





Article

Measuring the Risk of Vulnerabilities Exploitation

Maria de Fátima Brilhante ^{1,2} , Dinis Pestana ^{2,3,4,*} , Pedro Pestana ^{5,6}  and Maria Luísa Rocha ^{7,8} 

- ¹ Faculdade de Ciências e Tecnologia, Universidade dos Açores, Rua da Mãe de Deus, 9500-321 Ponta Delgada, Portugal; maria.fa.brilhante@uac.pt
- ² Centro de Estatística e Aplicações, Universidade de Lisboa (CEAUL), Campo Grande, 1749-016 Lisboa, Portugal
- ³ Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- ⁴ Instituto de Investigação Científica Bento da Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250-012 Lisboa, Portugal
- ⁵ Departamento de Ciências e Tecnologia, Universidade Aberta, Rua Almirante Barroso 38, 1000-013 Lisboa, Portugal; pedro.pestana@uab.pt
- ⁶ Centro de Investigação em Ciência e Tecnologia das Artes (CITAR), Rua de Diogo Botelho 1327, 4169-005 Porto, Portugal
- ⁷ Faculdade de Economia e Gestão, Universidade dos Açores, Rua da Mãe de Deus, 9500-321 Ponta Delgada, Portugal; maria.ls.rocha@uac.pt
- ⁸ Centro de Estudos de Economia Aplicada do Atlântico (CEEApLA), Rua da Mãe de Deus, 9500-321 Ponta Delgada, Portugal
- * Correspondence: ddpestanda@ciencias.ulisboa.pt
- † Current address: Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal.

Abstract: Modeling the vulnerabilities lifecycle and exploitation frequency are at the core of security of networks evaluation. Pareto, Weibull, and log-normal models have been widely used to model the exploit and patch availability dates, the time to compromise a system, the time between compromises, and the exploitation volumes. Random samples (systematic and simple random sampling) of the time from publication to update of cybervulnerabilities disclosed in 2021 and in 2022 are analyzed to evaluate the goodness-of-fit of the traditional Pareto and log-normal laws. As censoring and thinning almost surely occur, other heavy-tailed distributions in the domain of attraction of extreme value or geo-extreme value laws are investigated as suitable alternatives. Goodness-of-fit tests, the Akaike information criterion (AIC), and the Vuong test, support the statistical choice of log-logistic, a geo-max stable law in the domain of attraction of the Fréchet model of maxima, with hyperexponential and general extreme value fittings as runners-up. Evidence that the data come from a mixture of differently stretched populations affects vulnerabilities scoring systems, specifically the common vulnerabilities scoring system (CVSS).

Keywords: cybervulnerabilities; CVSS metrics; vulnerabilities lifecycle; heavy-tailed models; extremes and thinned extremes; risk management

MSC: 68M25; 62G32



Citation: Brilhante, M.d.F.; Pestana, D.; Pestana, P.; Rocha, M.L. Measuring the Risk of Vulnerabilities Exploitation. *AppliedMath* **2024**, *4*, 20–54. <https://doi.org/10.3390/appliedmath4010002>

Received: 7 September 2023

Revised: 18 November 2023

Accepted: 14 December 2023

Published: 24 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A vulnerability is a weakness, for instance a design flaw or an implementation bug, that can be exploited to threaten a computer system, resulting in the stealing of data, installation of malware, or the running of malicious code. Cybervulnerabilities are a major economic concern. Their severity and the risk of consequent exploitation guide the prioritization of those needing urgent patching or remediation. From 2005 onwards, the common vulnerabilities scoring system (CVSS) [1], using base, temporal, and environmental metrics, has been a standard approach for assessing the severity of vulnerabilities, particularly in the banking sector, although its usefulness is compromised, since the use of temporal and environmental metrics is optional. Moreover, the fact that it is static is subject to

harsh criticism, since it is clear that risk changes over time. Modeling the vulnerabilities lifecycle and exploitation frequency are at the core of security of networks evaluation. Pareto, Weibull, and log-normal models have been widely used to model exploit and patch availability dates, the time to compromise a system, the time between compromises, and the exploitation volumes. As malware rarely causes significant damage, propagates at high speed, and is difficult to remove, the risk associated for most vulnerabilities is low, while there can be many missing values in the records, compromising the validity of the analysis of data.

As the existing data arguably result from drastic censoring and eventual thinning, thus representing top- or bottom-order statistics, other heavy-tailed distributions in the domain of attraction of extreme value or geo-extreme value laws are investigated as suitable alternatives to the Pareto model, power law models, and log-normal models.

