mathbb{E}[min(Y_{v_1}, ..., Y_{v_{D^v}}, Z_v)] = \mathbb{E}\left[\int_0^\infty \bar{F}^{D_v}(x)\bar{G}_v(x)dx\right]$$

avec

- $\bar{F}^{D_v}(x, .., x) = P(Y_{v_1} > x, ..., Y_{v_{D_v}} > x)$ où F correspond à la distribution marginale du temps d'infection quand la menace provient de **l'intérieur** du réseau ;
- $\bar{G}_v(x) = P(Z_v > x)$ où G correspond à la distribution marginale du temps d'infection quand la menace provient de **l'extérieur** du réseau.

D'après l'inégalité de Jensen, on a

$$\mathbb{E}\left[\int_0^\infty \bar{F}^{D_v}(x)\bar{G}_v(x)dx\right] \geq \int_0^\infty \bar{F}^{\mathbb{E}[D_v]}(x)\bar{G}_v(x)dx$$

D'où la probabilité d'infection

$$p_v \leq \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + \int_0^\infty \bar{F}^{\mathbb{E}[D_v]}(x)\bar{G}_v(x)dx}$$

On considère l'équation d'infection suivante :

$$p_v^* = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + \int_0^\infty \bar{F}^{\sum_{j=1}^N a_{vj}p_j^*(t)}(x)\bar{G}_v(x)dx} \tag{3.1}$$

3.7 Processus de sévérité d'une infection

Lorsqu'un ordinateur est infecté, celui-ci va impliquer deux types de pertes :

- Les pertes causées par l'infection elle-même, telles que le vol de données à caractère personnel, le dommage des données, les frais juridique pour payer les dommages et intérêts aux tiers ou les frais pour notifier les personnes infectées. Le premier type de perte est modélisé par un coût aléatoire $\eta_v(L_{v,1})$, où $(L_{v,1})$ signifie les pertes d'information (comme par exemple les données personnelles) et elles peuvent aussi être utilisées pour modéliser les coûts de responsabilité civile.

- Les pertes causées par la restauration du nœud infecté à un état sain. Ce deuxième type de perte est assimilée à la durée de réparation ou de hors-service du nœud. Il peut être modélisé par un coût de l'activité $C_v(R_{v,1})$, où $(R_{v,1})$ est la durée de hors-service.

La figure 3.5 schématise les coûts d'une infection distribuées dans le temps.

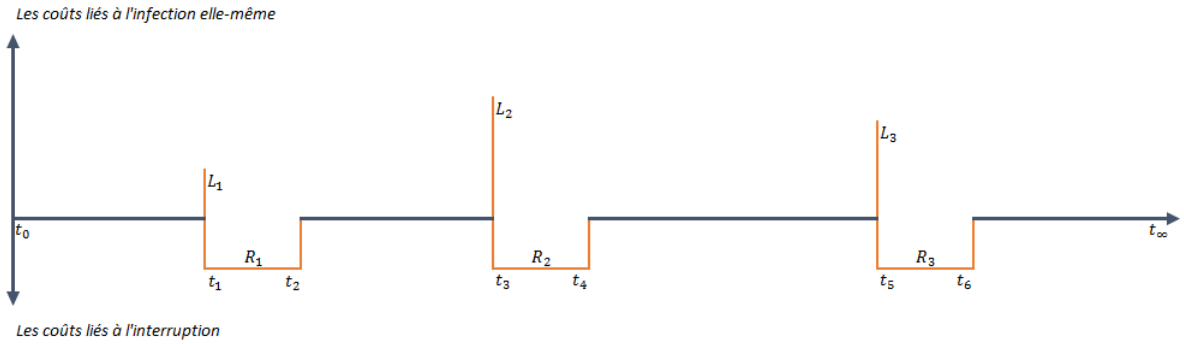

FIGURE 3.5 – Exemple des pertes liées à une cyber-attaque pour le nœud v

Au temps $T = t_2$, le nœud v est sain mais vulnérable à d'autres attaques et il pourra être encore une fois infecté aux temps t_3 et t_5. Par conséquence, pour le nœud v, la perte cumulée au temps t peut être représentée de la manière suivante :

$$s_v(t) = \sum_{i=1}^{M_v(t)} [\eta_v(L_{v,i}) + C_v(R_{v,i})]$$

Où $\eta_v(.)$ représente le coût due à l'infection et $C_v(.)$ représente le coût de l'activité associé avec la durée de la période $R_{v,i}$ en hors service. Pour chaque nœud v, en fait, il s'agit d'un processus qui se répète à chaque période. La perte totale à laquelle une entreprise peut faire face pour une période comprise entre $[0; t]$ est :

$$S(t) = \sum_{v=1}^{N} s_v(t) = \sum_{v=1}^{N} \sum_{i=1}^{M_v(t)} [\eta_v(L_{v,i}) + C_v(R_{v,i})] \tag{3.2}$$

Où $M_v(t)$ est le nombre total d'infections du nœud v jusqu'au temps t. L'équation précédente montre que la quantité clé est le vecteur d'infection $[I_1(t), \ldots, I_N(t)]$, lequel requiert la théorie de l'épidémie.

3.8 Conclusion

Depuis une décennie, les recherches pour modéliser le cyber-risque se sont intensifiées de part l'ampleur des cyber-attaques, en terme de coût et de fréquence. La modélisation de ce risque présente un certain nombre d'obstacles, le principal étant celui concernant les données : leur quantité et leur granularité rendent difficiles l'application de modèles statistiques standards. C'est pour cette raison, que des chercheurs ont axé leurs recherches en assimilant ce risque au risque pandémique ; ces deux risques présentent en effet des similitudes :

- un réseau informatique peut s'assimiler à une population, où un ordinateur infecté peut alors contaminer l'ensemble du réseau comme n'importe quelle maladie infectieuse ;
- les différents états liés à une cyber-attaques peuvent s'assimiler à ceux d'une maladie infectieuse, à savoir sain/en état de marche, infecté/hors service, guéri/réparé.

Les premiers modèles pandémiques datent du XVIIIe siècle et depuis, de nombreux modèles ont été créés, chacun s'adaptant à une épidémie spécifique. Parmi ceux-là, un modèle est couramment utilisé pour modéliser les réseaux informatiques et les services de télécommunication. Il s'agit du modèle comportemental SIS, qui est un modèle simple avec seulement deux états : sain et infecté. Pour passer d'un état à un autre, des taux passages sont ainsi définis : β pour le passage d'un état sain à infecté et δ pour le passage d'un état infecté à sain. Des chercheurs ont généralisé ce modèle afin de l'adapter aux particularités du cyber-risque : ainsi en plus du taux β, ils ont ajouté un autre taux, le taux ϵ. Ces deux taux sont complémentaires et peuvent s'interpréter de la manière suivante :

- le taux ϵ correspond à une menace provenant de l'extérieur. Ce taux peut faire référence à un cyber-criminel qui va introduire un virus dans un ordinateur ou bien à une personne qui va consulter un site internet malicieux ;

- le taux β correspond à une menace provenant de l'intérieur. Une fois le virus introduit dans le réseau, celui-ci va se propager avec le taux β et contaminer ainsi d'autres ordinateurs.

Les paramètres des deux modèles, pandémiques et cyber, peuvent être ainsi résumés dans le tableau suivant :

Paramètre	Modèle Pandémique SIS	Modèle Cyber
- taux de passage d'un état sain à un état infecté	β	$\beta \sum_{j=1}^{N} a_{vj} I_j(t) + \epsilon$
- taux de passage d'un état infecté à un état sain	δ	δ
- Probabilité d'infection	$\frac{\beta}{N} S(t) I(t) - \delta I(t)$	$S_v(t) \left(\beta \sum_{j=1}^{N} a_{vj} I_j(t) + \epsilon \right) - \delta I_v(t)$

TABLE 3.1 – Récapitulatif des paramètres pour les modèles pandémique SIS ou Cyber

Un autre paramètre est utilisé dans notre modèle : il s'agit du temps d'infection. Il est défini comme suit :

$$T_v = min(Y_{v_1}, ..., Y_{v_{Dv}}, Z_v)$$

où $Y_{v_1}, ..., Y_{v_{Dv}}$ correspond au temps qu'il faut pour que le noeud v soit infecté par une menace intérieure (i.e. par ses noeuds voisins) et Z_v le temps qu'il faut pour que le noeud v soit infecté par une menace extérieure.

La probabilité d'infection et les temps d'infection sont des variables aléatoires dans le sens où un noeud peut être infecté par ses noeuds voisins qui, eux-mêmes peuvent être infectés par leurs voisins et ainsi de suite. Or, pour appliquer le processus continue de Markov, cela requiert un nombre réel et non une variable aléatoire. Pour cela, la méthode *Mean Field Apprxomiation* (MFA) permet d'approximer une variable aléatoire par sa moyenne et ainsi d'appliquer le processus de Markov. On obtient ainsi la probabilité d'infection suivante sous sa forme matricielle :

$$P(t) = e^{Qt} P(0) + Q^{-1}[e^{Qt} - I]\epsilon$$

avec $Q = \text{diag}\left(\frac{\beta\delta}{\delta+\epsilon}\right) A - \text{diag}(\delta + \epsilon)$

Une fois la probabilité d'infection approximée, on applique le processus de renouvellement afin de déterminer le nombre de fois qu'un noeud v sera infecté et sain. La probabilité d'infection est définie selon le cas de figures où les processus d'infection sont indépendants :

$$p_v^* = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + \int_0^\infty \bar{F}^{\sum_{j=1}^{N} a_{vj} p_j^*(t)}(x) \bar{G}_v(x) dx}$$

Notre modèle prend en compte également les pertes liées aux cyber-attaques. On en distingue deux types :

- les pertes liées à l'infection elle-même modélisée par un coût aléatoire $\eta_v(L_{v,1})$ où $(L_{v,1})$ correspond à la perte d'information ;
- les pertes liées à la restauration du noeud infecté à un état sain modélisé par un coût de l'activité : $C_v(R_{v,1})$ où $(R_{v,1})$ correspond à la durée de la restauration.

