DET
RAN
OTE

DETRA NOTE 2021-4

MODELING ACCUMULATION SCENARIOS IN CYBER RISK

By Louise d’Oultremont, Olivier Lopez
and Brieuc Spoorenberg

Detralytics

DISCLAIMER

The content of the Detra Notes for a pedagogical use only. Each business case is so specific that a careful analysis of the situation is needed before implementing a possible solution. Therefore, Detralytics does not accept any liability for any commercial use of the present document. Of course, the entire team remain available if the techniques presented in this Detra Note required your attention.

Detralytics
Rue Belliard/Belliardstraat 2
1040 Brussels
www.detralytics.com
info@detralytics.eu

ABSTRACT

In this work, we consider a simple way to model accumulation episodes (i.e., large number of claims occurring in a short amount of time) in the context of cyber risk. This provides a generic way to design stochastic scenarios measuring the impact of such an event on an insurance portfolio. A particular attention is devoted to providing a way to model network effects (contagion between actors connected with each other) while producing a model needing relatively few data to be calibrated. Illustrations of this method are performed to determine the impact of different behaviors of the policyholders and of the insurance company on the spread of the cyber epidemic.

Keywords: Cyber insurance; accumulation scenarios; epidemiological models; multi-group SIR models.

Short title: Accumulation scenarios in cyber insurance.

1 Introduction

Cyber-risk is the corollary of the dependence of the society to digital tools. With the increase of cyber-attacks (see Lallie et al. (2021), the report from Agence Nationale de la Sécurité des Systèmes d'Information (2021)), the development of insurance products appears as a promising solution to enhance the resilience of the economy. However, the uncertainty linked to this new risk is important, see for example Eling and Loperfido (2017) or Farkas et al. (2021). A major concern is the systemic nature of the risk (see Welburn and Strong (2019)): due to the connectivity between the policyholders, a cyber infection can spread fast and lead to many claims in a small amount of time. This endangers the mutualization process, which is at the core of the insurance activity. This paper aims at defining a standard framework to design stochastic scenarios of such accumulation episodes. The aim is to understand how reactions of the actors can have impact on the severity of such a massive attack, and how the shape of the network linking the policyholders can impact the outcome.

The history of such massive contagious cyber events is relatively poor. However, WannaCry (see Mohurle and Patil (2017)) and NotPetya (see Tourny (2017), Fayi (2018)), in 2017, have been warning signs that show that the financial impact of such events can be huge. The threat of a "cyber-hurricane" (see Institut Montaigne (2018)) being important, designing scenarios is a way to dimension the response against such an event, and to constitute the proper amount of reserve. If designing deterministic scenarios is a proper way to prepare for identified vulnerabilities, stochastic scenarios are a way to rapidly consider a large scope of situations. In this paper, we consider the general framework of Hillairet and Lopez (2021) used to design contagion scenarios in cyber-risk at a portfolio level. This approach is based on the separate modeling of the spread of the infection among the national (or even global) population, of the time required to assist an affected policyholder, and of the reaction (how fast people protect themselves against the threat). We here mostly focus on the spread of the virus, with a particular attention devoted to building a model which takes the potential links between policyholders (hence the connectivity of the network they form). The importance of connectivity has been considered for example in Fahrenwaldt et al. (2018), while the approach that we propose is somehow rougher but easier to calibrate from data.

The paper is organized as follows, a first section summarizes the general framework and the assumptions made on the portfolio. Then, the next section goes in the heart of the model by exposing the assumptions and the way it works. Finally, the last section illustrates this model on a portfolio, testing different scenarios before concluding.

2 General framework

In this section, we describe the general framework studied in this paper. The way we describe the contagion episode at a portfolio level is shown in Section 2.1, which summarizes the setting of Hillairet and Lopez (2021). The assumptions are discussed in Section 2.2.

2.1 Description of the portfolio

We follow the general framework of Hillairet and Lopez (2021) to model our portfolio. We have n policyholders, whose status with respect to the cyber event is described by a random vector $\mathbf{V}_i = (T_i, C_i, U_i)_{1\leq i\leq n}$. The vectors $(\mathbf{V}_i)_{1\leq i\leq n}$ are assumed to be independent, and represent the following time-to-event variables:

- T_i is the time at which policyholder i is hit by the cyber virus;
- U_i is the time required to deliver immediate assistance to the policyholder (for example restarting its activity), hence the claim is "solved" at time $T_i + U_i$;
- C_i is the time at which policyholder i manages to implement a protection that prevents the attack.

If $C_i \leq T_i$, the protection is achieved before the infection, and the potential victim is never hit. Let us also note that $\mathbb{P}(\{T_i = \infty\})$ may be strictly positive: in many attacks, some potential victims will never be hit, even if they do not implement protection.

The total number of victims in the portfolio is then $\mathfrak{N} = \sum_{i=1}^{n} \mathbf{1}_{T_i\leq C_i}$, while the number of policyholder to assist at a given time t is

$$\mathfrak{I}_t = \sum_{i=1}^{n} \mathbf{1}_{T_i\leq C_i}\mathbf{1}_{T_i\leq t} - \sum_{i=1}^{n} \mathbf{1}_{T_i\leq C_i}\mathbf{1}_{T_i+U_i\leq t}.$$

One is typically interested in the distribution of $\mathfrak{N}$, and on the distribution of $\sup_t(\mathfrak{I}_t)$, which represents the maximum number of victims to assist at some given time. This quantity is important, since the insurance company may fail to deliver its assistance service if this number exceeds its capacity of response.

2.2 Discussion on the assumptions

In this simple framework, we assume that, after time C_i, the protection against the threat is perfect. This assumption may be questionable, especially for large organisations. Some of them, even after identifying the vulnerability and knowing how to solve it, may have difficulties to implement a solution that protects their whole information system. Moreover, the protection is not only a matter of technique: human behavior (for example, in the case of phishing campaigns, individuals clicking on malicious weblinks) may contribute to lower the protection. To solve this issue, a possible and simple way to proceed is to consider a random variable C_i such that $\mathbb{P}(C_i = \infty) \neq 0$, which is a way to lower the ability of protecting oneself.

Let us also notice that the way the population protects itself from the threat can also be incorporated in the modeling of T. Hence, C can represent the additional protection that will implement the portfolio community (thanks to the specific help and attention devoted by the insurer) compared to the general population.

The strongest assumption is the independence between policyholders. This assumption is questionable, because connected entities may be more likely to be stroke by the same attack. Nevertheless, the portfolio is supposed to be a small population compared to the global one, so that the contagion is more likely to come from outside than from inside, which is, apart from the simplicity, the reason for relying on this assumption. Moreover, as it will be shown in Section 3.1, we can introduce covariates to materialize the fact that policyholders belong to some similar classes of risk following different temporal evolutions. This is a way to introduce dependence through this characteristic, without needing to model all the connections between policyholders as in Fahrenwaldt et al. (2018).