Section 2 sketches some of the main features of vulnerabilities threats, and of the main scoring systems, including the CVSS and the exploit prediction scoring system (EPSS), evaluating their severity and the consequent risk of exploitation. The sophisticated machine learning algorithm, EPSS, for computing the probability of a vulnerability being exploited in the wild, uses more than one thousand variables in its algorithm black box, as well as the number of days since disclosure. The time variables in the lifecycle of vulnerabilities are crucial to understand the exposure risk of a network, and a main feature of this work is a critical appraisal of the goodness-of-fit of heavy-tailed distributions to the time from publication to the update of vulnerabilities.

In Section 3, progress in the statistical analysis of heavy-tailed empirical data from variables arising in the management of vulnerabilities and the associated risk and remediation are analyzed. The detailed discussion provided by Frei et al. [2], Holm [3], and Allodi [4] highlight the modeling of several time variables in the lifecycle of vulnerabilities, while Ruohonen [5] and references therein, serve as a basis for understanding the evolution of the state of the art in this matter.

In Section 4, several heavy-tailed laws for modeling the time variables in the lifecycle of vulnerabilities are discussed. The main idea is that extreme value models may have advantages over traditional power laws. On the other hand, as there exists strong evidence that thinning in the reporting of data certainly does occur, instead of use of the traditional general extreme value (GEV) model, or of the associated generalized Pareto model, use of the geo-max laws introduced by Rachev and Resnick [6], namely the log-logistic heavy-tailed law in the domain of attraction of the Fréchet law of long-tailed maxima, may represent a valuable approach. As the data exhibit some features typical of mixture models, and in light of the claims of Feldmann and Whitt [7] in favor of long-tail mixtures of exponentials, hyperexponential models are also discussed.

In Section 5, information on the process of obtaining the samples of vulnerabilities reported in [8] that are used in the analyzes is provided. Some issues concerning model fitting and parameter estimation, and the criteria for model choice, namely the Akaike information criterion (AIC) [9], the Schwarz–Bayesian information criterion (BIC) [10], and Vuong’s test [11], are briefly sketched.

In Section 6, the goodness-of-fit of the models under evaluation is compared with regards to the time from publication to the update of vulnerabilities sampled from 2021 and 2022. The general result, both for the 2021 systematic sample and for the 2022 simple random sample without replacement, is that power law fitting is inadequate, and that, as predicted from considerations on extreme values and thinning, the log-logistic model provides the best fit. On the other hand, a hyperexponential fit can be strongly recommended, since there is evidence that the data come from at least two differently stretched distributions.

Section 7 draws the conclusions from the statistical analysis, with suggestions provided regarding overcoming the static status of the CVSS, and discusses the evolution of metrology towards the consideration of virtual metrics.

In the Appendices, some further information on the CVSS metrics and evolution towards version 4.0, to be released soon, and on the Pareto laws is provided.

2. Vulnerabilities and Scoring Systems

Vulnerabilities are weaknesses that can be exploited by cybercriminals to access a computer system, with the intention of running malicious code, stealing sensitive data, and/or installing malware (Tunggal [12]), creating risks of possible intrusion and exploitation with potentially serious economic losses (Eling et al. [13]).

The first two vulnerabilities were reported in 1988. Nowadays, the Vulnerability Database [14] contains 236,311 vulnerabilities, with 25,227 vulnerabilities registered in 2022 in the National Vulnerability Database [15] from NIST, the National Institute of Standards and Technology. Information held in the Exploit Database [16] shows that until 7 August 2023, 5710 vulnerabilities exploitation codes had been detected, and that 33,750 had been verified. In Figure 1, a huge increase in the yearly disclosure of vulnerabilities can be observed, mainly occurring since 2017.

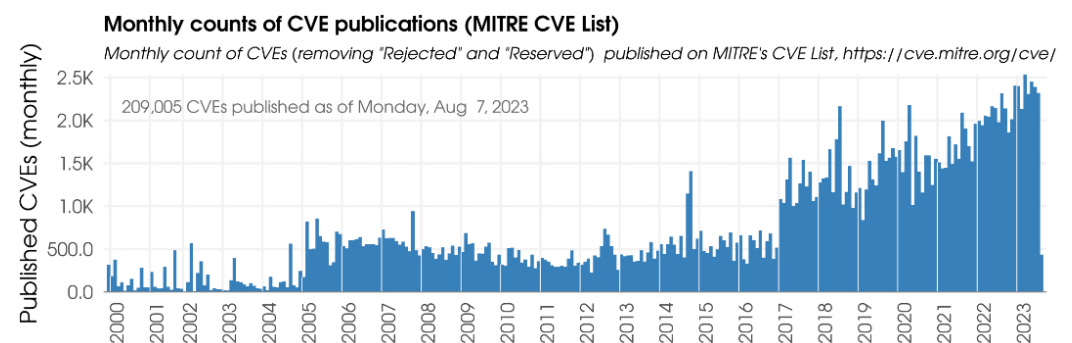


Figure 1. Monthly registered vulnerabilities (source: https://first.org/epss/data_stats (accessed on 7 August 2023). Copyright © 2023 Forum of Incident Response and Security Teams, Inc. All Rights Reserved).

