Cyber risk modeling using a two-phase Hawkes Process with external excitation

Yousra Cherkaoui, Milliman R&D - CREST Ensae

Joint work with :

Alexandre Boumezoued, Milliman R&D
Caroline Hillairet, CREST Ensae

European Congress of Actuaries

June 7th, 2024

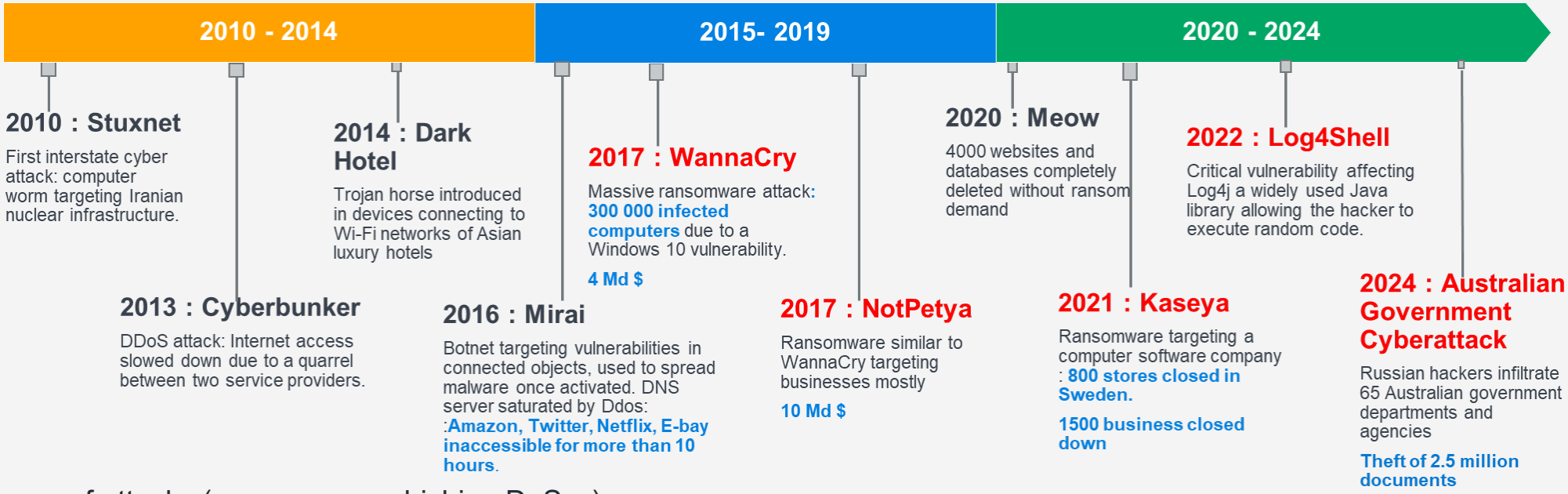
Milliman

Agenda

- Context
- Cyber risk modelling using Hawkes processes with vulnerabilities
- Cyber attacks and vulnerability databases
- Calibration of the One-Phase Hawkes process
- Response measures using the second phase of the Hawkes process
- Future research questions

Cyber risk

Context



- **Various types** of attacks (ransomware, phishing,DoS…)
- Focus on **contagious** cyber incidents, by taking into account exogenous excitation
- Regular publications of **vulnerabilities** that may cause **cyber pandemics :** EternalBlue (Wannacry, NotPetya), Log4Shell etc
- **Quantifying** impact of **protection measures** to **limit the effect of a cyber attack** (patching vulnerabilities for instance)

Cyber risk modelling

Hawkes processes

The **number of claims** up to time **t (Frequency component)**

Cumulative loss process of a cyber-insurance portfolio

$$L(t) := \sum_{i=1}^{N(t)} Y_i$$

The cost of i^{th} claim (Severity component)

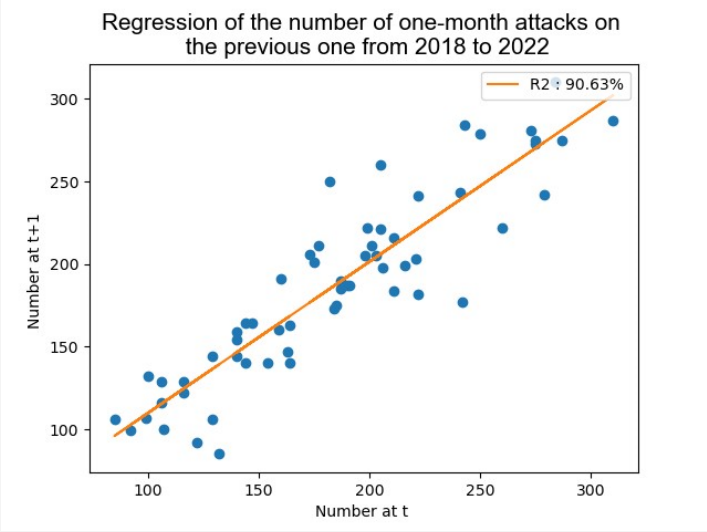
- A **Hawkes process** is a counting process $N(t) = \sum_{i\geq 1} 1_{T_i \leq t}$ which has the following intensity:

$$\lambda_t = \lambda_0 + \sum_{T_i<t} \phi(t - T_i)$$

Baseline intensity

Excitation kernel

- The sum $\sum_{T_n<t} \phi(t - T_n)$ represents **the impact of past events** and captures the **self-excitation** property. Hawkes processes allow the modelling of **intense arrivals** and **regime changes**.
- **Autoexcitation**, causality between attacks
- **Parametric** and **tractable**
- Adapted to some **cyber datasets**