3 Model for the cyber pandemic

This section is devoted to a way to consider the connectivity between classes of actors through a simple epidemiological model, and to take it into account to model the infection variable T. In section 3.1, we introduce the multi-group SIR model, which is a commonly used tool (see for example Brauer et al. (2012)) to represent the spread of a contagion at a macroscopic level. Section 3.2 explains how to translate this type of macroscopic models in the framework of section 2.1. Finally, section 3.3 provides a simple formula to compute the final size of such a contagious episode that may be useful for calibration.

3.1 Multi-group SIR

Compartmental models are a common way to describe contagious events in epidemiology, since McKendrick (1925) and Kermack and McKendrick (1927). In Hillairet and Lopez (2021), a SIR model is used to describe the evolution of the cyber epidemic at a global level (hence to determine the distribution of T). The SIR (for "Susceptible - Infected - Recovered") model is one of the most simplest of compartmental model. Each member of the population belongs to one of the three categories:

- $S =$ the Susceptibles, that is the entities that can be stroke by the attack;
- $I =$ the Infected, that should be understood as the entities that are currently contagious, in the sense that they contribute to the spread of the epidemic;
- $R =$ the Recovered, sometimes called "Removed", are the ones for which contagion has been stopped.

Let us note that the status R does not mean that the crisis is over for the entity. The complete recovery can take much longer than the whole duration of the epidemic. In a biological epidemic, the category R gathers individuals who recovered, and the ones who died.

The evolution of each category through time is then described by a system of differential equations. When the population is heterogeneous (in our case, the targeted population may include entities from various activity sectors, like mining, construction, finance, ...), multi-group SIR models are a way to generalize this approach. Let's consider that we have d categories in the targeted population, and define, for all $j = 1, ..., d$, the quantities $s_j(t)$, $i_j(t)$, $r_j(t)$, denoting respectively the number of Susceptibles in category j at time t, the number of Infected in category j at time t, and the number of Removed in category j at time t. Next, following Magal et al. (2018), the evolution of each of these groups is given by

$$\frac{ds_j(t)}{dt} = -\left(\left\{\alpha_j(t) + \sum_{k=1}^{d} \beta_{k,j} i_k(t)\right\}\right) s_j(t), \tag{3.1}$$

$$\frac{di_j(t)}{dt} = \left\{\alpha_j(t) + \sum_{k=1}^{d} \beta_{k,j} i_k(t)\right\} s_j(t) - \gamma_j i_j(t), \tag{3.2}$$

$$\frac{dr_j(t)}{dt} = \gamma_j i_j(t). \tag{3.3}$$

The coefficient $\beta_{k,j}$ materializes how much the category k infects the category j. Let us note that the matrix $\mathbf{B} = (\beta_{k,j})_{1\leq k,j\leq d}$ may not necessarily be symmetric (a category k can protect itself against category j more than the category j protects itself from category k). Compared to a classical multi-group SIR model as in Magal et al. (2018), we added the functions $\alpha_j(t)$. This is to model the fact that the initialization of the attack may be caused by a burst of continuous attack. Typically, in the following, we will assume that $\alpha_j(t) = \alpha_j \mathbf{1}_{t\leq t_0}$, where $\alpha_j > 0$ and t_0 is the time after which the hackers stop spreading the virus (which propagates itself only through the contagion after t_0).

Let us note that the matrix $\mathbf{B}$ can also be used to simultaneously describe the risk of direct cross-infection as well as the risk of operating loss resulting from the attack affecting the activities of companies connected to the infected one.

3.2 From the multi-group SIR to the impact on an insurance portfolio

Let $\mathbf{x}_i$ denote the category to which the $i-$th policyholder belongs. The hazard rate function is defined as

$$\lambda_{T_i}(t) = \lim_{dt\to 0^+} \frac{\mathbb{P}(T_i \in [t, t+dt]|T_i \geq t)}{dt}.$$

If the portfolio is considered like a random sample from a population governed by the multi-group SIR, we have

$$\lambda_{T_i}(t) = \left\{ \alpha_j(t) + \sum_{k=1}^{d} \beta_{k,j} i_k(t) \right\}, \text{ if } \mathbf{x}_i = j.$$

The hazard rate entirely defines the distribution of a random variable. In our case, the considered variables T_i are durations. Hence the hazard rate is more adapted to describe their dynamical evolution than the density. Indeed, if $\lambda_{T_i}(t)$ is high, this means that the current (i.e. at time t) level of infection is high. We clearly see in this formula that this hazard function is logically higher when the current number of infected is high.

3.3 Total number of victims

Given a set of parameters describing the contagion according to equations (3.1) to (3.3), the real time trajectory of the infection is only known by numerically plotting each of the curves (s_j, i_j, r_j) which can be time-consuming in absence of closed formula. On the other hand, if one only wants to focus on the final number of infected computers, this total number of victims in each class can be retrieved much faster, solving a fixed-point problem. It can also help to perform a faster calibration of a model.

Let $r_j(\infty) = \lim_{t\to\infty} r_j(t)$. Since every Infected ultimately becomes Removed after a finite amount of time, $r_j(\infty)$ represents the total number of Infected in class j, that is the quantity we want to determine. Define $\mathbf{r}(\infty) = (r_j(\infty))_{1\leq j\leq d}$. Let $\mathcal{A}_j = \int_0^\infty \alpha_j(t)dt$, for $j = 1, ..., d$. For $\mathbf{x} = (x_1, ..., x_d)^{tr}$, where ^{tr} denotes the transpose, let

$$\Phi_j(\mathbf{x}) = i_j(0) + s_j(0)\left\{1 - \exp\left(-\mathcal{A}_j - \sum_{k=1}^{d} \frac{\beta_{k,j}}{\gamma_j} x_k\right)\right\},$$

and $\Phi(\mathbf{x}) = (\Phi_j(\mathbf{x}))_{1\leq j\leq d}$. The vector $\mathbf{r}(\infty)$ is the unique solution of the equation

$$\mathbf{r} = \Phi(\mathbf{r}),$$

on $\mathcal{R} = \{\mathbf{r} : 0 \leq r_j \leq s_j(0) + i_j(0)\}$.

This result can be retrieved from Magal et al. (2018), on multi-group SIR models, with the only modification that we introduced the presence of the terms α_j. Finding the solution of this fixed point equation can be done relatively fast by taking an arbitrary value $\mathbf{r}_0$ and studying the sequence $\mathbf{r}_{n+1} = \Phi(\mathbf{r}_n)$ which can be shown to converge to the solution.