Les pertes totales cumulées d'un réseau informatique sont définies par la formule suivante :

$$S(t) = \sum_{v=1}^{N} s_v(t) = \sum_{v=1}^{N} \sum_{i=1}^{M_v(t)} [\eta_v(L_{v,i}) + C_v(R_{v,i})]$$

Où $M_v(t)$ correspond au nombre total d'infection durant une période t.

Dans les prochaines parties on analysera ce modèle puis on l'appliquera dans le cas d'une assurance Cyber sur le portefeuille de SMACL Assurances.

Chapitre 4

Analyse et calibration des différents paramètres

4.1 Les différents types de réseaux informatiques

Dans le premier chapitre, on avait définit un réseau informatique comme un ensemble d'équipements informatique reliés entre eux. Il existe, cependant, différents types de réseaux que l'on catégorise selon deux caractéristiques : leur topologie et leur typologie.

Topologie du réseau : la topologie du réseau fait référence à sa taille et à sa portée. On distingue les topologies suivantes :

- réseau personnel (PAN pour *Personal Area Network*) : il s'agit des connexions ne couvrant que quelques mètres, comme par exemple le bluetooth où la connexion est possible seulement au sein d'une même pièce voire d'une maison.
- réseau local (LAN pour *Local Area Network*) : il s'agit des connexions couvrant une distance plus importante que le réseau personnel, comme par exemple le WIFI où la connexion est possible dans un même bâtiment. Ce type de réseau est utilisé par les entreprises mais également par les institutions publiques comme les administrations, les écoles ou les universités.
- réseau métropolitain (MAN pour *Metropolitan Area Network*) : ce type de réseau permet de relier plusieurs LAN sur une zone géographique relativement proche.
- réseau étendu (WAN pour *Wide Area Network*) : ce type de réseau permet de couvrir une zone géographique beaucoup plus large que la MAN, à l'échelle d'un pays voire d'un continent.
- réseau global (GAN pour *Global Area Network*) : ce type de réseau couvre plusieurs WAN et permet ainsi à des entreprises internationales de se connecter dans le monde entier. Les GAN utilisent les infrastructures de fibre optique des WAN et les relie entre elles par des câbles sous-marins internationaux ou par des transmissions satellites.

Typologie du réseau : La typologie du réseau fait référence à la manière dont sont reliés entre eux les équipements informatiques. On peut les classer en cinq grands types :

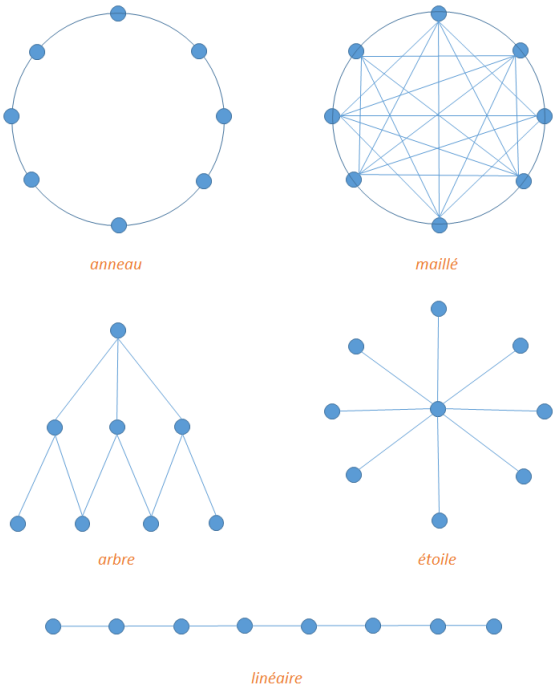

FIGURE 4.1 – Les différents types de connexions possibles d'un réseau informatique

4.2 Analyse du processus de propagation d'une infection

Dans cette partie, on analysera la probabilité d'infection selon différents paramètres : selon le degré de liberté des noeuds, le taux de la menace intérieure/extérieure, le taux de guérison/réparation et le nombre d'équipements informatiques. Mais avant toute chose, on va partir d'une hypothèse.

4.2.1 Hypothèse

Pour le processus de propagation d'une infection, on fait l'hypothèse qu'il suit la distribution exponentielle. Cette loi est généralement utilisée pour modéliser la durée de vie d'un phénomène sans vieillissement ou sans usure. Dans notre cas, la durée de vie correspond à la durée pour laquelle un ordinateur reste dans un état sain. A partir de la formule (3.1) du chapitre précédent, qu'on rappelle ci-dessous, on obtient la probabilité d'infection suivante * :

$$p_v^* = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + \mathbb{E}[T_v]} = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + \int_0^\infty \bar{F}^{\sum_{j=1}^N a_{vj} p_j^*(t)}(x) \bar{G}_v(x) dx}$$

Où $F(x)$ et $G(x)$ correspondent aux distributions marginales du temps d'infection quand la menace provient respectivement de l'intérieur et de l'extérieur du réseau. On fait donc

∗. d'autres lois existent pour modéliser le processus de propagation d'une infection, comme les lois de Weibull ou Log-normale et dont les calculs sont détaillés en annexe A

l'hypothèse que les deux distributions marginales suivent la loi exponentielle. On obtient ainsi les fonctions de survie suivantes : $\bar{F}(x) = e^{-\beta x}$ et $\bar{G}(x) = e^{-\epsilon_v x}$, d'où la durée moyenne pour que le noeud v devienne infecté :

$$\begin{aligned}
\mathbb{E}[T_v] &= \int_0^\infty \bar{F}^{\sum_{j=1}^N a_{vj} p_j^*(t)}(x)\bar{G}_v(x)dx \\
&= \int_0^\infty (e^{-\beta x})^{\sum_{j=1}^N a_{vj} p_j^*(t)} e^{-\epsilon_v x} dx \\
&= \int_0^\infty e^{-x(\beta \sum_{j=1}^N a_{vj} p_j^*(t) + \epsilon_v)} dx \\
&= \left[-\frac{1}{\left(\beta \sum_{j=1}^N a_{vj} p_j^*(t) + \epsilon_v\right)} e^{-x(\beta \sum_{j=1}^N a_{vj} p_j^*(t) + \epsilon_v)} \right]_0^\infty \\
&= \frac{1}{\beta \sum_{j=1}^N a_{vj} p_j^* + \epsilon_v}
\end{aligned}$$

Comme $\lim_{x \to -\infty} e^x = 0$.

On obtient alors la formule suivante :

$$p_v^* = \frac{\mathbb{E}[R_v]}{\mathbb{E}[R_v] + 1/\left(\beta \sum_{j=1}^N a_{vj} p_j^* + \epsilon_v\right)} \tag{4.1}$$

Si on assume par conséquent que le processus de guérison/réparation suit également une loi exponentielle, alors :

$$\bar{S}(x) = P(R_v > x) = e^{-\delta_v x}$$

On en déduit le modèle Markovien, selon la formule du processus d'infection. On obtient ainsi :

$$p_v^* = \frac{1/\delta_v}{1/\delta_v + 1/\left(\epsilon_v + \beta \sum_{j=1}^N a_{vj} p_j^*\right)} = \frac{\beta \sum_{j=1}^N a_{vj} p_j^* + \epsilon_v}{\beta \sum_{j=1}^N a_{vj} p_j^* + \epsilon_v + \delta_v}$$

4.2.2 Analyse selon le degré de liberté

On va, dans un premier temps, analyser la probabilité d'infection selon le degré de liberté, qui correspond au nombre de connexions pour un noeud. On va prendre le cas où les taux de la menace intérieure et extérieure sont identiques $\beta = \epsilon = 0.5$ et le taux de guérison variera $\delta = (2, 3, 4)$:

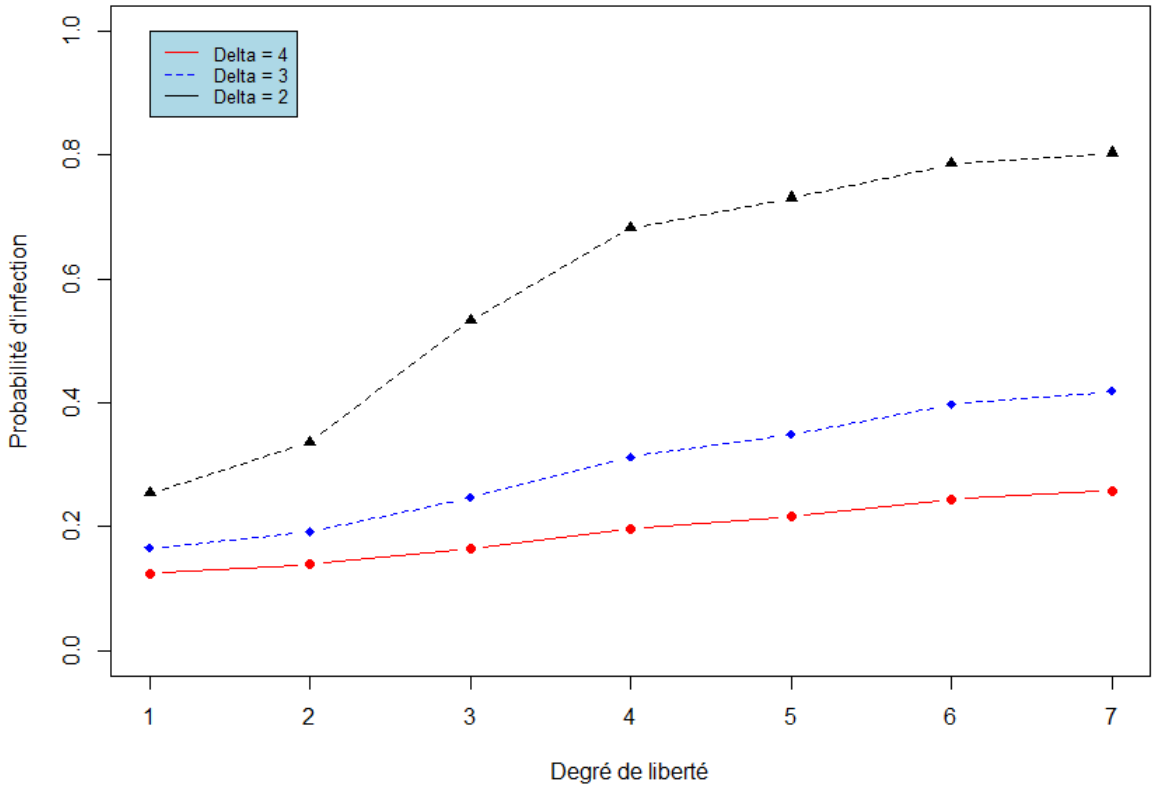

FIGURE 4.2 – Analyse selon le degré de liberté

La figure 4.2 montre que la probabilité d'infection augmente en fonction du nombre de connexions, et plus l'écart entre le taux de guérison (δ) et les taux de menace (β, ϵ) est faible, plus la probabilité d'infection est élevée.