Frequency modelling of cyber attacks

- Grasp **internal excitations** through **Hawkes frequency process** $N(t)$
- Add **external excitation** through **cyber vulnerabilities**

Cyber risk modelling

Hawkes processes with external excitation

$$\lambda_t = \lambda_0 + \sum_{\overline{T_k}<t} \overline{\phi}\left(t - \overline{T_k}\right) + \sum_{T_i<t} \phi(t - T_i)$$

Baseline intensity

External excitation

Vulnerabilities dates

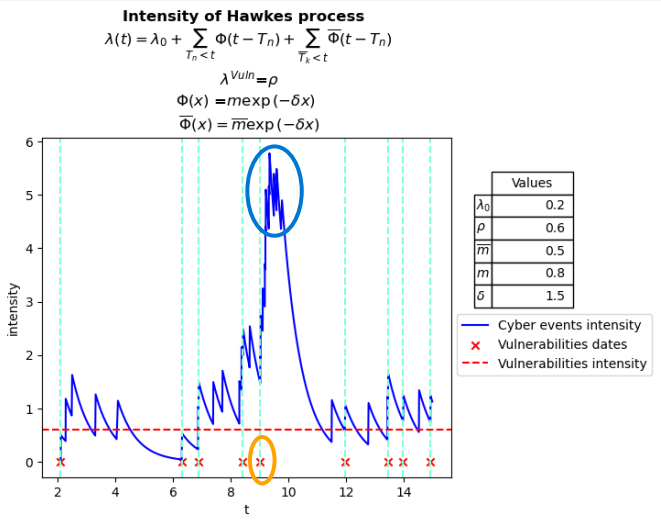
Internal excitation

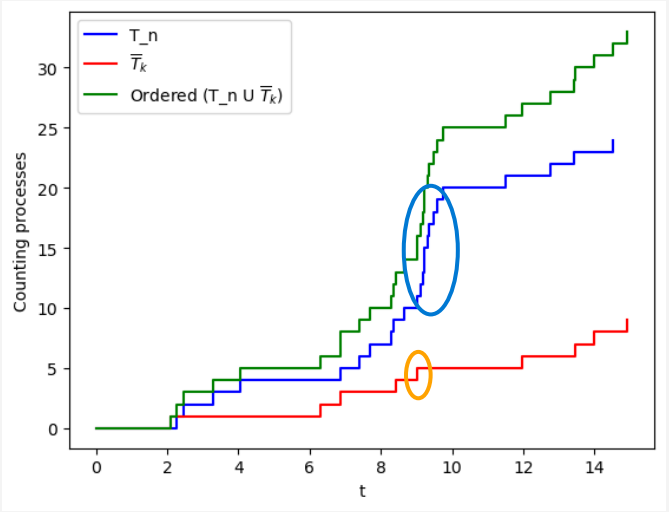
Cyber attacks dates

- Here the **external excitation** represents the regularly published **vulnerabilities** in public datatabases (**NVD database** for example)

An illustrative **simulated** example using the **thinning algorithm**

- External events (here **cyber vulnerabilities**) increase the intensity of the **cyber attacks** process leading to internal **contagion**





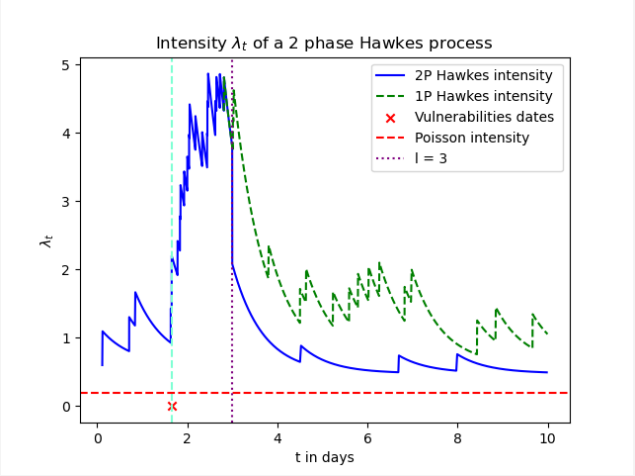
Cyber risk modelling

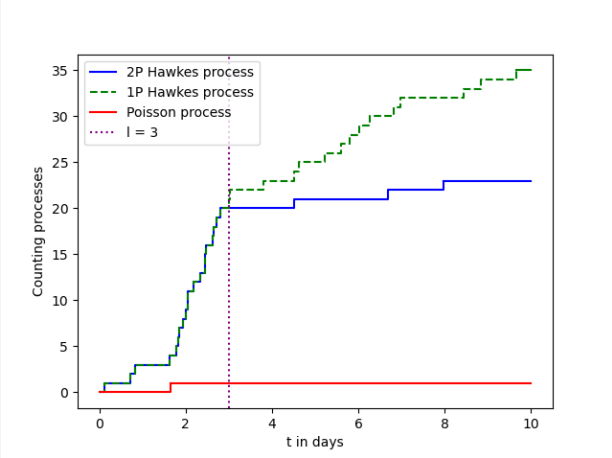
A Two-Phase Hawkes process with external excitation