4 Illustration

The use of the multi-group SIR can be illustrated by studying the propagation of a cyber event across different sectors of activity from an economy. This allows to model the contagion effects and the interactions between sectors. Another interest of this model is the ability to quantify the impact of different initial scenarios on the size of the crisis. For instance, one can assume that only one sector is initially targeted with a given intensity.

As in Hillairet and Lopez (2021), we will model a WannaCry-type event to illustrate the use of this multi-group SIR model. For this empirical analysis we consider a portfolio made of policyholders over five different sectors of activity. These sectors of activity have been chosen according to categories from OECD, namely:

- Mining and quarrying

- Manufacturing

- Electricity, gas, water supply, sewerage, waste and remediation services
- Construction
- Total business sector services

In Section 4.1, we show how macroeconomic data can help to build a reasonable contagion matrix $\mathbf{B}$. Models for the reaction C are provided in Section 4.2 while the empirical results are gathered in Section 4.3.

4.1 Example of calibration of the matrix B.

To use a multi-group model, we need some statistics of the interactions between sectors. To obtain credible values, statistics from the OECD[1] about the origin of the added value in function of the origin and destination sector in 2015 are used. The raw matrix of the flows of added values (in millions of USD) is represented in Table 1 with the origin sector in rows and the destination in columns. Our key assumption is that these flows of added values also reflect the connectivity between sectors. This is of course questionable but let us recall that our aim is only to get a reasonable idea of this connectivity to create a benchmark.

	Mining	Manufacturing	Energy	Construction	Services
Mining	225.52	1026.27	154.72	506.18	412.55
Manufacturing	14.86	8654.41	94.61	1709.06	1362.29
Energy	4.92	342.46	674.89	165.10	284.47
Construction	1.41	58.85	12.55	3685.20	197.56
Services	33.62	4396.65	249.46	2164.84	22206.97

Table 1: Exchange of added value between sectors - OECD data, 2015. A line represents the flow of added value sent from the corresponding sector to the sectors in columns.

In addition to the amount of the exchanges, we need the distribution of the companies across sectors. Again the data from the OECD[2] in 2015 are used. Table 2 presents the number of companies for each sector.

[1] Trade in Value Added (TiVA): Origin of value added in final demand

[2] SDBS Structural Business Statistics (ISIC Rev. 4) : Total number of enterprises, by sector

Sector	Number of companies	Percentage
Mining	66’492	0.20%
Manufacturing	3’068’178	9.02%
Energy	220’892	0.65%
Construction	4’874’747	14.34%
Services	25’768’765	75.79%

Table 2: Distribution of companies between sectors - OECD data, 2015

To obtain an interaction matrix from these data, some transformations must be performed. First, we can assume that the flow of information between two sectors do not depend on the direction of the economic exchanges. For this reason, the value of the exchange between two specific sectors is taken as the sum of the exchanges in both directions. Second, to model the amount of risk per company, the absolute value of the exchanges is divided by number of companies in the origin sector. Finally, the matrix is normalized such that the sum of the coefficients is equal to one. The final interaction matrix $\mathbf{B}$ is then assumed to be proportional to the matrix $\mathbf{B}_0$ defined in Table 3, that is $\mathbf{B} = \beta \mathbf{B}_0$, where $\mathbf{B}_0$ contains all the information about the connectivity, and β represents a strength of the contagion.

	Mining	Manufacturing	Energy	Construction	Services	Total
Mining	0,0634	0,2927	0,0449	0,1427	0,1255	0,6692
Manufacturing	0,0063	0,0527	0,0027	0,0108	0,0351	0,1076
Energy	0,0135	0,0370	0,0571	0,0150	0,0452	0,1679
Construction	0,0019	0,0068	0,0007	0,0141	0,0091	0,0326
Services	0,0003	0,0042	0,0004	0,0017	0,0161	0,0227
Total	0,0855	0,3934	0,1057	0,1844	0,2309	1

Table 3: Normalized Interaction matrix $\mathbf{B}_0$.

According to this table, we see that the Mining & Quarrying sector is the most contagious one, followed by the Energy sector. This is quite intuitive, as an attack stopping the activity of those sectors will have strong effects on the activities of other sectors. This high contagiousness is however to be nuanced by the small population of these sectors. The sector receiving most of the cross-infection are the Services and Manufacturing ones. The Services sector is indeed highly connected and represents a significant part of the economy, explaining that it is a target of interest. As for the manufacturing sector, it depends a lot on the supplies from other sectors, as from the Mining & Quarrying one for instance. Moreover, companies in this sector often operate complex supply chains, making them highly inter-connected as well.

4.2 Model for the reaction C

The reaction C represents the time when the policyholders manage to implement a protection that prevents the attack, making them immune to the attack. It describes the fastness to identify the incident and to communicate countermeasures that may prevent the spread. It also reflects the behavior of the policyholders, in their reactivity concerning the alerts. We will denote C_j the random time at which the policyholder j gains immunity. There is then a competition between two time-variables, with C_j acting like a right-censoring variable to the time of infection T_j. We implicitly assume that we consider the part of the portfolio which is exposed to the risk (that is, none of the policyholders is protected against the risk at the beginning of the episode).

As in Hillairet and Lopez (2021), we consider three scenarios for the response variable:

- From the time of response (τ_1), the rate at which policyholders update their security system is constant through time. This is described by a translated exponential distribution:
$$\lambda_C^{(1)}(t) = c_1 \mathbb{1}_{t \geq \tau_1}$$
- From the time of response (τ_2), the vigilance of the policyholders decreases through time. This is modelled using a Pareto-type distribution:
$$\lambda_C^{(2)}(t) = c_2 (t - \tau_2 + 1/2)^{-\alpha_2} \mathbb{1}_{t \geq \tau_2}, \text{ for } \alpha_2 > 0$$
- Finally, the vigilance of the policyholders increases through time after the time of response (τ_3). A Weibull-type distribution is chosen for this scenario:
$$\lambda_C^{(3)}(t) = c_3 (t - \tau_3)^{\alpha_3} \mathbb{1}_{t \geq \tau_3}, \text{ for } \alpha_3 > 0$$

The parameters of the three types of responses are taken such that $\mathrm{E}[C_j - \tau_j | C_j \geq \tau_j] = 1$. This makes the comparisons more legitimate, as the response distributions only differ by the shape of its hazard function. In addition, for each scenario, different delays are considered:

- A fast response: $\tau_j = 3$ days after the start of the event;
- A medium response: $\tau_j = 5$ days after the start of the event;
- A slow response: $\tau_j = 7$ days after the start of the event.

4.3 Empirical analysis

4.3.1 Multi-group SIR model: WannaCry-type scenario

To illustrate the use of the multi-group SIR model in the modeling of a cyber event, we calibrate it on a WannaCry-type event.