On a vu précédemment qu'il existe différents types de réseau. On va maintenant comparer la probabilité de leur infection. Prenons l'exemple d'un réseau composé de 5 ordinateurs ou serveurs* avec les taux de passage suivants : $\beta = \epsilon = 0,1$ et $\delta = 0,5$. On obtient les résultats dans le tableau suivants :

Noeuds	étoile	maille	anneau	linéaire	arbre
Noeud 1	0,4715	0,8667	0,5238	0,4390	0,5632
Noeud 2	0,4715	0,8667	0,5238	0,5054	0,6042
Noeud 3	0,6024	0,8667	0,5238	0,5158	0,6042
Noeud 4	0,4715	0,8667	0,5238	0,5054	0,5632
Noeud 5	0,4715	0,8667	0,5238	0,4390	0,5632

TABLE 4.1 – Probabilité d'infection selon le type de réseau - p_v^*

∗. l'annexe B répertorie les matrices adjacentes de chaque réseau.

Noeud	étoile	maille	anneau	linéaire	arbre
Noeud 1	1,1207	0,1538	0,9091	1,2778	0,7755
Noeud 2	1,1207	0,1538	0,9091	0,9787	0,6551
Noeud 3	0,6599	0,1538	0,9091	0,9388	0,6551
Noeud 4	1,1207	0,1538	0,9091	0,9787	0,7755
Noeud 5	1,1207	0,1538	0,9091	1,2778	0,7755

TABLE 4.2 – Durée moyenne pour qu'un noeud soit infecté selon le type de réseau - $\mathbb{E}[T]$

Le réseau *maillé* est celui dont la probabilité d'infection est la plus élevée. Tous les noeuds de ce réseau étant connectés entre eux, la durée moyenne pour qu'ils deviennent infectés est la plus rapide ($\mathbb{E}[T_{maille}] = 0,1538$). Le degré de liberté de chaque noeud est de 4, soit le maximum. A l'inverse, les réseaux de type *anneau* et *linéaire* ont une probabilité beaucoup plus faible, puisque le degré de liberté de chaque noeud est inférieur ou égal à 2, d'où la durée beaucoup plus longue pour qu'ils deviennent infectés. Pour le réseau de type *anneau*, la probabilité d'infection est identique à tous les noeuds contrairement au réseau *linéaire* qui voit sa probabilité d'infection augmenter de manière pyramidale, avec une probabilité faible pour les noeuds aux extrémités de la chaîne et une probabilité élevée pour celui qui est au milieu. Pour le réseau de type *étoile*, tous les noeuds ont une seule connexion, à part le noeud 3 qui est connecté avec tous les autres noeuds, ce qui explique la probabilité d'infection élevée de ce noeud. Pour le dernier réseau, de type *arbre*, la probabilité d'infection peut-être très volatile selon la taille de l'arbre et sa composition.

4.2.3 Analyse selon les taux β, ϵ et δ

Pour cette analyse, on va faire varier les différents taux de passage (β, ϵ et δ) pour voir l'évolution de la probabilité d'infection. Pour cela, prenons un réseau de type *étoile* composé de 5 équipements informatiques et dont la matrice adjacente est la suivante :

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Les graphiques 4.3 et 4.4 compare l'évolution de la probabilité d'infection pour le noeud 3, celui qui a le plus de connexions, et les autres noeuds qui en ont seulement une. Pour le premier graphique, la probabilité d'infection augmente en fonction des taux de passage β et ϵ alors qu'elle diminue en fonction du taux δ dans le second graphique. Le noeud 3 voit clairement sa probabilité d'infection plus élevé que les autres noeuds.

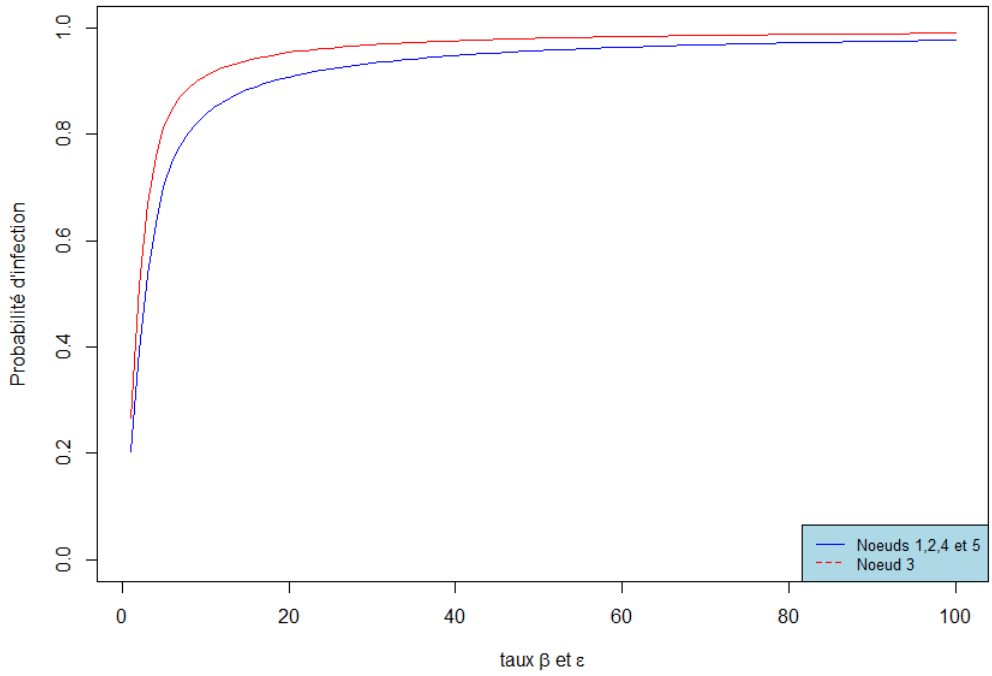

FIGURE 4.3 – Analyse des taux de passage β (menace intérieure) et ϵ (menace extérieure)

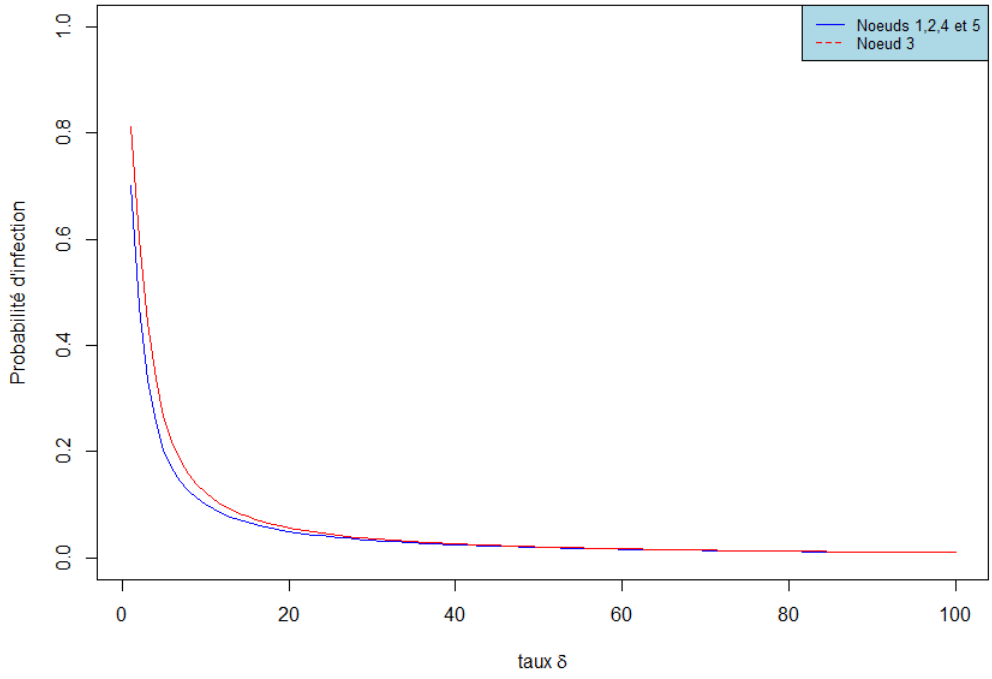

FIGURE 4.4 – Analyse du taux de passage δ (réparation)

4.2.4 Analyse selon la taille du réseau informatique

On va maintenant faire varier le nombre d'équipement informatique pour voir l'impact sur la probabilité d'infection. Prenons encore le cas du réseau *étoilé* où le noeud central (celui qui est connecté à tous les autres noeuds) correspondrait au serveur et les autres noeuds à des ordinateurs. Le graphique 4.5 montre que plus le réseau est grand et plus la probabilité d'infection est élevée. Cela s'explique par le fait qu'un noeud sain va être infecté beaucoup plus rapidement dans un réseau de grande taille que dans un réseau plus petit, alors que la durée d'infection reste identique quelle que soit la taille du réseau. Comme pour l'analyse précédente, la probabilité d'infection pour le serveur est plus élevée que pour les ordinateurs, ceci étant dû au nombre de connexions.