$$\lambda_t = \begin{cases} \underbrace{\lambda_0}_{\text{Baseline intensity}} + \underbrace{\sum_{T_k<t} \overline{m}\, e^{-\delta(t-T_k)}}_{\text{External excitation : cyber vulnerabilities}} + \underbrace{\sum_{T_i<t} m^{bl}\, e^{-\delta(t-T_i)}}_{\text{Self excitation : cyber attacks}} & \text{if } t \leq \underbrace{\ell}_{\text{Reaction time}} \\ \underbrace{\alpha_0}_{\text{Reaction parameter}}\lambda_0 + \underbrace{\alpha_1}_{\text{Reaction parameter}}(\lambda_{\ell^-} - \lambda_0) & \text{if } t = \ell \\ \alpha_0\lambda_0 + \alpha_1(\lambda_{\ell^-} - \lambda_0)e^{-\delta(t-\ell)} + \sum_{\ell<T_i<t} m^{al}\, e^{-\delta(t-T_i)} & \text{if } t > \ell \end{cases}$$

The **response phase** is characterized by :

- Cutting off the arrival of **external events**
- Modulating the baseline intensity $\boldsymbol{\lambda_0}$ (through $\boldsymbol{\alpha_0}$ parameter) and the **self-excitation** component from past attacks (through $\boldsymbol{\alpha_1}$ parameter)
- Reducing the impact of future attacks (after $\boldsymbol{\ell}$) through $\boldsymbol{m^{al}}$

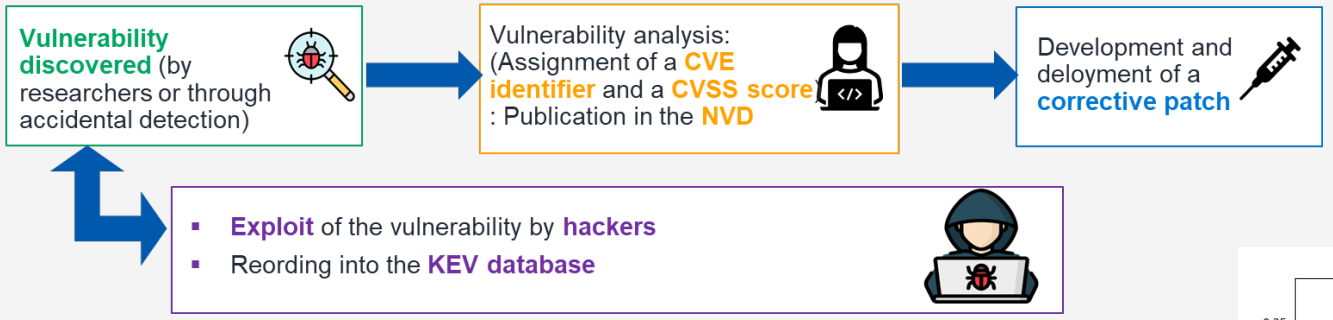




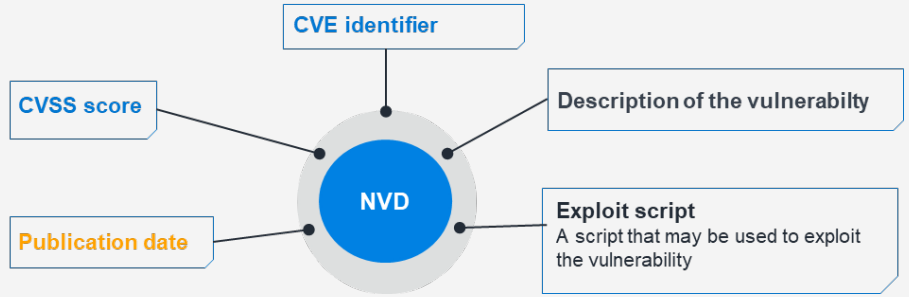
Cyber databases

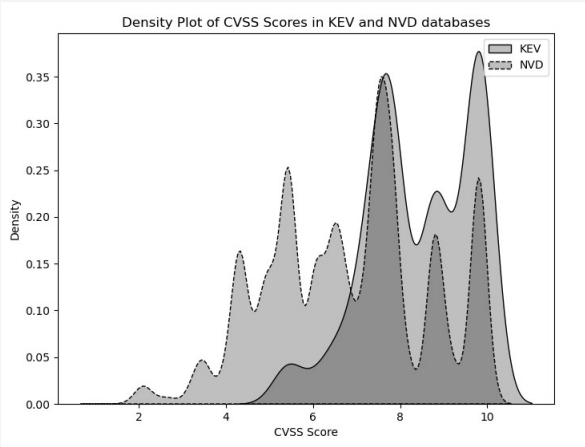
Vulnerability databases

- A cyber **vulnerability** is a flaw in an IT system that allows for **cyber attacks** or **unauthorized access**. These weaknesses can be due to mistakes in coding, wrong configurations, or not updating software properly.
- The **NVD** (National Vulnerability Database) is a US governoment database that lists **computer vulnerabilities**, while **KEV** (Known Exploited Vulnerabilities) lists **vulnerabilities that hackers are actively exploited**.