The first wave of infections being relatively fast for this type of events, the event is modeled over a period of ten days.

We first consider functions $\alpha_j(t) = \alpha \mathbf{1}_{t \leq t_0}$. This corresponds to a situation where there is an initial burst of attacks during a short period (before t_0, here we will take $t_0 = 1$ day), after which only the contagion between victims is involved in the propagation of the cyber epidemic. Moreover, since the functions α_j are identical, this means that the attack targets all categories homogeneously.

We then need to calibrate the coefficient β for the contagion, and the burst intensity α such that the final size of the epidemic is of the same order as the WannaCry event. Following the steps of the calibration method in Hillairet and Lopez (2021) (with a slightly different model), the number of victims is taken as 300'000 infections. We also use the same total number of Susceptible as in Hillairet and Lopez (2021), estimated at 4'064'279 companies.

This leads to the following values:

$$\alpha = 7 \times 10^{-3}$$

$$\beta = 1.845 \times 10^{-5}$$

The corresponding estimated function $t \to i_t$ during the ten first days of the WannaCry crisis is shown in Figure 1. These parameters lead to 300'362 infections after ten days, with a maximum of 31'112 victims in the Infected state at a given time.

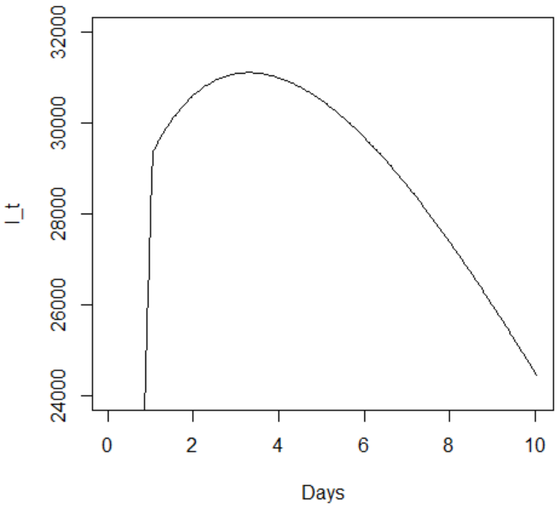

Figure 1: Function $t \to i_t$ obtained for the first ten days using the estimated β and α coefficients.

In this situation, the attackers strike the victims homogeneously. But due to the differences in terms of connectivity, one may try to figure out what happens if the same strength of attack is targeted against a specific category. Applying the first burst intensity on one sector will allow to assess the impact of one sector on the size of the epidemic. To analyze the effect of each sector we kept the same value for β. Note that to make things comparable, the coefficient α describing intensity must be adjusted to account for the number of companies in each sector. To keep the intensity of the initial attack similar similar (i.e., the same number of victims on the first day), a smaller value of α is needed if the number of Susceptibles in that class is high and vice versa. The table 4 summarizes the results for each scenario. The results are ordered starting from the sector having the most negative impact on the size of the epidemic, to the one leading to the least impact.

Targeted sector	β	α	Total infected	Peak
Mining	1.845×10^{-5}	3.5000	702'566	88'404
Manufacturing	1.845×10^{-5}	0.0776	577'998	69'401
Energy	1.845×10^{-5}	1.0769	469'942	53'263
Services	1.845×10^{-5}	0.0092	258'737	27'722
Construction	1.845×10^{-5}	0.0488	213'020	24'489

Table 4: Results of the scenarios targeting a single sector. The length of the initial attack (t_0) is still 1 day.

We notice that the initial burst targeting the Mining and Quarrying sector leads to the most important increase in the size of the epidemic. This is expected given the high contagiousness of this sector represented by the coefficients of table 3. Moreover, as this sector regroups a small number of companies, the burst intensity is very high leading to almost all the sector's companies being infected. A burst targeting the Manufacturing or the Energy sector also increases the size of the epidemic. Note that the effect observed is not totally surprising. This has been illustrated recently by the impact of the cyber-attack of Colonial Pipeline[3] in May 2021, cutting off the distribution of oil in part of the US. In the considered SIR model, the contagion matrix is based on the economic exchanges, that might also capture the impact of a cyber-attack in term of operating loss, in addition to the pure contagion by the cyber-virus.

On the opposite side of the table, we find the Construction sector. A burst targeting this sector leads to a decrease of the size of the epidemic. Looking at the coefficients from table 3, we notice that this sector is in average less contagious that the other sectors.

It is of course possible to analyze the effect of the different scenarios, at segment level. Table 5 represents the proportion of infected in each sector in function of the initial target of the hackers. Overall, we notice that the Mining & Quarrying and Energy sectors suffer less when they are not directly targeted. On the opposite, the Manufacturing and Services sectors see more infections when the intensity of the attack is focused in one of the sectors interacting with them. It is also interesting to emphasize that the Services sector is more affected when the other sectors are targeted directly, apart from the construction sector, than when it is itself targeted by the initial burst.

[3] https://www.bloomberg.com/news/articles/2021-05-08/u-s-s-biggest-gasoline-and-pipeline-halted-after-cyberattack

Targeted sector	Mining	Manufacturing	Energy	Construction	Services
Uniform attack	1,06%	4,11%	0,99%	2,07%	8,86%
Attack on Mining	97,77%	12,47%	1,34%	5,39%	20,04%
Attack on Manufacturing	1,00%	15,70%	0,65%	2,99%	16,32%
Attack on Energy	0,97%	6,23%	67,53%	2,46%	13,47%
Attack on Construction	0,31%	2,37%	0,20%	6,26%	5,45%
Attack on Services	0,26%	2,61%	0,22%	1,02%	7,89%

Table 5: Proportion of each sector affected by the epidemic

Looking at the infections between sectors we notice that the epidemic has the tendency to be concentrated in the Services sector and, to a lesser extent, in the Manufacturing sector. To illustrate this, figure 2 shows for each origin sector the proportion of infections according to the destination sector. In this figure the case of a uniform attack across sectors is considered. Note, that similar trends are observed in the different attack scenarios. This figure also confirms that a minority of the cross-contamination are targeting the Mining & Quarrying and Energy sectors.