Nombre d'équipement	5	7	9	11	21
$E[T_{Ordinateurs}]$	1,4648	1,3669	1,2711	1,1772	0,7349
$E[T_{Serveur}]$	0,9833	0,7563	0,5999	0,4856	0,1903

TABLE 4.3 – Durée moyenne pour qu'un ordinateur/serveur soit infecté selon la taille du réseau

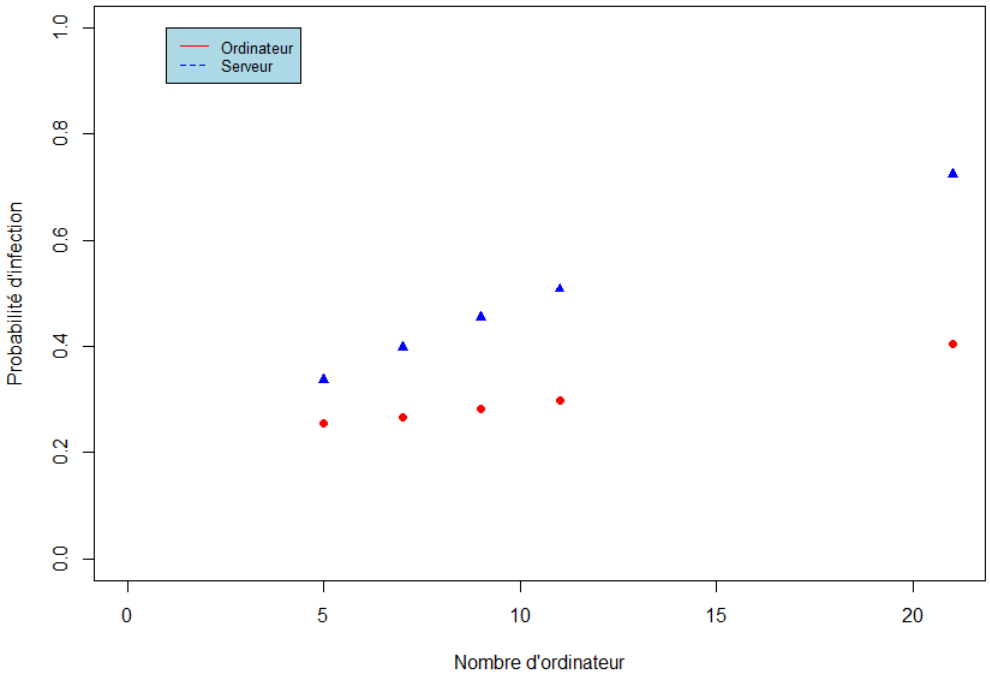

FIGURE 4.5 – Analyse selon la taille du réseau informatique, avec $\beta = \epsilon = 0.5$ et $\delta = 2$

4.3 Analyse du processus de sévérité d'une infection

Dans cette partie on analysera le processus de sévérité d'une infection. Après avoir défini les différents paramètres, on modélisera les coûts basés sur des simulations de type Monte Carlo.

4.3.1 Hypothèse

Dans le cadre du processus de sévérité d'une infection, on avait déterminé au chapitre précédent deux types de pertes : celles causées par l'infection elle-même et celles causées par la restauration de l'équipement informatique à un état fonctionnel. À partir de la formule (3.2) sur les pertes totales cumulées d'un réseau informatique, on obtient les résultats suivants :

$$S(t) = \sum_{v=1}^{N} s_v(t) = \sum_{v=1}^{N} \sum_{i=1}^{M_v(t)} [\eta_v(L_{v,i}) + C_v(R_{v,i})]$$

Où $\eta_v(L_{v,i})$ correspond à la modélisation des pertes causées par l'infection, $C_v(R_{v,i})$ à la modélisation des pertes causées par la restauration et $M_v(t)$ au nombre total d'infection pendant une période t.

Les pertes causées par l'infection

Les pertes causées par l'infection correspondent aux vols de données à caractère personnel ou aux dommages des données. Ces pertes sont modélisées par un coût aléatoire, tel que :

$$\eta_v(L_{v,i}) = c_1 L_{v,i} \tag{4.2}$$

Où c_1 correspond au taux lié à la pertes d'information et $(L_{v,i})$ correspond à la pertes de données. Pour modéliser (L_v), on utilisera la loi Bêta. En assurance Non-vie, cette distribution permet de modéliser de manière très flexible - grâce à ses deux paramètres de forme a et b - les taux de dommages qui correspondent au rapport entre le montant du sinistre et les sommes assurées du contrat concerné. Dans notre cas, ces taux de dommages correspondent à la probabilité d'infection.

$$f_{L_v}(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

Où a et b sont des paramètres de forme et B est la fonction Bêta.

Les pertes causées par la restauration

Les dépenses causées par la réparation de l'équipement sont de deux ordres :

- les pertes fixes comme les dommages des équipements informatique ;
- les pertes variables qui dépendent de la durée de réparation ou de remplacement des équipements comme la perte de revenu ou l'interruption d'activité ;

Ces pertes sont modélisées par la fonction suivante :

$$C_v(R_v) = c_2 + c_3 R_v$$

Où c_2 représente le taux pour les pertes fixes, c_3 le taux pour les pertes variables et (R_v) correspond à la durée de hors-service ou de restauration.

4.3.2 Simulation et analyse des coûts

Dans cette partie on va mettre en place un outil basé sur des simulations de type Monte Carlo. On déterminera le nombre d'infection sur une année ainsi que les coûts associés. Prenons le cas du réseau *étoilé*, composé de 4 ordinateurs et un serveur, avec les taux de passage suivants : $\beta = \epsilon = 0,1$ et $\delta = 0,5$. Supposons que les taux pour les pertes causés par l'infection et la restauration des équipements informatiques sont les suivants : $c_1 = c_2 = c_3 = 0,1$. On obtient alors les résultats dans les tableaux 4.4 et 4.5.

Le nombre d'infection est plus élevé pour le serveur dû au nombre élevé de connexions. Son coût, de 33,45, est par conséquent plus important par rapport aux ordinateurs qui ont chacun une seule connexions. Cela rejoint l'analyse faite sur la probabilité d'infection selon le degré de liberté (tableau 3.1) où la probabilité d'infection était plus élevé pour le serveur que les ordinateurs.

Nombre d'infection	Moyenne	écart-type	Minimum	Maximum
Ordinateur 1	36,58	5,67	25	53
Ordinateur 2	36,31	5,81	24	52
Ordinateur 3	36,75	5,05	25	53
Ordinateur 4	36,42	6,22	21	50
Serveur	47,33	6,74	34	67

TABLE 4.4 – Simulations du nombre d'infection avec la méthode Monte Carlo sur une année

Le tableau 4.6 récapitule les différents coûts pour un réseau selon son type. Comme pour l'analyse sur la propagation d'une infection selon le type de réseau, le coût suit la même tendance. Le coût est plus élevé pour le réseau *maillé* que les réseaux *étoilé*, *anneau* et *linéaire*.

Coût	Moyenne	écart-type	Minimum	Maximum
Ordinateur 1	28,38	6,34	12,41	53,84
Ordinateur 2	29,68	5,01	15,18	40,55
Ordinateur 3	28,30	5,70	15,09	44,19
Ordinateur 4	28,17	5,58	15,46	39,96
Serveur	33,45	5,75	16,47	49,58

TABLE 4.5 – Simulations du coût de l'infection avec la méthode Monte Carlo sur une année

Coût moyen	*étoilé*	*maillé*	*anneau*	*linéaire*	*arbre*
$\eta_v(L_{v,i})$	75,83	109,29	67,88	87,96	89,63
$C_v(R_{v,i})$	63,38	81,80	55,99	54,44	74,40
$S(t)$	147,63	191,73	155,41	151,70	160,78

TABLE 4.6 – Les différents coût selon le type de réseau

4.4 Conclusion

Dans cette partie on a analysé les processus de propagation et de sévérité d'une infection.

Dans le premier cas, on a fait l'hypothèse que le processus de propagation suivait une distribution exponentielle. Plusieurs facteurs ont été analysés, leurs impacts peuvent-être résumés dans le tableau ci-dessous :

Facteurs	Impact
Type de réseau	Plus le nombre de connexions pour un équipement informatique est élevé, plus la probabilité d'infection est élevée. Le nombre de connexions augmente la rapidité d'infection. Le réseau de type *maillé* aura une probabilité d'infection plus élevée que les autres types de réseau : *étoile*, *linéaire*, *anneau*.
Taille du réseau	Plus il y a d'équipements informatiques dans un réseau, plus la probabilité d'infection est élevée. Le nombre d'équipements augmente la rapidité d'infection.
Taux de passage β et ϵ	Plus les taux de passage sont élevés, plus la probabilité d'infection croît de manière exponentielle.
Taux de passage δ	Plus le taux de passage est élevé, plus la probabilité d'infection diminue.

TABLE 4.7 – Les facteurs impactant la probabilité d'infection

Dans le deuxième cas, on a fait l'hypothèse que la sévérité de l'infection engendrait plusieurs coûts :

- les coûts liés à l'infection qui correspond à la perte d'information. Ces coûts sont modélisés par la distribution Bêta.
- les coûts liés à la restauration de l'équipement. Ces coûts comprennent les pertes fixes

comme les dommages des équipements informatique et les pertes variables qui dépendent de la durée de réparation ou de remplacement des équipements, noté R_v.

On a simulé, à partir de la méthode de Monte Carlo, le nombre d'infection sur une année complète ainsi que les coûts associés. Les résultats obtenus sont au niveau micro, il s'agit d'un coût pour chaque équipement informatique. Ces coûts suivent la même tendance que l'analyse faite sur la propagation d'une infection. Le coût augmente selon le nombre de connexions. Un réseau de type *maillé* sera plus cher que celui de type *étoilé*.

A partir de ces analyses, on estimera dans la partie suivante les différents paramètres du modèle selon la structure tarifaire d'un contrat d'assurance Cyber que l'on définira.