Life cycle of a computer vulnerability

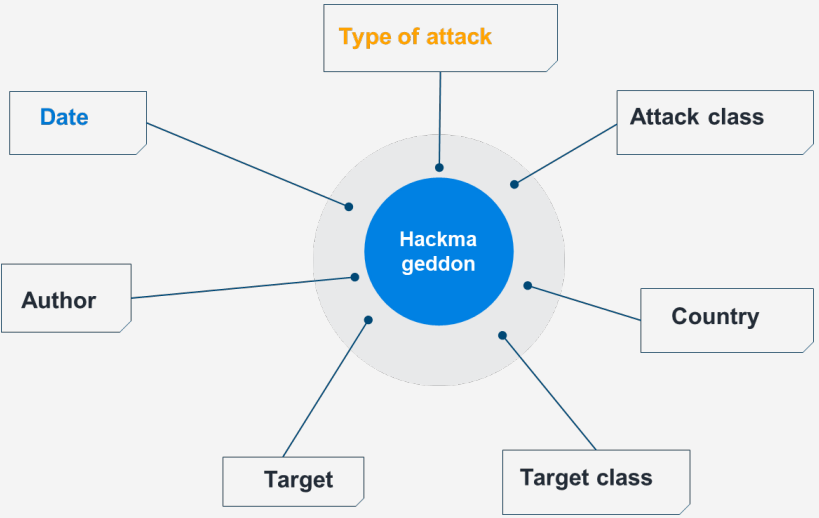




- **CVSS** scores **distribution** differ from one database to another : KEV (concentrated around 7) vs NVD
- Exploited vulnerabilities have a CVSS score above 5

Cyber databases

Hackmageddon database



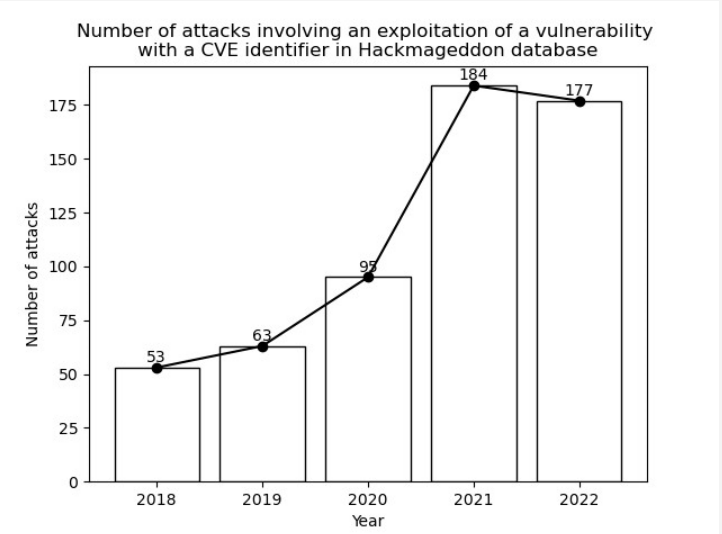
$$\lambda_t = \underbrace{\lambda_0}_{\text{Baseline intensity}} + \underbrace{\sum_{T_k<t} \overline{m}\, e^{-\delta(t-\overline{T_k})}}_{\text{External excitation : cyber vulnerabilities}} + \underbrace{\sum_{T_i<t} m\, e^{-\delta(t-T_i)}}_{\text{Self excitation : cyber attacks}}$$

CVE (Common Vulnerability Exposure) **publication dates**

Cyber attacks dates

- **Hackmageddon** database found at (https://www.hackmageddon.com/)
  - Gathered by **Paolo Passeri** according to his **expert judgment**
  - **Different incidents** types, among these, incidents the known **CVE identifiers**, which we link to **vulnerability databases**

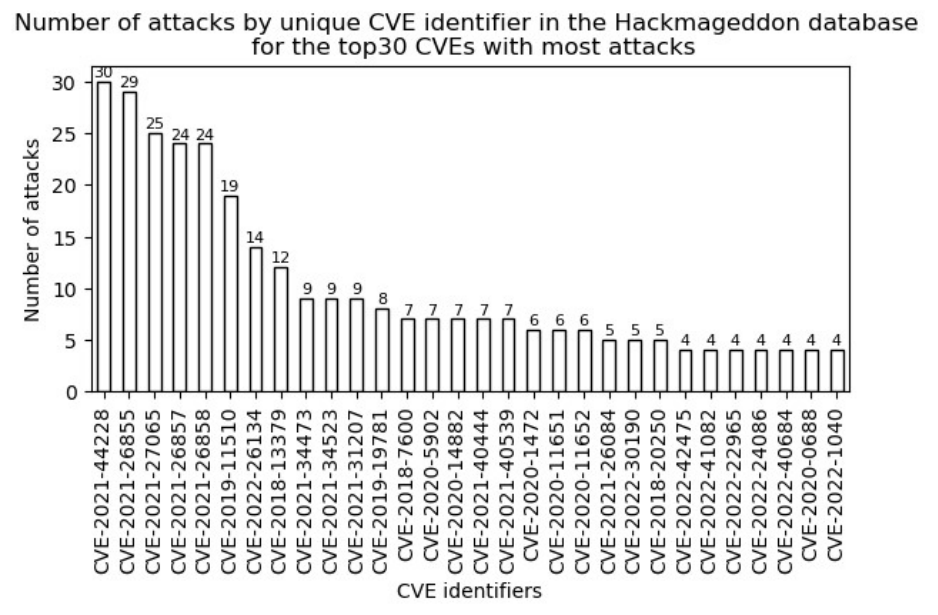
Different countries are represented in this database - The US is still the most represented