The probability of a vulnerability being maliciously exploited and the impact of eventual exploits (meaning the likelihood of an intrusion occurring) are decisive in assessing the risk it represents. It is assumed that on a monthly basis from 5 to 20% of known vulnerabilities are corrected, while only 2 to 7% of vulnerabilities are actually exploited (Using data from [8], the percentage of exploited vulnerabilities is only 1.87%, with very high percentages observed between 2009 and 2012 (see Appendix C). However, apparently these data refer to vulnerabilities following assignment of a common vulnerabilities and exposures ID (CVE-ID), and the greatest risk occurs in the pre-disclosure period). Allodi [4] indicates that between 2009 and 2012, approximately 95% of the 75.7×10^6 attacks on 374 vulnerabilities targeted only 5% of them, which led him to comment that panic reactions are inappropriate, stating, "[...] most vulnerabilities may carry negligible risk. This indicates that the classical approach 'I have a vulnerability' → 'I must fix it' may be a largely disproportionate reaction to the real threat."

The percentage of vulnerabilities subject to cyberattacks seems low, but as nowadays more than 20,000 new vulnerabilities appear annually, it is anticipated that around 400 to 1700 of them may be subject to exploitation.

The eventual losses for companies, expenses associated with cybersecurity, and the existence of a flourishing market for instruments facilitating the exploitation of vulnerabilities, has drawn attention to the economic importance of the issue, with the potential for extreme negative outcomes (Eling et al. [13]). Therefore, it is essential to highlight the vulnerabilities that may potentially represent greater risks in order to prioritize those requiring urgent remediation or mitigation (which may not represent conclusive solutions), or, ideally, patching.

CVSS and Other Tools for Assessing the Severity and Risk of Vulnerabilities

Vulnerabilities have a strong bearing on the security level of networks, and CVSS has been widely used to assess the severity rating of security risks since version 1 appeared in 2005.

CVSS version 3.1 (released in June 2019) considers a base metric group (exploitability metrics and impact metrics) of eight nominal variables, a temporal metric group of three nominal variables, and an environmental metric group of 11 nominal variables, producing string vectors (check the CVSS v3.1 calculator at [1]). Each metric category has a corresponding numerical value, and a score is produced using the CVSS v3.1 equations (see details in Appendix A). Although the number of possible string vectors, 191,102,976,000, is very large, due to roundup procedures and limitation of the range to [0,10], only 101 scores are theoretically possible, leading to a qualitative severity rating scale considering the level of risk as follows: None \leftrightarrow 0, Low \leftrightarrow [0.1, 3.9], Medium \leftrightarrow [4.0, 6.9], High \leftrightarrow [7.0, 8.9], and Critical \leftrightarrow [9.0, 10.0].

The ‘High’ and ‘Critical’ severity ratings indicate vulnerabilities that can have a significant economic impact. Papers on the economic consequences of vulnerability exploits, such as Anderson [17], Bollinger [18], Dubendorfer et al. [19], and Kannan and Telang [20], had surely some bearing on moving from the 2005 Qualys [21] qualitative evaluation of the laws of vulnerabilities towards the quantitative score assessment of CVSS. With regard to the economic effect of vulnerabilities, the paper by Anderson and Mood [22] appeared before the release of CVSS v2 in 2007, while Frei’s comprehensive doctoral dissertation [23] had already been submitted in 2009.

Although CVSS scores are widely used to assess the severity of vulnerabilities, it must be noted that the optional temporal and environmental metrics are seldom used. Moreover, the temporal metrics do not reflect the fact that risk associated with vulnerabilities changes over time. As the optional temporal metrics of CVSS v3.1 did not effectively impact the CVSS score, v4.0 (to be released later in 2023) substitutes these with threat metrics. Additional attributes for vulnerability response are also introduced in v4.0 (cf. Appendix B). The effect of the temporal metrics equation in CVSS v3.1 is to produce a temporal score $TS = RU_p(BS \times ECM \times RL \times RC)$, with BS denoting the basic score, ECM, the exploit code maturity, RL, the remediation level, RC, the report confidence, and RU_p , the ceiling decimal round up, which can reduce the base score by up to 79.5334% (cf. Appendix A). As a consequence, other methodologies to evaluate the vulnerability severity and ensuing exploitation risk have been devised. Mention should also be made of other tools described below.