Chapitre 5

Application numérique sur le portefeuille de SMACL Assurances

5.1 Objectif

La méthode qu'on utilise, basée sur le modèle pandémique, ne permet pas d'obtenir un tarif reflétant la sinistralité moyenne, contrairement aux méthodes statistiques standards qui se basent, elles, sur les données disponibles. Comme on l'a énoncé au chapitre trois, la quantité et la granularité des données actuellement disponibles, que ce soit au niveau de notre portefeuille ou même au niveau marché, ne permettent pas l'application de ces méthodes statistiques. Notre méthode modélise le processus de propagation d'une infection et sa sévérité, mais pour l'appliquer à un cas concret plusieurs paramètres doivent être estimés. Ils peuvent être estimés de différentes manières, soit à partir de données internes soit externes. Dans notre cas, les paramètres liés à la composition du réseau informatique et aux taux de passage seront estimés à partir de données externes, que l'on peut trouver soit sur internet soit dans des études faites par des spécialistes dans le domaine de la cyber-criminalité. Pour les paramètres liés aux pertes, ils seront estimés dans un premier temps, de manière à obtenir un tarif qui soit cohérent à la fois :

- par rapport aux contraintes techniques de SMACL Assurances. Le tarif doit être suffisamment élevé pour pouvoir être solvable en cas de sinistre. Sachant que le coût moyen d'une cyber-attaque se chiffre en millions d'euros, SMACL Assurances a donc limité son montant de garantie : il va de 50000 € pour les plus petites collectivités et jusqu'à 600000 € pour les plus importantes.
- et par rapport à ses contraintes commerciales. Le tarif ne doit pas être disproportionné par rapport au tarif des autres produits d'assurances proposés aux collectivités. SMACL Assurances juge que le tarif doit être compris entre 5% et 10% du tarif moyen d'un contrat d'assurances en dommages aux biens et en responsabilité civile.

L'objectif est de proposer une nouvelle structure tarifaire qui prendra en compte des critères qui ne faisaient pas partie de la tarification jusqu'à présent. Ainsi les résultats de la modélisation seront utilisés pour différencier nos sociétaires et non pour calculer une prime pure. A l'avenir, les paramètres de ce modèle pourront s'ajuster en fonction des nouvelles informations que l'on obtiendra.

Le tableau 5.1 propose une échelle de prix cible souhaité par SMACL Assurances, définie selon le type de collectivité :

Type de collectivité	Tarif moyen	Tarif cible	
		Borne min (5%)	Borne max (10%)
Communes de moins de 2000 habitants	1 649	82	165
Communes de 2000 à 4999 habitants	4 996	250	500
Communes de 5000 à 9999 habitants	8 341	417	834
Communes de 10000 à 19999 habitants	21 457	1 073	2 146
Communes de 20000 à 49999 habitants	33 087	1 654	3 309
Communes de 50000 à 79999 habitants	62 269	3 113	6 227
Communes de 80000 à 100000 habitants	74 188	3 709	7 419
Communes de plus de 100000 habitants	90 176	4 509	9 018
Communautés urbaines / métropoles	55 325	2 766	5 532
Communautés d'agglomération	24 180	1 209	2 418
Communautés de communes	5 561	278	556
Départements	83 879	4 194	8 388
Régions	148 775	7 439	14 877

TABLE 5.1 – Tarif cible pour un contrat d'assurance Cyber

5.2 Structure tarifaire

Dans cette partie on proposera une structure tarifaire pour un produit d'assurance Cyber. Précédemment, plusieurs paramètres influençant sur la propagation d'une infection ou sa sévérité ont été définis. Il s'agit :

- du type et de la taille du réseau,
- de la vitesse de réparation et de propagation de l'infection, représentées par les taux de passage δ, β et ϵ,
- des taux liés aux pertes causées par l'infection et la restauration.

Ces paramètres vont être estimés à partir des analyses faites dans la partie précédente et du tarif cible attendu par SMACL Assurances.

5.2.1 Le type et la taille du réseau informatique

Il est actuellement difficile de connaître précisément la composition d'un réseau informatique dans une collectivité. C'est lors de l'établissement du contrat d'assurance que l'assuré renseignera précisément sur la composition exacte de son réseau informatique. Mais dans un premier temps, on supposera que le type de réseau sera celui « étoilé » car c'est celui le plus courant, où tous les ordinateurs sont reliés à un seul serveur. Concernant la taille du réseau, on peut l'estimer à

partir du nombre d'agent travaillant dans les collectivités. Sachant qu'un agent ne travaille pas systématiquement sur un ordinateur, on fait l'hypothèse que seuls les agents des catégories A et B en sont équipés. La catégorie C, quant à elle, correspond plus à une fonction d'exécution ou de terrain, qui n'a pas vocation à travailler devant un ordinateur, comme par exemple un agent d'entretien ou un surveillant de prison (etc.). Pour obtenir cette information on se base sur les rapports de la fonction publique territoriale *. Ces rapports fournissent plusieurs statistiques comme le nombre de collectivités existantes et le nombre d'agent selon le type de collectivité (voir annexe C). A partir de ces informations on peut en déduire le nombre moyen d'agent par collectivité et la proportion d'entre eux travaillant sur un ordinateur. On obtient les résultats dans le tableau suivants (Table 5.2) :

Type de collectivité	Nombre moyen d'agents	Ratio ordinateur par agent	Taille estimée	Taille retenue
Communes de moins de 2000 habitants	4	17,4%	1	**5**
Communes de 2000 à 4999 habitants	37	17,4%	6	**10**
Communes de 5000 à 9999 habitants	110	17,4%	19	19
Communes de 10000 à 19999 habitants	427	17,4%	46	46
Communes de 20000 à 49999 habitants	678	17,4%	118	118
Communes de 50000 à 79999 habitants	1 465	17,4%	255	255
Communes de 80000 à 100000 habitants	2 168	17,4%	377	377
Communes de plus de 100000 habitants	4 617	17,4%	803	803
Communautés urbaines / métropoles	1 835	30,7%	563	563
Communautés d'agglomération	390	30,7%	120	120
Communautés de communes	81	30,7%	25	25
Départements	2 975	36,8%	1 095	1 095
Régions	5 849	19,4%	1 135	1 135

TABLE 5.2 – Nombre d'ordinateurs retenus par type de collectivité

Le nombre d'ordinateurs estimés paraît cohérent selon le type et la taille de la commune, mis à part pour les plus petites d'entre elles, inférieures à 5000 habitants. Pour ces petites collectivités, le nombre d'ordinateurs paraît beaucoup trop faible par rapport à la réalité. On prendra donc les valeurs suivantes : 5 ordinateurs pour les communes de moins de 2000 habitants et 10 pour celles de 2000 à 5000 habitants.

5.2.2 La vitesse de réparation et de propagation de l'infection

La vitesse de réparation et de propagation de l'infection, représentée par les taux de passage δ, β et ϵ, vont influer sur le nombre d'infection dans un réseau informatique. Pour estimer ces paramètres, on se base sur l'étude de l'institut Ponemon et d'Accenture, publié en 2019, qui montrent qu'une entreprise subit, en moyenne, 145 violations de sécurité par an [15]. A partir de cette information et des analyses faites dans le chapitre précédent, on peut en déduire les

∗. https ://www.collectivites-locales.gouv.fr

différents taux de passage.

Mais avant cela, on va d'abord implémenter un critère qui viendra majorer ou minorer le tarif en fonction de la vitesse de propagation. Ce critère correspondra au niveau de sécurité informatique de la collectivité. Plus la collectivité aura un réseau informatique sécurisé, plus la vitesse de propagation sera faible et donc un nombre d'infection bas, d'où une minoration de son tarif. Cinq niveaux de sécurité sont définis, où le niveau 1 correspond au niveau de sécurité le plus faible. Les différents niveaux peuvent être décrits dans le tableau 5.3.

Niveau de sécurité	Description
Niveau 1	Protection des ordinateurs (anti-virus), habilitation restreintes
Niveau 2	Protection des serveurs (anti-virus)
Niveau 3	Protection des réseaux (pare-feu internet, anti-spam intelligent)
Niveau 4	Gestion régulières de sauvegarde
Niveau 5	Séparation des réseaux / plusieurs fournisseurs internet

TABLE 5.3 – Description des différents niveaux de sécurités

Le tableau 5.4 montre les valeurs retenues pour les taux passage β et ϵ selon le niveau de sécurité. Ainsi, plus ils sont élevés, plus la probabilité d'infection augmente et le tarif de la collectivité sera élevé. Concernant le taux de passage δ, on considère qu'il ne dépend d'aucun critère.

Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
β et ϵ	0,38	0,34	0,30	0,26	0,22
δ	1				

TABLE 5.4 – Taux de passage β, ϵ et δ selon le niveau de sécurité

A partir de ces valeurs, on obtient le nombre moyen d'infection par ordinateur/serveur dans le cas d'un réseau « étoilé » avec 10 ordinateurs (tableau 5.5). Pour le niveau 1, le niveau le moins sécurisé, le nombre est de 146, soit le nombre moyen de violations d'après l'étude de l'institut Ponemon et d'Accenture.

Nombre d'infection	Moyenne	écart-type	Minimum	Maximum
Niveau 1	146	11	118	178
Niveau 2	132	10	109	158
Niveau 3	116	9	92	141
Niveau 4	100	9	80	125
Niveau 5	84	8	64	106

TABLE 5.5 – Nombre d'infections selon le niveau de sécurité

5.2.3 Les taux liés aux pertes causées par l'infection et la restauration

Comme il a été précisé dans le chapitre précédent, les pertes causées par l'infection correspondent à la perte d'information (c_1) que l'on modélise à partir de la loi Bêta. Les pertes causées

par la restauration correspondent aux dommages des équipements informatiques (pertes fixes - c_2) et aux pertes de revenu ou liées à l'interruption d'activité (pertes variables - c_3). Les pertes variables dépendent du taux de passage δ qui correspond à la vitesse de réparation.

Comme pour la vitesse de propagation de l'infection, on va implémenter un critère tarifant qui dépendra, cette fois-ci, du taux lié à la perte d'information et qui viendra majorer ou minorer le tarif. Ce critère correspondra au niveau de sensibilité des données. Plus les données traitées par la collectivité seront sensibles, plus le taux lié à la perte d'information sera élevé, d'où une majoration de son tarif. Cinq niveaux seront également définis, où le niveau 1 correspond au niveau de sensibilité le plus faible. Les différents niveaux peuvent être définis selon les activités contrôlées par le système informatique (SI) de la collectivité :

Niveau de sensibilité	Description
Niveau 1	Urbanisme, foyers des personnes âgées
Niveau 2	Gestion des établissements scolaires et des crèches
Niveau 3	Gestion des prestations sociales et du personnel
Niveau 4	Gestion des listes électorales,
Niveau 5	Comptabilité, marchés publics, état civil

TABLE 5.6 – Description des différents niveaux de sensibilités selon l'activité exercée par le SI de la collectivité

Si le SI d'une collectivité exerce plusieurs activités, c'est la plus risquée qui sera retenue dans la tarification.