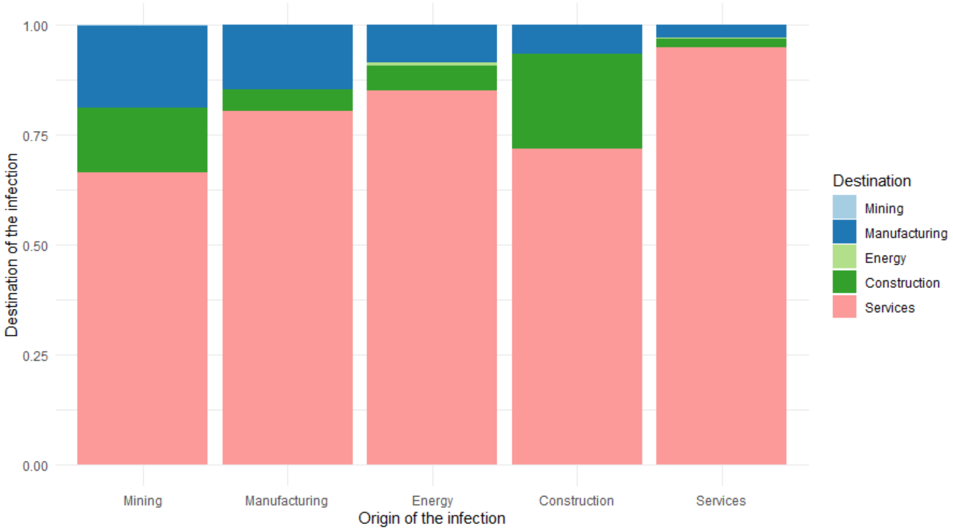

Figure 2: Cross-infection between sectors after a uniform burst (proportion by destination)

To have a better understanding of the behavior of the epidemic, it is also interesting to look at the time evolution of the number of infected inside each sector. On figure 3 we can see that epidemic continues to grow only in the Services sector after a uniform attack on the portfolio. In the other sectors, a slow decrease of the number of infected is observed from the time where the burst of attacks stops.

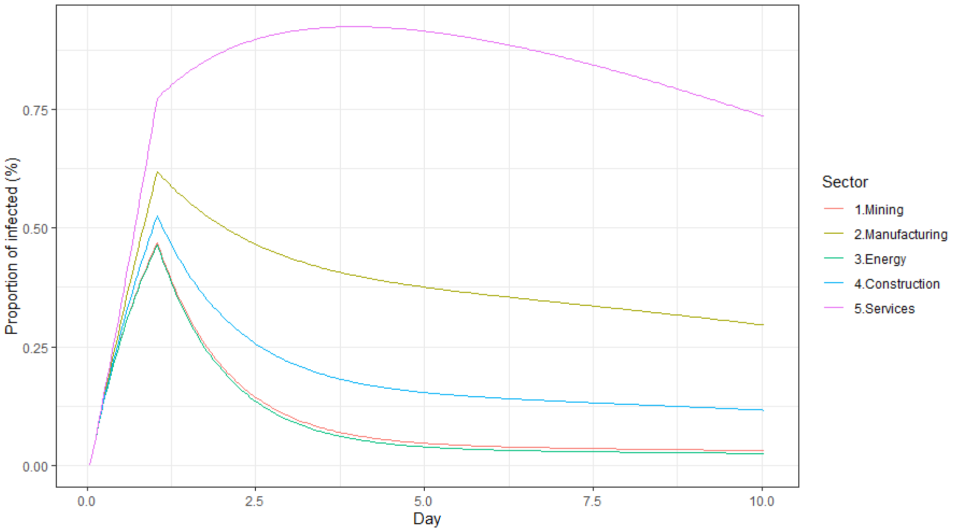

Figure 3: Evolution of the proportion of infected - Uniform attack

Figure 4 represents the evolution of the proportion of infected for each sector after a first burst targeting the Mining & Quarrying sector. Note that the y-axis has been bounded to 4%, so that the effects of the infection remain readable for all sectors. Indeed, the value of the peak for the Mining & Quarrying sector is very high at 70% of infected companies after 10 hours of epidemic. Let us recall that it was the scenario having the biggest effect on the size of the epidemic (Table 4). This graph shows that the number of infected in the Manufacturing sector reacts quickly. With the epidemic developing, the most hit sector is the Services one. The Energy sector suffers very little from the infection. As said previously, similar observations are made on other scenarios, with the Mining & Quarrying and Energy sectors suffering little from the cyber-attack when they are not directly targeted.

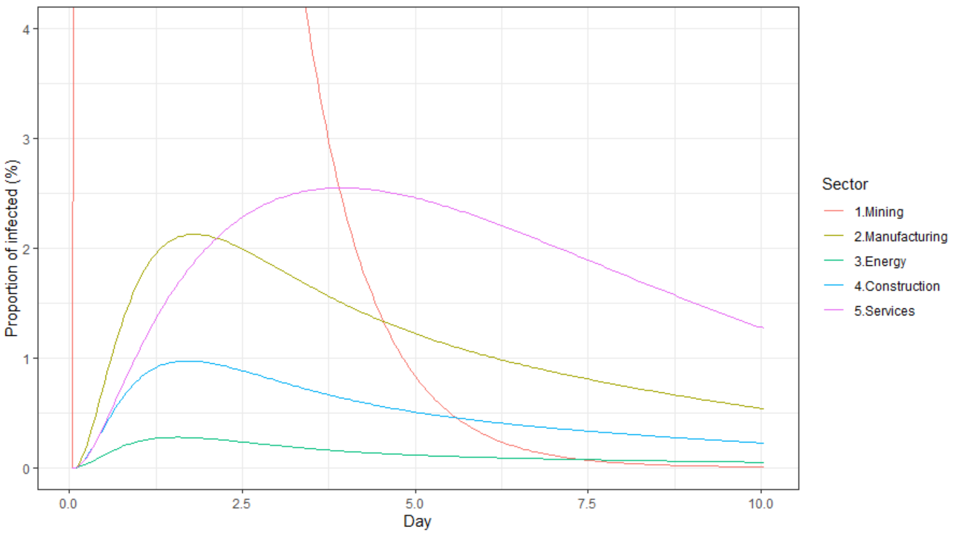

Figure 4: Evolution of the proportion of infected - Attack on Mining & Quarrying

4.3.2 Simulations results

To analyze the evolution and behavior of the epidemic inside a portfolio, we consider 10'000 exposed policies. Using the parameters estimated for the SIR model and the response variable, 10'000 simulations are performed. The objective of this exercise is to observe the sensitivity of the epidemic to different response types. This is done considering the sector dimension and different initial attack scenario. The responses parameters used are the ones described previously in section 4.2. In addition, we suppose that the victims need in average 3 days of assistance (exponential variable U) to restart their activity.

Finally, we assume that the considered portfolio has the same composition as the economy. Thus, the same sector distribution as the one used for the SIR model is applied. Table 6 gives the number of policies inside each sector.

Sector	Number of companies	Percentage
Mining	20	0.20%
Manufacturing	902	9.02%
Energy	65	0.65%
Construction	1'434	14.34%
Services	7'579	75.79%

Table 6: Distribution of companies between sectors

First, we look at the impact of the different response's scenarios on the final size of the epidemic. For the time being, the uniform initial burst scenario is considered. Table 7 presents different statistics about the total number of infected resulting from the simulations. In addition, the average total number of infected is given for each sector.