Growing number of attacks involving a **CVE identifier**

Cyber databases

Hackmageddon database and calibration configurations



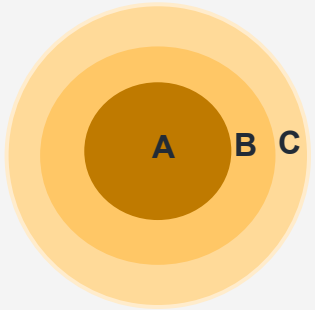
$$\lambda_t = \underbrace{\lambda_0}_{\text{Baseline intensity}} + \underbrace{\sum_{T_k<t} \overline{m}\, e^{-\delta(t-\overline{T_k})}}_{\text{External excitation : cyber vulnerabilities}} + \underbrace{\sum_{T_i<t} m\, e^{-\delta(t-T_i)}}_{\text{Self excitation : cyber attacks}}$$

CVE (Common Vulnerability Exposure) **publication dates**

Cyber attacks dates

Three **vulnerability** databases configurations :

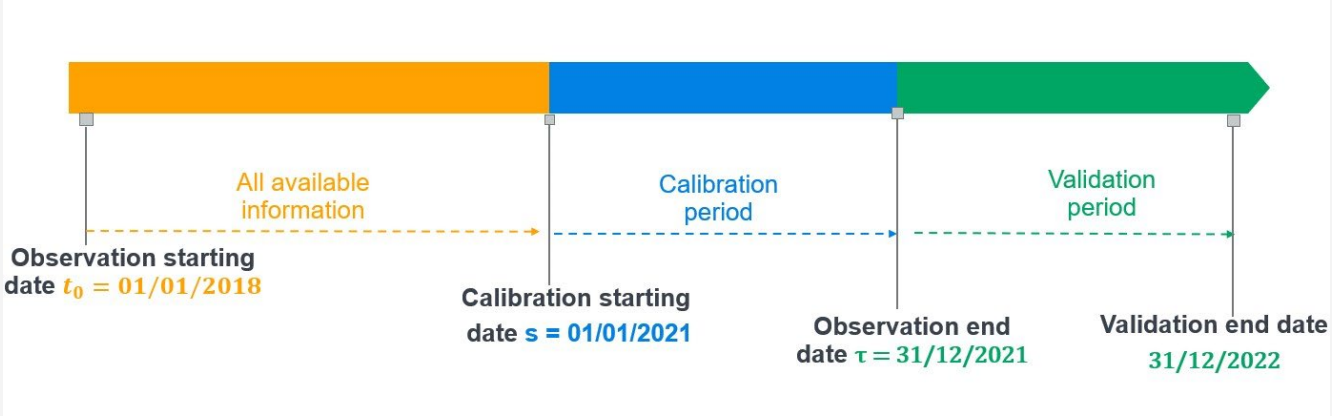
A. **Hackmageddon**
B. **KEV**
C. **NVD**



- **Log4Shell vulnerability** is the most represented in the **Hackmageddon database**

Calibration of the one-phase Hawkes process

Calibration periods



Year	Nb. of attacks	Nb. of Hackmaged don vuln.	Nb. of KEV vulnerabilit ies	Nb. of NVD vulnerabilit ies
2018	1310	31	63	8402
2019	1788	62	102	11932
2020	2321	63	125	15979
2021	2628	128	175	17829
2022	2649	91	113	22288

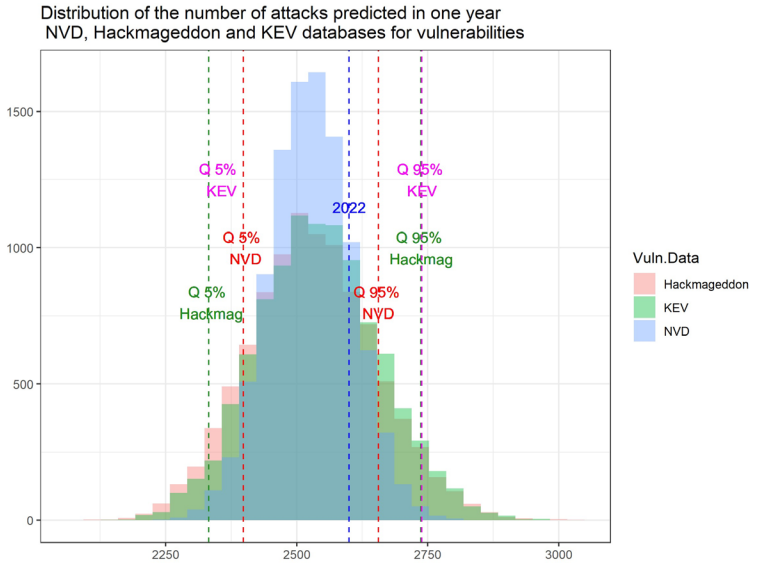
- Cyber risk is **rapidly evolving** :
  - **Calibration** is done on **2021** knowing **historical events** from **2018 to 2021**
  - **Validation** is done on **2022**
  - **Three vulnerability** configurations are **compared** one with the other