Comme il a été précisé dans la partie 5.1, on va estimer ces paramètres de façon à établir un tarif cohérent tenant compte à la fois de la solvabilité de SMACL Assurances et de ses attentes commerciales.

On va cependant se baser encore une fois sur l'étude de Ponemon et d'Accenture [15] de façon à obtenir un tarif pour chaque type de perte qui soit représentatif de la réalité. Pour cela, on reprend le tableau 1.1 du chapitre 1 sur les différents coûts que peut engendrer une cyber-attaque et on utilise la même répartition qui est faite entre ces types de pertes pour estimer le tarif. On obtient la répartition suivante :

Type de perte	Taux associé	Coût	Proportion
Perte d'information	c_1	5.9M$	45%
Dommage équipement	c_2	0.5M$	4%
Perte de revenue Interruption de l'activité	c_3	2.6M$ 4.0M$	51%
Total	c_{Total}	13.0M$	100%

TABLE 5.7 – Proportion des types de pertes calculée à partir de l'étude de Ponemon et d'Accenture (c.f. tableau 1.1)

A partir du tarif cible et de la répartition faite dans le tableau 5.7, on obtient un tarif cible pour chaque type de perte.

Type de collectivité	Tarif cible - c_1		Tarif cible - c_2		Tarif cible - c_3	
	min	max	min	max	min	max
Communes de moins de 2000 habitants	37	75	3	7	42	84
Communes de 2000 / 4999 habitants	113	227	10	20	127	254
Communes de 5000 / 9999 habitants	189	379	17	33	212	423
Communes de 10000 / 49999 habitants	487	974	43	86	545	1 089
Communes de 20000 à 49999 habitants	751	1 502	66	132	840	1 680
Communes de 50000 à 79999 habitants	1 413	2 826	125	249	1 581	3 161
Communes de 80000 à 100000 habitants	1 683	3 367	148	297	1 883	3 766
Communes de plus de 100000 habitants	2 046	4 093	180	361	2 289	4 578
Communautés urbaines / métropoles	1 255	2 511	111	221	1 404	2 809
Communautés d'agglomération	549	1 097	48	97	614	1 228
Communautés de communes	126	252	11	22	141	282
Départements	1 903	3 807	168	336	2 129	4 258
Régions	3 376	6 752	298	595	3 777	7 553

TABLE 5.8 – Tarif cible pour chaque type de perte

Une fois le tarif cible défini par type de perte, on peut estimer les différents taux.

$$
\begin{aligned}
c_{1,estimé} &= \frac{TC_1}{N_{Infection} \times N_{Ordinateur} \times \mathbb{E}[L_{v,i}]} \\
c_{2,estimé} &= \frac{TC_2}{N_{Infection} \times N_{Ordinateur}} \\
c_{3,estimé} &= \frac{TC_3}{N_{Infection} \times N_{Ordinateur} \times \mathbb{E}[R_v]}
\end{aligned}
$$

Où

- TC correspond au tarif cible selon le type de perte,
- $N_{Infection}$ correspond au nombre d'infections,
- $N_{Ordinateur}$ correspond au nombre d'ordinateurs,
- $\mathbb{E}[L_{v,i}]$ correspond à la sévérité moyenne lorsqu'un ordinateur est infecté (voir équation (4.2)),
- $\mathbb{E}[R_v]$ correspond à la durée moyenne pour qu'un ordinateur soit réparé (voir équation (4.1)).

Le taux c_1 lié à la perte d'information est ainsi défini selon le niveau de sensibilité des données traitées par la collectivité. A partir du taux $c_{1,estimé}$, on décline ce taux pour chaque niveau de

sensibilité. Les taux c_2 et c_3 liés respectivement aux dommages des équipements informatiques et à la perte de revenu ne dépendent, eux, d'aucun critère. Les tableaux 5.9 et 5.10 fournissent les estimations de ces taux.

Taux $c_{1,estimé}$	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Communes de moins de 2000 habitants	0.267	0.278	0.288	0.299	0.310
Communes de 2000 / 4999 habitants	0.405	0.421	0.437	0.453	0.469
Communes de 5000 / 9999 habitants	0.354	0.368	0.382	0.396	0.410
Communes de 10000 / 49999 habitants	0.374	0.389	0.404	0.419	0.433
Communes de 20000 à 49999 habitants	0.227	0.236	0.245	0.254	0.263
Communes de 50000 à 79999 habitants	0.198	0.206	0.214	0.222	0.229
Communes de 80000 à 100000 habitants	0.159	0.166	0.172	0.178	0.185
Communes de plus de 100000 habitants	0.091	0.095	0.098	0.102	0.105
Communautés urbaines / métropoles	0.080	0.083	0.086	0.089	0.092
Communautés d'agglomération	0.163	0.170	0.176	0.183	0.189
Communautés de communes	0.180	0.187	0.195	0.202	0.209
Départements	0.062	0.065	0.067	0.069	0.072
Régions	0.106	0.110	0.115	0.119	0.123

TABLE 5.9 – Taux lié à la perte d'information selon la sensibilité des données (c_1)

Type de collectivité	Taux $c_{2,estimé}$	Taux $c_{3,estimé}$
Communes de moins de 2000 habitants	0.008	0.107
Communes de 2000 / 4999 habitants	0.013	0.162
Communes de 5000 / 9999 habitants	0.011	0.142
Communes de 10000 / 49999 habitants	0.012	0.150
Communes de 20000 à 49999 habitants	0.007	0.091
Communes de 50000 à 79999 habitants	0.006	0.079
Communes de 80000 à 100000 habitants	0.005	0.064
Communes de plus de 100000 habitants	0.003	0.036
Communautés urbaines / métropoles	0.003	0.032
Communautés d'agglomération	0.005	0.065
Communautés de communes	0.006	0.072
Départements	0.002	0.025
Régions	0.003	0.043

TABLE 5.10 – Taux lié aux dommages des équipements informatiques (c_2) et lié à la perte de revenu ou à l'interruption d'activité (c_3)

Dans la partie suivante, on déterminera un tarif à partir de ces paramètres et selon les critères définis.

5.3 Les tarifs d'un contrat d'assurance Cyber

A partir des estimations précédentes, on va pouvoir déterminer un tarif selon le type de la collectivité et selon les différents critères prédéfinis. Les tarifs ainsi obtenus prendront en compte les chargements liés au frais de SMACL Assurances, que l'on détaillera dans le paragraphe suivant.

5.3.1 Les chargements

La prime d'assurance que verse l'assuré pour se couvrir en cas de sinistre se compose de la prime pure (coût probable des sinistres), des chargements et des taxes. Concernant les chargements, SMACL Assurances en distingue 3 types :

- les frais généraux qui correspondent à tous les frais inhérents à la gestion des contrats, à leurs coûts d'acquisition (commissionnement du réseau de distribution) et à la gestion des sinistres ;
- le coût des traités de réassurance ;
- la marge qui correspond aux bénéfices attendus.

Ces chargements sont différents selon le type de produit (dommages aux biens, véhicules à moteurs, responsabilité civile...) et selon si l'offre est « standard » ou « sur mesure ». Pour l'offre « sur mesure », la durée du contrat est beaucoup plus longue, environ 4/5 ans, ce qui a tendance à diminuer ses frais d'acquisition, contrairement à l'offre « standard » où les contrats ont une durée limitée, de 1 à 3 ans environ.

5.3.2 La grille tarifaire selon les critères

Dans cette partie, on présentera seulement la grille tarifaire pour les communes entre 20000 et 50000 habitants (les autres grilles se trouvent en annexe D). Les grilles concernant les types de pertes y seront également détaillées : perte d'information, dommages aux équipements informatiques et perte de revenu ou interruption d'activité. Comme précisé précédemment, on se base sur un réseau de type *étoilé*. Le tarif d'un contrat d'assurance Cyber pour une collectivité dépendra donc de son niveau de sécurité informatique et du niveau de sensibilité de ses données.

Les tarifs obtenus correspondent bien à ceux attendus : ils sont compris dans la tranche du tarif cible souhaité par SMACL Assurances. Pour la collectivité en question, le tarif au global est compris entre 1657 € et 3058 €, ce qui correspond bien au tarif cible compris entre 1654 € et 3309 €. Ce constat peut se vérifier également pour les grilles de chaque type de perte et ce, quel que soit le type de collectivité.

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	2 865	2 578	2 269	1 960	1 657
Niveau 2	2 922	2 597	2 303	1 988	1 695
Niveau 3	2 962	2 633	2 326	2 025	1 722
Niveau 4	3 014	2 693	2 392	2 071	1 740
Niveau 5	3 058	2 749	2 425	2 091	1 772

TABLE 5.11 – Tarif total pour les communes entre 20000 et 50000 habitants avec 118 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	1 235	1 104	984	851	725
Niveau 2	1 298	1 150	1 004	886	748
Niveau 3	1 354	1 186	1 058	912	781
Niveau 4	1 400	1 229	1 106	968	806
Niveau 5	1 452	1 301	1 151	995	819

TABLE 5.12 – Tarif lié à la perte d'information (c_1) pour les communes entre 20000 et 50000 habitants

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	114	103	91	79	67
Niveau 2	114	103	89	78	66
Niveau 3	115	102	91	79	66
Niveau 4	115	102	92	79	66
Niveau 5	116	104	92	78	65

TABLE 5.13 – Tarif lié aux dommages des équipements informatiques (c_2) pour les communes entre 20000 et 50000 habitants

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	1 491	1 336	1 190	1 017	873
Niveau 2	1 527	1 345	1 170	1 041	870
Niveau 3	1 509	1 356	1 184	1 033	838
Niveau 4	1 494	1 330	1 185	1 023	848
Niveau 5	1 514	1 349	1 207	1 029	850

TABLE 5.14 – Tarif lié à la perte de revenu/interruption d'activité (c_3) pour les communes entre 20000 et 50000 habitants

5.4 Conclusion

Dans cette partie, le modèle présenté au chapitre 4 est appliqué au portefeuille de SMACL Assurances. Pour cela, différents paramètres doivent être estimés : ceux liés à la composition du réseau et aux taux de passages sont estimés à partir de données externes qui ont été trouvées sur internet ou dans des études réalisées par des spécialistes dans le domaine de la cyber-criminalité. Pour les paramètres liés aux coûts, ils sont estimés dans un premier temps de manière à obtenir un tarif qui soit cohérent à la fois, par rapport à la solvabilité de SMACL Assurances et par rapport à ses objectifs commerciaux. Concernant le deuxième point, un tarif cible a donc été déterminé selon le type de collectivité. Il est compris entre 5% et 10% des tarifs moyens actuellement pratiqués pour les autres contrats d'assurances.