From the portfolio figures, we see that the reduction of the total size of the epidemic is relatively similar between the different type of responses. The exponential response, corresponding to a constant awareness and rate of countermeasures implementation, seems to have slightly better performances. As expected, faster response time leads to a bigger reduction of the epidemic size. This reduction is not uniform across sectors. The biggest reaction is observed in the Manufacturing and Services sector. Whereas the Mining & quarrying and Energy sectors are less impacted. It is interesting to note that, at sector level, the exponential response is not necessarily the most efficient.

	Portfolio statistics				Mean scenario - Sectors				
Reaction	Mean	σ	Min	Max	Mining	Manuf.	Energy	Constr.	Services
No reaction	741.75	26.31	647	840	0.21	37.20	0.63	29.89	673.81
Slow expo.	618.30	24.19	618	717	0.20	31.75	0.60	26.49	559.26
Slow Pareto	619.61	24.21	528	715	0.20	31.81	0.60	26.50	560.50
Slow Weibull	622.55	24.18	528	717	0.20	31.93	0.60	26.59	563.22
Med. expo.	486.23	21.54	410	570	0.18	25.94	0.57	22.88	436.65
Med. Pareto	486.34	21.51	406	583	0.18	25.92	0.57	22.90	436.76
Med. Weibull	487.96	21.49	418	578	0.18	26.00	0.57	22.93	438.27
Fast expo.	340.71	18.08	272	415	0.17	19.54	0.53	18.90	301.56
Fast Pareto	340.65	18.05	273	414	0.17	19.53	0.54	18.91	301.50
Fast Weibull	341.63	18.13	268	412	0.17	19.57	0.53	18.95	302.39

Table 7: Statistics on the number of victims depending on the reaction. Portfolio statistics provide figures for the total number of victims (all sectors are gathered). The second part shows the decomposition of the central scenario.

The peak of the epidemic is also of interest for an insurance company, as it represents the maximum number of policyholders needing assistance at the same time. This can be used to determine the needed response capacity. Table 8 presents the statistics related to this peak. As for the size of the epidemic, the exponential response seems to have the bigger effect on the peak. At sector level, the peak is most sensible to the response in the Mining & Quarrying and Services sectors. On the other hand, it varies very little in the Construction sector. Note that the slowest responses, after 5 and 7 days, have a very limited impact on the peak. Indeed, the peak of the epidemic happens earlier in most cases.

	Portfolio statistics				Mean scenario - Sectors				
Reaction	Mean	σ	Min	Max	Mining	Manuf.	Energy	Constr.	Services
No reaction	216.04	32.97	74	239	0.20	13.70	0.58	12.38	199.26
Slow expo.	212.39	20.24	73	192	0.19	13.29	0.56	12.29	195.27
Slow Pareto	213.49	22.26	73	191	0.19	13.35	0.56	12.30	196.45
Slow Weibull	213.48	22.50	73	195	0.19	13.37	0.56	12.30	196.44
Med. expo.	200.58	10.85	58	150	0.18	12.66	0.54	12.15	182.85
Med. Pareto	203.79	11.62	58	148	0.18	12.74	0.54	12.18	186.14
Med. Weibull	203.94	12.35	58	149	0.18	12.78	0.54	12.18	186.34
Fast expo.	171.23	6.68	49	107	0.17	11.50	0.51	11.89	153.00
Fast Pareto	177.93	5.50	49	103	0.17	11.66	0.51	11.93	159.60
Fast Weibull	178.83	6.79	49	104	0.17	11.73	0.51	11.95	160.65

Table 8: Statistics on the peak of the epidemic depending on the reaction. Portfolio statistics provide figures for the total peak (all sectors are gathered). The second part shows the decomposition of the central scenario

From these simulations we observe that the deployment of countermeasures has a strong impact on the epidemic. This is particularly true in early responses scenario. Figure 5 illustrates this with the evolution of the number of Infected i_t for a given simulation.

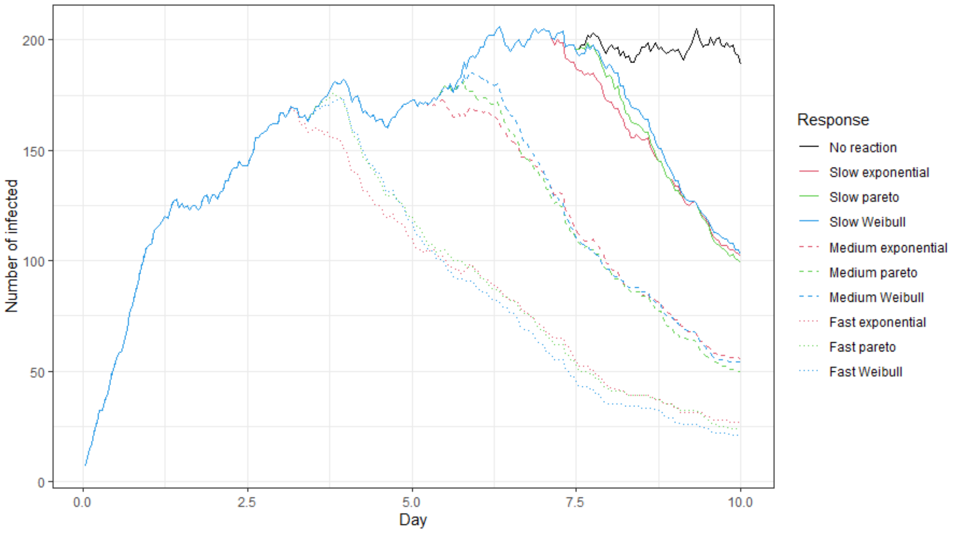

Figure 5: Evolution of the proportion of Infected - Uniform burst.

This impact is however not uniform across sectors. If the Services sector drives the behavior of the portfolio, leading to very similar evolution, it is not necessarily the case for the other ones. For instance, in the Construction sector the peak is in general observed at the beginning of the epidemic. Figure 6 represents a typical evolution of the epidemic inside this sector. We notice that the peak appears before the deployment of countermeasures, thus limiting their effect.