- Dates of **Hackmageddon CVE** Vulnerabilities are retrieved from the **NVD database**
- The **KEV database** contains all known **exploited** vulnerabilities
- The **NVD database** contains **all** known **vulnerabilities**

Calibration of the one-phase Hawkes process

Calibration results

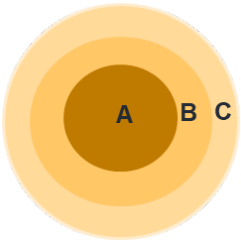
Model	Vuln. database	λ_0	ρ	$\overline{m}$	m	δ	$\|\|\phi\|\|$
No external events	-	2.7031	-	-	0.9182	1.5047	0.61
	95% C.I	[2.4863,2.9199]	-	-	[0.8608, 0.9756]	[1.1723, 1.8371]	-
With external events	Hackmageddon	2.7081	0.3636	0.5941	0.8891	1.5080	0.58
	95% C.I	[2.4873,2.9289]	[0.3180, 0.4092]	[0.3484, 0.8398]	[0.6909, 1.0873]	[1.1649, 1.8511]	-
With external events	KEV	2.6964	0.5057	0.9774	0.8529	1.5061	0.56
	95% C.I	[2.4229, 2.9699]	[0.4527, 0.5587]	[0.4388, 1.2282]	[0.6734, 1.1048]	[1.1921, 1.8239]	-
With external events	NVD	2.4195	48.849	0.077413	0.67139	1.8697	0.36
	95% C.I	[2.1573,2.6817]	[48.2987,49.1993]	[0.01211,0.1427]	[0.4985,0.8442]	[1.3998,2.3396]	-



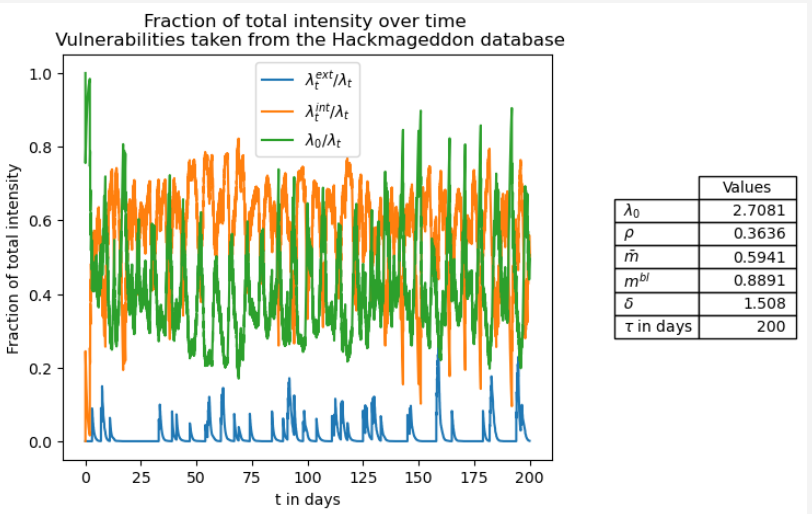
- $\| \phi \|$ (the **endogeneity degree** of the system) represents the **average number** of attacks **an attack** will lead to.
- $\| \phi \|$ is nearly **halved** between the **model with no external excitation** and the model with the **external excitation taken from the NVD database.**
- The distributions seem to **capture the dynamics of cyber attacks in 2022** for the **Hackmageddon** database.
- The **distribution of the number of attacks** with vulnerabilities from the **NVD** database has the **smallest variance**.
- This **decrease in variance** has significant implications in **insurance reserve calculations**, for example.