EPSS: EPSS registers the probability that a vulnerability will be exploited, with an additional indication of the corresponding percentile. The calculation of these probabilities is performed by an algorithm based on machine learning using 1164 variables covering a wide spectrum of information, including the CVSS base score and the number of days that have elapsed since disclosure (see detailed description in [24], and in Jacobs et al. [25]). The training for version 2 used gradient boosting and Poisson regression. Version 3 was released in March 2023.

The probability indicated by EPSS is a guide to identifying vulnerabilities that should be prioritized in the remediation effort. Note that, since it is a probability, false negatives and false positives must be taken into account.

The strengths of EPSS are its ability to calculate the probability of risk of a set of vulnerabilities, and consequently the probability that a company is at risk of exploitation when using software subject to a set of vulnerabilities. In addition, the fact that the calculation is dynamic, which is a consequence of the algorithm being automatically updated taking into account new information, represents a further advantage. For example, vulnerability CVE-2021-44228 (CVSS score 10), published on 10 December 2021, started with an exploit probability estimated by EPSS at 0.355 on 11 December. On 13 December, it became 0.384, on 14 December, 0.300, on 15 December, it increased to 0.633, on 17 December to 0.944, on 18 December, it dropped to 0.633, but on 12 January 2022, it increased again to 0.944.

However, the EPSS only exists for previously published vulnerabilities—there is no calculator that can be used in the CVE-ID discovery/reservation phase as exists for the CVSS.

An efficiency comparison of CVSS v3.1, EPSS v1, and EPSS v2 considering the November 2021 vulnerabilities showed, respectively, scores greater than 8.6, probabilities greater than 0.066, and probabilities exceeding 0.149, in order to provide coverage of around 50% in the three scenarios, as described in the documentation on the EPSS model. These thresholds led to signaling for corrections of 25.3%, 9.3%, and 4.7% of the vulnerabilities, respectively.

With regard to CVSS v3.1, there were attempts to exploit 5% of the vulnerabilities identified, while with EPSS v1, the efficiency was 12.9%, and with EPSS v2, it was 42.5%, as shown in Figure 2.

It is, thus, observed that EPSS compares favorably with CVSS, and has also evolved significantly from version 1 to the current version 3 (the team has ambitions to provide quarterly updates of the algorithm).

- CTI: Cyber threat intelligence [26], which is one of the indexes used in the excellent database [14] for daily updating of risk, takes into account, on the one hand, alarms and attacks, and, on the other hand, the existence of corrections, as well as the economic impact.
- CWSS: The common weakness scoring system [27], which has a well-documented methodology for classifying “weaknesses”, should be taken into account in any normalization project assessment of the severities of vulnerabilities and the risks of malicious exploitation.
- SSVC: The stakeholder-specific vulnerability categorization system [28] offers the cyber community a vulnerability analysis methodology that can account for a vulnerability’s exploitation status, impacts on safety, and the prevalence of the affected product in a singular system. In 2020, a customized SSVC decision tree to examine relevant vulnerabilities was developed.

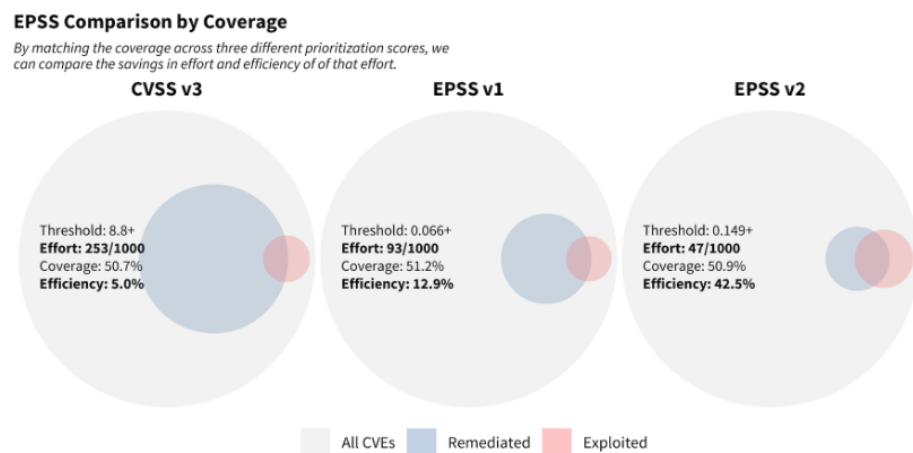


Figure 2. Efficiency of CVSS and EPSS (source: <https://www.first.org/epss/model> (accessed on 1 September 2023) Copyright© 2023 Forum of Incident Response and Security Teams, Inc. All Rights Reserved).