L'objectif principal de ce premier tarif Cyber est de définir les bases d'une nouvelle structure tarifaire qui prend en compte des critères qui n'existaient pas auparavant. A l'avenir, ces paramètres pourront s'ajuster en fonction des nouvelles informations qui seront obtenues.

Les estimations faites sur les différents paramètres sont résumées dans le tableau 5.15. Concernant les taux liés aux pertes, ils sont estimés à partir des tarifs cibles prédéfinis et selon la répartition des coûts moyen à partir de l'étude Ponemon et d'Accenture.

Paramètres	Estimation
Taille du réseau	estimation du nombre d'ordinateurs dans un réseau à partir du nombre moyen d'agent par collectivité et des catégories A et B.
Taux de passage	estimation des taux de passage δ, β et ϵ à partir du nombre moyen d'infection par an, fournit par l'étude de l'institut Ponemon et d'Accenture.
Taux liés aux coûts	estimation des taux c_1, c_2 et c_3 selon le tarif souhaité par SMACL Assurances. La répartition entre ces taux est déduite à partir de la répartition faites dans l'étude de l'institut Ponemon et d'Accenture sur les différents coûts que peuvent engendrer une cyber-attaque.

TABLE 5.15 – Récapitulatif des estimations faites pour chaque paramètre

Les formules ci-dessous permettent d'estimer les taux liés aux coûts selon le type de perte :

$$c_{1,estimé} = \frac{TC_1}{N_{Infection} \times N_{Ordinateur} \times \mathbb{E}[L_{v,i}]}$$
$$c_{2,estimé} = \frac{TC_2}{N_{Infection} \times N_{Ordinateur}}$$
$$c_{3,estimé} = \frac{TC_3}{N_{Infection} \times N_{Ordinateur} \times \mathbb{E}[R_v]}$$

Les tarifs obtenus à partir de ces estimations permettent d'obtenir une nouvelle structure tarifaire avec des tarifs correspondant à ceux attendus par SMACL Assurances.

Conclusion générale

L'objectif de ce mémoire était d'utiliser une nouvelle approche pour modéliser et tarifer le cyber-risque car les méthodes statistiques standards ne sont, actuellement, pas compatible avec la granularité et la quantité des données disponibles. Cette nouvelle approche consiste donc à modéliser le cyber-risque en s'inspirant de la modélisation pandémique, car ces deux risques présentent des similitudes. Parmi les modèles pandémiques existants, celui sur lequel se base le cyber-risque est le modèle compartimental SIS. C'est un modèle simple avec seulement deux états, sain et infecté, et avec des taux de passages β et δ pour passer d'un état à un autre. Des chercheurs ont généralisé ce modèle afin de l'adapter aux particularités du cyber-risque et un autre taux de passage a été ajouté, le taux ϵ qui fait référence à une menace extérieure d'un réseau informatique.

Cette méthode présente plusieurs avantages. Le premier, et qui est d'ailleurs la raison pour laquelle beaucoup de recherches ont été initiées, permet de pallier le manque de données disponibles grâce à un jeu de simulations. On a utilisé la distribution exponentielle pour modéliser la propagation d'une infection d'un réseau informatique et la distribution Bêta pour sa sévérité (coûts liés à la perte d'information). Ensuite, à partir des simulations de la méthode Monte Carlo, on a pu déterminer le nombre d'infections sur une année et le tarif correspondant. Cette méthode utilise un processus stochastique pour décrire les effets dynamiques de propagation épidémique dans le temps.

L'autre avantage de cette méthode, est le fait de prendre en compte la composition d'un réseau informatique dans la tarification. La probabilité d'infection dépend de plusieurs facteurs comme la taille et le type de réseau informatique. Plus la taille d'un réseau est grande, avec plusieurs connexions entre chaque équipement informatique, plus la probabilité d'infection est élevée, parce que la propagation d'un virus informatique sera alors plus rapide. Cette analyse au niveau micro pour déterminer le tarif d'une assurance Cyber, peut s'assimiler à celle qui est faite en assurance automobile avec les caractéristiques du véhicule : sa puissance, le nombre de kilomètres parcourus...

Un troisième et dernier avantage concerne les coûts liés aux cyber-attaques. Cette nouvelle approche tarifaire permet non seulement de modéliser les processus d'infection et de restauration mais aussi les pertes associées. Les coûts peuvent inclure les pertes ou le vol des données mais aussi les dommages potentiels comme l'interruption de l'activité ou la perte de revenu.

Cette méthode a été appliquée sur le portefeuille de SMACL Assurances. Sa particularité a permis de définir une nouvelle structure tarifaire qui prend en compte plusieurs critères dont notamment le niveau de sécurité informatique et le niveau de sensibilité des données. Les paramètres du modèle ont été estimés à partir de données externes et selon les attentes de SMACL Assurances. Les paramètres ainsi définis pourront, à l'avenir, être ajustés à partir de nouvelles informations disponibles au niveau de notre portefeuille ou au niveau du marché.

Plusieurs axes doivent être étudiés afin d'affiner le modèle développé dans ce mémoire, en particulier :

- l'utilisation du modèle SIR dans le cadre du cyber-risque permet d'étudier la propagation d'une infection dans un réseau bien identifié, où l'on connaît la typologie et ses différents points de connexions. Or les collectivités connaissent peu la composition de leur réseau. Pour contourner cet obstacle nous avons fait plusieurs hypothèses : le type de réseau est celui « étoilé », c'est celui le plus courant où tous les ordinateurs sont reliés à un serveur. Le nombre d'ordinateur d'une collectivité est estimé selon le nombre d'agent des catégories A et B. La structure tarifaire est ainsi définie de sorte qu'une collectivité n'est pas nécessairement besoin de sa carte de réseau. Des échanges doivent avoir lieu avec les collectivités afin de fiabiliser ces hypothèses et d'éventuellement construire un modèle de tarification plus personnalisé.
- les paramètres sont estimés à partir de données externes qui proviennent d'études réalisées par des organismes privés et étrangers. Or ces données ne sont pas forcément représentatives des collectivités locales en France. Une base de données correspondant à la cible étudiée doit être constituée afin d'intégrer dans le modèle les caractéristiques propres aux collectivités.

Bibliographie

[1] Van Der Aalst. *A network approach to interrelated insurance risk.* PhD thesis, Université d'Amsterdam, 2018.

[2] Matthias A. Fahrenwaldt, Stefan Weber, and Kerstin Weske. Pricing of cyber insurance contracts in a network model. *Article publié par l'Université de Cambridge sur le site The Journal of the IAA*, 2018.

[3] Hugo Falconet and Antoine Jego. Rapport universitaire de l'École normale supérieure. *Modéliser la propagation d'une épidémie*, 2015.

[4] Alison L. Hill, David D. Rand, and Nicholas A. Christakis Martin A. Nowak. Emotions as infectious diseases in a large socialnetwork : the sis model. *Proc Royal Soc B 277 :3827–383*, 2010.

[5] Anne Le Hénanff, Didier Danet, and Gérard De Boisboissel. Etude pour le creogn : Données personnelles et collectivités territoriales : usages actuels et recommandations. *Proc Royal Soc B 277 :3827–383*, 2017.

[6] Laetitia Laguzet. *Modélisation mathématique et numérique des comportements sociaux en milieu incertains.* PhD thesis, Université Paris-Dauphine, 2015.

[7] Khouloud Mandhouj. Mémoire d'actuariat de l'ensae. *Analyse du Risque Catastrophe d'une Pandémie en Assurance Prévoyance par une Approche Épidémiologique*, 2010.

[8] Piet Van Mieghem. The n-intertwined sis epidemic network model. *Delft University of Technology*, 2011.

[9] Piet Van Mieghem and Eric Cator. Epidemics in networks with nodal self-infection and the epidemic threshold. *Physical review E86, 016116 - Delft University of Technology*, 2012.

[10] Piet Van Mieghem and R. Van de Bovenkamp. Accuracy criterion for the mean-field approximation in susceptible-infected-susceptible epidemics on networks. *Physical review E91, 032812 - Delft University of Technology*, 2015.

[11] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. Virus spread in networks. *IEEE/ACM Transactions on networking, Vol. 17, n°1 - Delft University of Technology*, 2009.

[12] Jasmina Omic and Piet Van Mieghem. Epidemic spreading in networks - variance of number of infected nodes. 2015.

[13] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Review of modern physics, Volume 87, n°3*, 2015.

[14] Manon PIEREN. Mémoire d'actuariat de l'euria. *Estimation et suivi du risque cyber*, 2017.

[15] Larry Ponemon and Kelly Bissel. *The cost of cybercrime.* Etude réalisé par l'intitut Ponemon et Accenture Security, 2019.

[16] Florian Pons. Mémoire d'actuariat du cnam. *Étude actuarielle du Cyber-risque*, 2013.

[17] Romain Speisser. Évaluation du risque de pandémie et construction de deux modèles internes partiels en assurance de personnes dans le cadre de solvabilité 2. *Mémoire d'actuariat - ESSEC*, 2013.

[18] Maochao Xu and Lei Hua. Cybersecurity insurance : Modeling and pricing. *Article publié au North American Actuarial Journal*, 2019.

[19] Maochao Xu and Shouhuai Xu. An extended stochastic model for quantitative security analysis of networked systems. *The University of Texas at San Antonio*, 2012.