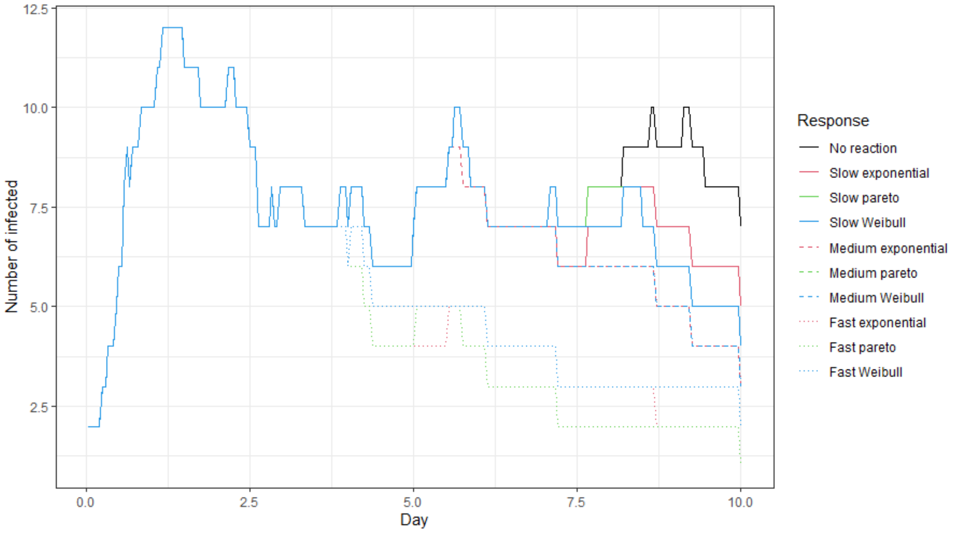

Figure 6: Evolution of the number of Infected in the Construction sector - Uniform burst.

The deployment of countermeasures also has variable efficiency depending on the initial scenario of the epidemic ignition. As we have seen before, in table 4, the size of the epidemic depends on the intensity and target of the initial burst. The latter also impacts the speed of spreading of the epidemic. Table 9 shows for the most populated segment (ie. Services) the impact of each response on the average final size of the epidemic as a function of the initial attack intensity (scenarios considered in table 4). Table 10 summarizes the impact on the average value of the peak of the epidemic. In both tables, impacts are given as the relative deviation with respect to the non-reaction scenario. To complete the analysis, tables 11 and 12 present the same results for the Construction sector, the second most populated one.

	Uniform	Mining	Manufacturing	Energy	Construction	Services
Slow expo.	-17,00%	-12,70%	-14,11%	-15,45%	-18,38%	-16,97%
Slow Pareto	-16,82%	-12,41%	-13,81%	-15,19%	-18,24%	-16,81%
Slow Weibull	-16,41%	-12,03%	-13,40%	-14,76%	-17,85%	-16,46%
Med. expo.	-35,20%	-29,39%	-31,55%	-33,29%	-36,99%	-34,76%
Med. Pareto	-35,18%	-29,17%	-31,37%	-33,21%	-36,99%	-34,75%
Med. Weibull	-34,96%	-28,80%	-31,05%	-32,90%	-36,84%	-34,55%
Fast expo.	-55,24%	-51,16%	-53,34%	-54,20%	-56,32%	-54,09%
Fast Pareto	-55,25%	-51,03%	-53,19%	-54,16%	-56,44%	-54,14%
Fast Weibull	-55,12%	-50,67%	-52,92%	-53,95%	-56,36%	-53,98%

Table 9: Services - Impact of the countermeasures on the final size of the epidemic in function of the attack scenario

	Uniform	Mining	Manufacturing	Energy	Construction	Services
Slow expo.	-2,01%	-0,02%	-0,12%	-0,61%	-4,90%	-2,21%
Slow Pareto	-1,41%	0,00%	-0,03%	-0,34%	-3,91%	-1,63%
Slow Weibull	-1,42%	-0,01%	-0,04%	-0,36%	-3,83%	-1,63%
Med. expo.	-8,24%	-2,53%	-3,90%	-5,56%	-13,05%	-7,62%
Med. Pareto	-6,58%	-1,35%	-2,40%	-3,93%	-11,28%	-6,26%
Med. Weibull	-6,48%	-1,42%	-2,43%	-3,93%	-11,07%	-6,15%
Fast expo.	-23,22%	-19,14%	-21,25%	-21,63%	-27,32%	-20,47%
Fast Pareto	-19,91%	-14,69%	-16,88%	-17,63%	-24,49%	-17,57%
Fast Weibull	-19,38%	-14,12%	-16,29%	-17,10%	-24,03%	-17,06%

Table 10: Services - Impact of the countermeasures on the peak of the epidemic in function of the attack scenario

First, it is interesting to note that the countermeasures have a bigger effect on the final size of the epidemic than on its peak of infections. This is particularly true for slower responses, leading to a marginal reduction of the peak value. As expected, the impact of countermeasures depends on the initial attack scenario. The biggest impact is observed for an attack targeting the Construction sector and uniform burst. The smallest ones are observed when the Mining & Quarrying, Manufacturing and Energy sectors are targeted. Observing the behavior of the epidemic in these cases, we notice that targeting these sectors leads to a quicker diffusion of the epidemic. The epidemic is then already well developed and often the peak is passed when countermeasures are deployed. This explains the smaller effect of the response in these scenarios.

	Uniform	Mining	Manufacturing	Energy	Construction	Services
Slow expo.	-11,37%	-7,35%	-10,90%	-11,45%	-7,72%	-12,27%
Slow Pareto	-11,35%	-7,20%	-10,83%	-11,16%	-7,70%	-12,11%
Slow Weibull	-11,04%	-7,00%	-10,40%	-10,87%	-7,43%	-11,92%
Med. expo.	-23,45%	-17,14%	-24,38%	-24,55%	-15,49%	-24,20%
Med. Pareto	-23,40%	-16,99%	-24,19%	-24,43%	-15,47%	-24,30%
Med. Weibull	-23,29%	-16,80%	-24,02%	-24,18%	-15,24%	-24,24%
Fast expo.	-36,77%	-30,66%	-41,07%	-40,09%	-24,83%	-37,42%
Fast Pareto	-36,73%	-30,34%	-41,01%	-39,98%	-24,84%	-37,46%
Fast Weibull	-36,60%	-30,14%	-40,72%	-39,81%	-24,53%	-37,34%

Table 11: Construction - Impact of the countermeasures on the final size of the epidemic in function of the attack scenario

	Uniform	Mining	Manufacturing	Energy	Construction	Services
Slow expo.	-0,72%	0,00%	-0,54%	-0,87%	-0,03%	-0,93%
Slow Pareto	-0,64%	0,00%	-0,38%	-0,73%	-0,02%	-0,85%
Slow Weibull	-0,60%	0,00%	-0,35%	-0,62%	-0,02%	-0,78%
Med. expo.	-1,80%	-0,13%	-2,81%	-3,07%	-0,11%	-2,27%
Med. Pareto	-1,64%	-0,09%	-2,35%	-2,67%	-0,08%	-2,10%
Med. Weibull	-1,58%	-0,08%	-2,19%	-2,51%	-0,08%	-2,08%
Fast expo.	-3,94%	-1,33%	-9,12%	-7,13%	-0,52%	-4,03%
Fast Pareto	-3,59%	-0,91%	-7,81%	-6,38%	-0,39%	-3,84%
Fast Weibull	-3,45%	-0,80%	-7,55%	-6,14%	-0,31%	-3,72%

Table 12: Construction - Impact of the countermeasures on the peak of the epidemic in function of the attack scenario

The same kind of observations as for the Services sector can be made for the Construction one. The loss of efficiency of the response is the most important when the Mining & Quarrying sector is targeted and when it is itself targeted. In those case the response has almost no impact on the peak. During these scenarios, the epidemic grows fast under the effect of the initial burst. However, when this attack stops, the epidemic tends to decrease slowly. Thus, most of the infections and the peak happen before the countermeasures, explaining their limited impact.