Calibration of the one-phase Hawkes process

Calibration results

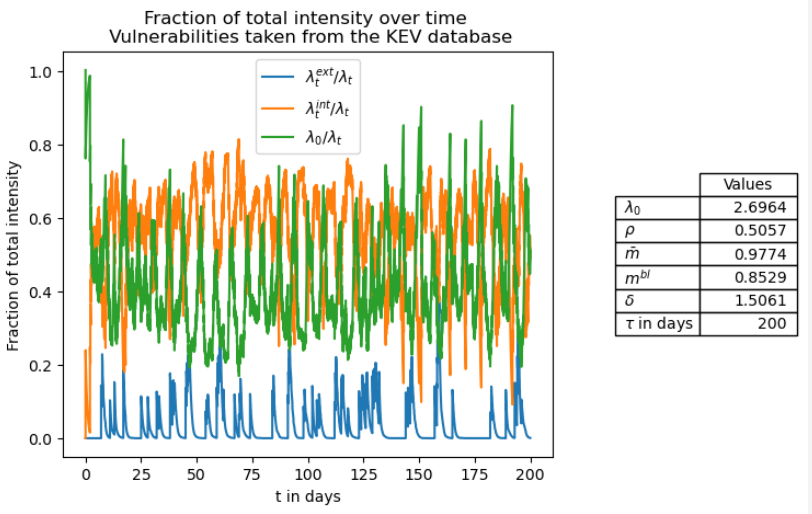


A : Hackmageddon



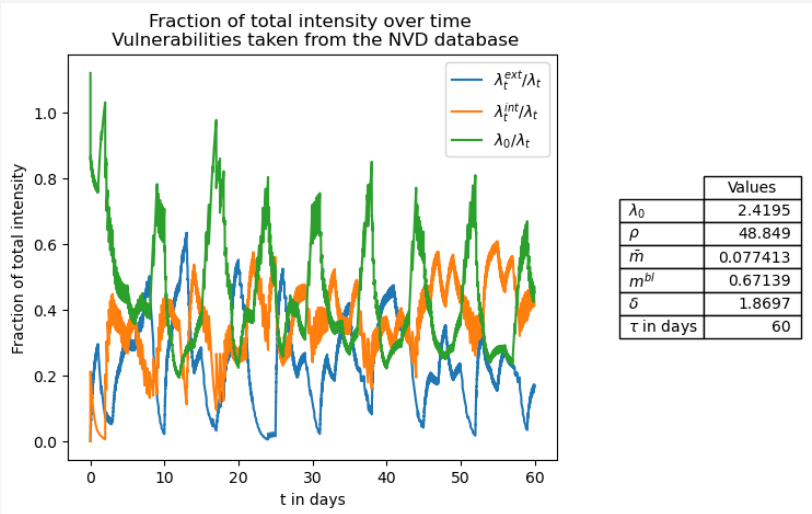
	Values
λ_0	2.7081
ρ	0.3636
$\bar{m}$	0.5941
m^{bl}	0.8891
δ	1.508
τ in days	200

B : KEV



	Values
λ_0	2.6964
ρ	0.5057
$\bar{m}$	0.9774
m^{bl}	0.8529
δ	1.5061
τ in days	200

C : NVD



	Values
λ_0	2.4195
ρ	48.849
$\bar{m}$	0.077413
m^{bl}	0.67139
δ	1.8697
τ in days	60

- The **fractions of intensity** attributed to **external, internal, and baseline** components are plotted.
- This **breakdown of the intensity** of the attacks process helps us **determine what is driving the intensity** of the Hawkes process and where the **observed attacks originate from**.
- This **decomposition** also allows for **selecting the appropriate response strategy** by activating the **appropriate measures** to **mitigate the number of attacks,** depending on **whether the threat is endogenous or exogenous.**
- **A and B** configurations are more **endogenous** than the **C configuration** where the **exogenous component is more pronounced**, meaning that a **significant portion of the excitation comes from the arrival of vulnerabilities**.

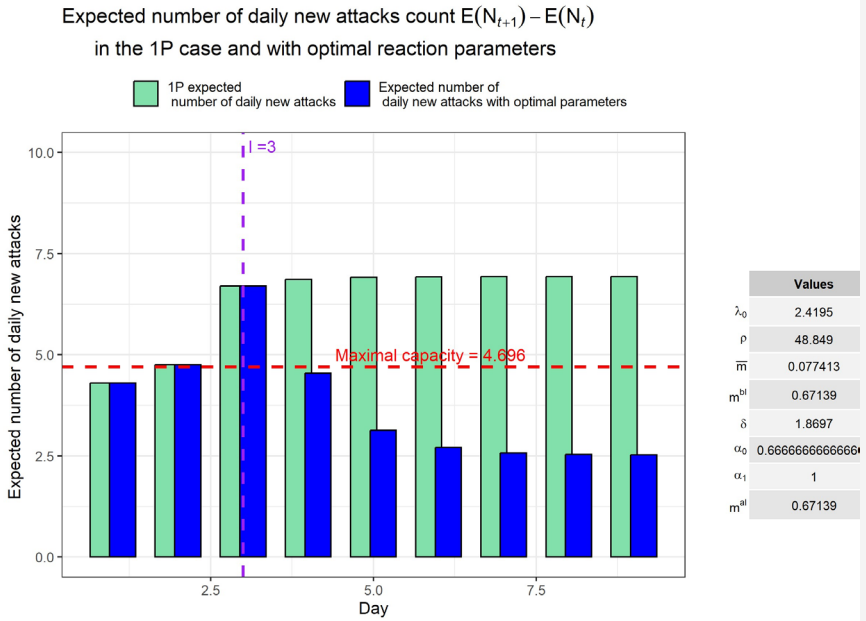
Response measures using the second phase of the process