Liste des tableaux

Table des figures

Annexes

Annexe A

Processus de propagation d'infection

A.1 Distribution Weibull

$$\bar{F}(x) = e^{(-\beta x)^{\alpha_1}}$$

et

$$\bar{G}(x) = e^{(-\epsilon_v x)^{\alpha_2}}$$

où β et ϵ_v sont les paramètres d'échelle, et $alpha_1$ et $alpha_1$ sont les paramètres de forme.

$$\begin{aligned} E[T_v^*] &= \int_0^\infty e^{-[(\epsilon_v x)^{\alpha_2} + (\beta x)^{\alpha_1} \sum_{j=1}^N a_{vj} p_j^*]} dx \\ &= \int_0^\infty e^{-[\epsilon_v^{\alpha_2} x^{\alpha_2} + \beta^{\alpha_1} x^{\alpha_1} \sum_{j=1}^N a_{vj} p_j^*]} dx \\ &= \phi(_v, \beta, \alpha_1, \alpha_2, P^*) \end{aligned}$$

Notons que si $\alpha_1 = \alpha_2 = \alpha$, alors on obtient :

$$\phi(_v, \beta, \alpha_1, \alpha_2, P^*) = \frac{1}{\left[\beta^\alpha \sum_{j=1}^N a_{vj} p_j^* + \epsilon_v^\alpha\right]^{1/\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)$$

Si on assume par conséquent que le processus de réparation suit également une loi de Weibull, alors

$$\bar{S}(x) = e^{-(\delta_v x)^{\alpha_3}}$$

alors

$$\mathbb{E}[R_v] = \frac{1}{\delta_v} \Gamma\left(1 + \frac{1}{\alpha_3}\right)$$

Alors, la probabilité d'infection peut alors s'écrire de la manière suivante :

$$p_v^* = \frac{\Gamma\left(1+\frac{1}{\alpha_3}\right)}{\Gamma\left(1+\frac{1}{\alpha_3}\right) + \phi(v, \beta, \alpha_1, \alpha_2, P^*)}$$

On peut constater que si $\epsilon = 0$, la formule correspond à la formule quand l'infection se stabilise.

A.2 Distribution log-normal

La fonction de densité de $Y_v j$ peut s'écrire de la manière suivante :

$$f_{vj}(x) = \frac{1}{x\sigma_1\sqrt{2\pi}} \exp\left[- \frac{\ln x - \mu_1}{2\sigma_1^2}\right]$$

Et la densité Z_v s'écrit :

$$g(x) = \frac{1}{x\sigma_2\sqrt{2\pi}} \exp\left[- \frac{\ln x - \mu_2}{2\sigma_2^2}\right]$$

Par conséquent, on a,

$$\begin{aligned}\mathbb{E}[T_v^*] =& \int_0^\infty \left[1 - \Phi\left(\frac{\ln x - \mu_2}{\sigma_2}\right)\right]\left[1 - \Phi\left(\frac{\ln x - \mu_1}{\sigma_1}\right)\right] \\ =& \Psi(\mu_1, \mu_2, \sigma_1, \sigma_2, p^*)\end{aligned}$$

Si on assume par conséquent que le processus de réparation suit également une loi de log-normal, on a :

$$S_v(x) = \Phi\left(\frac{\ln x - \mu_v}{2\sigma_v^2}\right)$$

Alors, on obtient le temps d'infection suivant :

$$\mathbb{E}[R_v] = \exp\left(\mu_v + \sigma_v^2/2\right)$$

On en déduit donc la probabilité d'infection suivante :

$$p_v^* = \frac{\exp\left(\mu_v + \sigma_v^2/2\right)}{\exp\left(\mu_v + \sigma_v^2/2\right) + \Psi(\mu_1, \mu_2, \sigma_1, \sigma_2, p^*)}$$

Annexe B

Matrice adjacente selon le type de réseau

$$A_{étoile} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} ; A_{maillé} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} ; A_{anneau} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$A_{linéaire} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} ; A_{arbre} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

Annexe C

Nombre d'agent par collectivité et par catégorie

Type de collectivité	Nombre d'agent	Nombre de collectivité	Nombre moyen d'agent
Communes de moins de 2000 habitants	133 543	30 105	4
Communes de 2000 à 4999 habitants	116 550	3 128	37
Communes de 5000 à 9999 habitants	124 823	1 138	110
Communes de 10000 à 49999 habitants	341 836	801	427
Communes de 50000 à 100000 habitants	120 406	75	1 605
Communes de plus de 100000 habitants	184 697	40	4 617
Communautés urbaines / métropoles	53 203	29	1 835
Communautés d'agglomération	85 849	219	390
Communautés de communes	82 965	1 018	81
Départements	291 512	98	2 975
Régions	81 885	14	5 849

TABLE C.1 – Nombre moyen d'agent par type de collectivité

Type de collectivité	catégorie A	catégorie B	catégorie C	Total	Proportion A+B
Communes	77,4	131,4	999,1	1 199,9	17,4%
Organismes inter-communaux	41,8	54,1	216,3	312,2	30,7%
Départements	49,9	81,1	225,4	356,4	36,8%
Régions	10,7	6,2	70,2	87,1	19,4%

TABLE C.2 – Nombre d'agents selon la catégorie et le type de collectivité (en milliers)

Annexe D

Grille tarifaire selon le type de collectivité

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	155	139	122	106	89
Niveau 2	158	141	124	108	90
Niveau 3	162	143	126	109	92
Niveau 4	163	146	129	112	93
Niveau 5	166	148	130	113	95

TABLE D.1 – Tarif total pour les communes de moins de 2000 habitants avec 5 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	459	412	363	312	268
Niveau 2	469	419	369	320	269
Niveau 3	479	425	375	326	274
Niveau 4	484	433	382	333	280
Niveau 5	494	438	389	337	282

TABLE D.2 – Tarif total pour les communes entre 2000 et 5000 habitants avec 10 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	742	668	598	510	429
Niveau 2	762	672	603	519	438
Niveau 3	779	694	605	518	446
Niveau 4	792	699	616	532	451
Niveau 5	796	711	628	546	456

TABLE D.3 – Tarif total pour les communes entre 5000 et 10000 habitants avec 19 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	1 865	1 690	1 466	1 289	1 077
Niveau 2	1 887	1 693	1 496	1 302	1 103
Niveau 3	1 932	1 724	1 523	1 323	1 125
Niveau 4	1 966	1 754	1 550	1 343	1 138
Niveau 5	1 993	1 784	1 576	1 359	1 146

TABLE D.4 – Tarif total pour les communes entre 10000 et 20000 habitants avec 46 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	2 865	2 578	2 269	1 960	1 657
Niveau 2	2 922	2 597	2 303	1 988	1 695
Niveau 3	2 962	2 633	2 326	2 025	1 722
Niveau 4	3 014	2 693	2 392	2 071	1 740
Niveau 5	3 058	2 749	2 425	2 091	1 772

TABLE D.5 – Tarif total pour les communes entre 20000 et 50000 habitants avec 118 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	5 355	4 796	4 241	3 649	3 103
Niveau 2	5 440	4 889	4 312	3 725	3 162
Niveau 3	5 564	4 943	4 375	3 803	3 215
Niveau 4	5 628	5 034	4 465	3 851	3 256
Niveau 5	5 726	5 122	4 502	3 921	3 320

TABLE D.6 – Tarif total pour les communes entre 50000 et 80000 habitants avec 255 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	6 394	5 724	5 057	4 364	3 711
Niveau 2	6 528	5 844	5 145	4 456	3 774
Niveau 3	6 622	5 919	5 228	4 533	3 832
Niveau 4	6 716	6 011	5 316	4 606	3 897
Niveau 5	6 858	6 128	5 399	4 684	3 968

TABLE D.7 – Tarif total pour les communes entre 80000 et 100000 habitants avec 377 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	7 850	7 027	6 209	5 358	4 556
Niveau 2	8 015	7 175	6 317	5 471	4 633
Niveau 3	8 130	7 267	6 418	5 566	4 705
Niveau 4	8 245	7 380	6 526	5 655	4 784
Niveau 5	8 420	7 524	6 629	5 751	4 871

TABLE D.8 – Tarif total pour les communes de plus de 100000 habitants avec 803 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	4 816	4 312	3 809	3 287	2 795
Niveau 2	4 917	4 402	3 876	3 356	2 843
Niveau 3	4 988	4 458	3 938	3 415	2 886
Niveau 4	5 059	4 528	4 004	3 470	2 935
Niveau 5	5 166	4 616	4 067	3 528	2 989

TABLE D.9 – Tarif total pour les communautés urbaines / métropoles avec 563 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	2 095	1 875	1 657	1 430	1 216
Niveau 2	2 138	1 914	1 686	1 460	1 236
Niveau 3	2 169	1 939	1 713	1 485	1 255
Niveau 4	2 200	1 969	1 741	1 509	1 277
Niveau 5	2 247	2 007	1 769	1 534	1 300

TABLE D.10 – Tarif total pour les communautés d'agglomération avec 120 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	498	447	391	337	289
Niveau 2	506	452	396	347	287
Niveau 3	517	454	401	353	295
Niveau 4	519	468	415	356	304
Niveau 5	527	472	422	361	310

TABLE D.11 – Tarif total pour les communautés de communes avec 25 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	7 266	6 504	5 747	4 959	4 217
Niveau 2	7 418	6 641	5 847	5 063	4 288
Niveau 3	7 525	6 726	5 941	5 151	4 355
Niveau 4	7 632	6 831	6 041	5 234	4 428
Niveau 5	7 793	6 964	6 135	5 323	4 509

TABLE D.12 – Tarif total pour les départements avec 1095 ordinateurs

Sensibilité \ Sécurité	Niveau 1	Niveau 2	Niveau 3	Niveau 4	Niveau 5
Niveau 1	12 920	11 565	10 219	8 818	7 498
Niveau 2	13 190	11 808	10 397	9 003	7 625
Niveau 3	13 380	11 959	10 563	9 159	7 743
Niveau 4	13 570	12 146	10 741	9 307	7 874
Niveau 5	13 858	12 382	10 909	9 464	8 017

TABLE D.13 – Tarif total pour les régions avec 1135 ordinateurs