From the observations on these two sectors, it is interesting to observe that the countermeasures have less effects in scenarios targeting the most contagious sectors. In those cases, the development of the epidemic tends to be faster, with a majority of infections happening before the time of response. For a given time of response, the different response type led to similar effects on the epidemic. However, the exponential response, corresponding to constant awareness of policyholders, seems to be slightly more efficient in most cases.

5 Conclusion

In this paper, we show how to design stochastic accumulation scenarios for cyber insurance through a flexible way that considers the connectivity between actors. Instead of requiring to exactly know how policyholders are connected with each other (which seems quite difficult), the model requires relatively few inputs. In the application, we showed how to calibrate such a model based on macro-level data that can be retrieved more easily. Of course, this calibration is not perfect since the indicators that we use are indirectly linked to the question of cyber connectivity. Nevertheless, let us point out that the aim of this scenario generation procedure is not necessarily prediction, but more about giving reasonable ideas of the pattern of a cyber pandemic, playing with the infection dynamic (here we gave the example of WannaCry, but other patterns can of course be considered) and the connectivity assumption. Let us also notice that the approach we develop can also be extended to model indirect consequences of a cyber incident: if an industrial sector is very dependent from another, an attack on this key sector can lead to business disruption, as it has been seen for example in the Colonial Pipeline attack (see Hobbs (2021)). In which case, the contagion may spread not only through digital channels.

References

Agence Nationale de la Sécurité des Systèmes d'Information (2021). Etat de la menace rancongiciel. `https://www.cert.ssi.gouv.fr/uploads/CERTFR-2021-CTI-001.pdf`.

Brauer, F., Castillo-Chavez, C., and Castillo-Chavez, C. (2012). *Mathematical models in population biology and epidemiology*, volume 2. Springer.

Eling, M. and Loperfido, N. (2017). Data breaches: Goodness of fit, pricing, and risk measurement. *Insurance: mathematics and economics*, 75:126–136.

Fahrenwaldt, M. A., Weber, S., and Weske, K. (2018). Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin: The Journal of the IAA*, 48(3):1175–1218.

Farkas, S., Lopez, O., and Thomas, M. (2021). Cyber claim analysis using generalized pareto regression trees with applications to insurance. *Insurance: Mathematics and Economics*, 98:92–105.

Fayi, S. Y. A. (2018). What petya/notpetya ransomware is and what its remidiations are. In *Information technology-new generations*, pages 93–100. Springer.

Hillairet, C. and Lopez, O. (2021). Propagation of cyber incidents in an insurance portfolio: counting processes combined with compartmental epidemiological models. *Scandinavian Actuarial Journal*, pages 1–24.

Hobbs, A. (2021). The colonial pipeline hack: Exposing vulnerabilities in us cybersecurity.

Institut Montaigne (2018). Cyber menace : avis de tempête. `https://www.institutmontaigne.org/ressources/pdfs/publications/cybermenace-avis-de-tempete-rapport.pdf`.

Kermack, W. and McKendrick, A. (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A.*, 115:700–721.

Lallie, H. S., Shepherd, L. A., Nurse, J. R., Erola, A., Epiphaniou, G., Maple, C., and Bellekens, X. (2021). Cyber security in the age of covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers & Security*, 105:102248.

Magal, P., Seydi, O., and Webb, G. (2018). Final size of a multi-group sir epidemic model: Irreducible and non-irreducible modes of transmission. *Mathematical biosciences*, 301:59–67.

McKendrick, A. G. (1925). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130.

Mohurle, S. and Patil, M. (2017). A brief study of wannacry threat: Ransomware attack 2017. *International Journal of Advanced Research in Computer Science*, 8(5):1938–1940.

Tourny, È. (2017). Cyber sécurité et attaques informatiques: les leçons à tirer de wanna cry et not petya. *Paix et sécurité européenne et internationale*, (7).

Welburn, J. W. and Strong, A. M. (2019). Systemic cyber risk and aggregate impacts. *Risk Analysis.*

6 About the serie and the authors...

6.1 The DetraNotes

The Detra Notes are a series of educational papers dedicated to the insurance sector. Those notes are published by members of the Detralytics team and written in a clear and accessible language. The team combines academic expertise and business knowledge. Detralytics was founded to support companies in the advancement of actuarial science and the solving of the profession's future challenges. It is within the scope of this mission that we make our work available through our DetraNotes and FAQctuary's series.

6.2 Authors' biographies

Louise d'Oultremont

Louise is part of Detralytics' Talent Accelerator Program (TAP), a unique opportunity for recently graduated actuaries to discover the different facets of the actuarial profession. Prior to joining Detralytics, Louise did an internship at Axa Belgium, where she was in the risk pricing team. Louise holds two Master's degree in Actuarial Science and Mathematics, both from UCLouvain. Her thesis focused on multi-state individual reserving and consisted in comparing an analytical method and a simulation method using an individual multi-state semi-Markovian model.

Olivier Lopez

Olivier is Scientific Director at Detralytics, as well as a Professor in Actuarial Science at Sorbonne Université, Paris. Since 2016, he is the director of ISUP (Institut de Statistique de l'Université de Paris), which is Sorbonne's actuarial department. He is a fully qualified member of the French institute of actuaries (Institut des Actuaires), and member of its scientific committee. He is member of the Education Committee of the European Actuarial Association and of the International Actuarial Association. His main research fields include survival analysis, applications of machine learning to insurance, cyber insurance. A full list of publications is available at `https://sites.google.com/view/sitepersonneldolivierlopez/home`.

Brieuc Spoorenberg

Brieuc is part of the Talent Accelerator Program (TAP). Prior to joining Detralytics, Brieuc worked as an intern in P&C Pricing at Axa Belgium. Brieuc holds a Master's degree in Actuarial Sciences from UCLouvain. For his thesis, he worked on the study the correlation between claims frequency and severity in motor insurance and its impact on the final premium.

Detr²alytics

Expertise and innovation at the service of your future

www.detralytics.com - info@detralytics.eu