Parameters selection

For $t > \ell > s$:

$$\mathbb{E}[N_t|\mathcal{F}_s] = \begin{cases} \mathbb{E}[N_\ell|\mathcal{F}_s] + \frac{\alpha_0\delta\lambda_0}{2}(t-\ell)^2 + \lambda_0(\alpha_0-\alpha_1)(t-\ell) + \alpha_1\mathbb{E}[\lambda_{\ell^-}|\mathcal{F}_s](t-\ell) & \text{if } \delta = m^{al} \\ \mathbb{E}[N_\ell|\mathcal{F}_s] + \frac{\alpha_0\delta\lambda_0}{\delta-m^{al}}(t-\ell) + \left((\alpha_0-\alpha_1)\lambda_0 + \alpha_1\mathbb{E}[\lambda_{\ell^-}|\mathcal{F}_s] - \frac{\alpha_0\delta\lambda_0}{\delta-m^{al}}\right)\frac{1}{(\delta-m^{al})}\left(1-e^{-(\delta-m^{al})(t-\ell)}\right) & \text{if } \delta \neq m^{al} \end{cases}$$

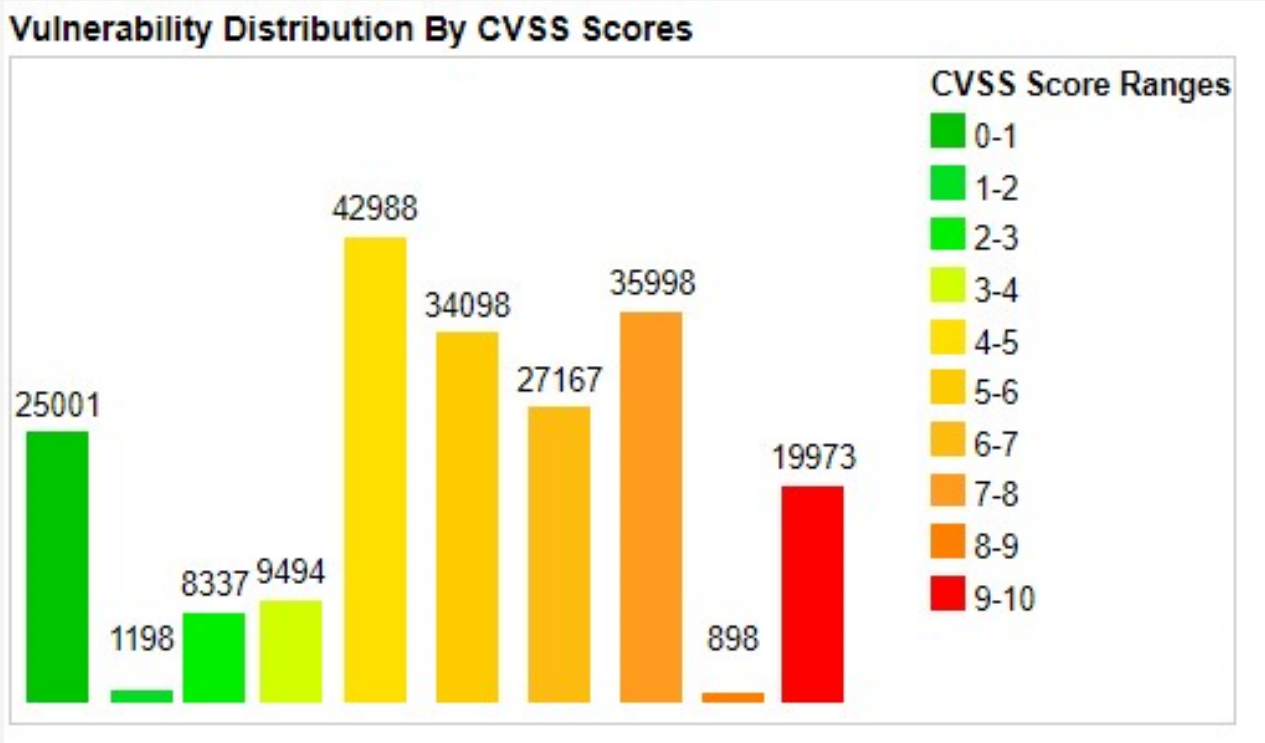
- **Fictional insurer** with a **limited reaction capacity** of 5 policyholders each day
- Compute **the adequate response parameters** such that the **response capacity** is **not exceeded on average**



	Values
λ_0	2.4195
ρ	48.849
$\bar{m}$	0.077413
m^{bl}	0.67139
δ	1.8697
α_0	0.6666666666666
α_1	1
m^{al}	0.67139

Future research questions

Paper available at :



- Extension to the **delay kernel** and **random marks**
- Develop **statistical classification and regression models** (such as CART trees) whose **classification criterion is based on the excitation of Hawkes processes**

Disclaimer

This presentation presents information of a general nature. It is not intended to guide or determine any specific individual situation and Milliman recommends that users of this presentation will seek explanation and/or amplification of any part of the presentation that they consider not to be clear. Neither the presenter nor the presenter's employer shall have any responsibility or liability to any person or entity with respect to damages alleged to have been caused directly or indirectly by the content of this presentation. All persons who choose to rely in any way on the contents of this presentation do so entirely at their own risk.

The contents of this presentation are confidential and must not be modified, copied, quoted, distributed or shown to any other parties without Milliman's prior written consent.