3. Vulnerabilities Lifecycle Management and the Importance of Modeling Inter-Occurrence Times

As the risk resulting from vulnerabilities changes over time, statistical modeling of time variables is relevant. As mentioned above, the number of days elapsed since disclosure of the vulnerability is one of the 1164 variables that the EPSS algorithm uses to compute the probability of exploitation in 30 days. Three important steps highlighted by Frei at al. [2], Holm [3], and Allodi [4] towards using appropriate model fitting to various time variables in the lifecycle of vulnerabilities are described. The main purpose was to devise effective algorithms to produce a dynamic CVSS score. Frühwirth and Männistö [29] proposed interrogative rules to change the settings in the temporal metrics, using the Frei at al. [2],

Pareto, and Weibull fittings. The aim of this work was to model the time from publication to update using samples from the 2021 and 2022 vulnerabilities, as published in [8].

3.1. Vulnerabilities Lifecycle: Can Pareto and Weibull Models Provide a Dynamic CVSS?

To understand the security risks inherent in the use and operation of today's large and complex information and communication systems, analysis of vulnerabilities' technical details alone is not sufficient. To assess the risk exposure of a network, one has to know and understand the lifecycle of vulnerabilities and their evolution.

Frei, May and Plattner [2]

The lifecycle of vulnerabilities, their discovery, disclosure, and patching, and of eventual exploits, has been thoroughly investigated by Frei et al. [2] (see also Frei [30] and Shahzad et al. [31]). In Figure 3, the lifecycle of vulnerabilities considered in Frei et al. [2] is shown.

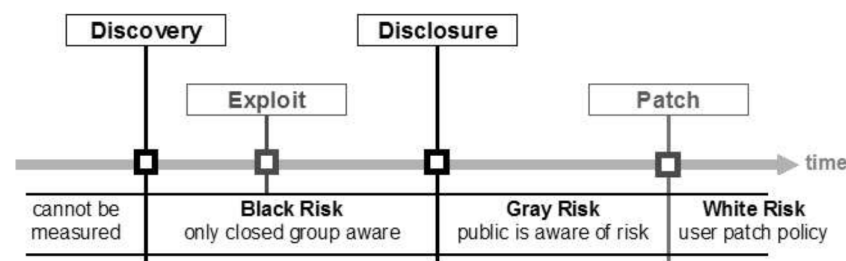


Figure 3. Vulnerabilitieslifecycle (source: Frei et al. [2]).

The strict definition of the time variables adopted in Frei et al. [2] will be used here. The time of disclosure is the first date a vulnerability description, including the risk rating information, such as its CVSS score, is freely available to the public via an independent trusted channel. The disclosure date of a vulnerability frequently occurs after the discovery date, which often will never be known by the public. Only a closed group is aware of the vulnerability from discovery to disclosure, and this is a period of major risk for exploits—any hacker-tool or intrusion can take advantage of the vulnerability. Two other important variables are the time of the exploit, the earliest date an exploit for a vulnerability is detected, and the time of patch availability, the earliest date the vendor or the originator of the software releases a fix, workaround, or a patch that provides protection against the exploitation of the vulnerability. As the public becomes aware of the vulnerability after disclosure, it is assumed that the risk of an exploit after disclosure is less than that before disclosure.

Frei et al. [2] investigated the lifecycle of vulnerabilities in the period 1996–2006. After systematically examining more than 80,000 advisories from publicly available sources, they gathered 14,326 vulnerabilities with non-empty common vulnerabilities and exposures (CVE) [32], entry and non-empty disclosure dates, 11,697 containing a discovery date, 3420 with an exploit date, and 1551 with a patch date, while 332 vulnerabilities had both an exploit and a patch date.

This is a clear indication that the records fail to consistently register the variables Frei et al. [2] suggest need modeling, since many vulnerabilities carry medium or low risk, and the exploits did not exist or had no relevant consequences. Although medium- and low-risk vulnerabilities are recorded, complete information on the discovery time, the disclosure date, the patch time availability, the exploit time, and on the frequency at which vulnerabilities are exploited, is often missing—as expected, detailed information exists when the attack processes have far-reaching consequences. In fact, there exists a probability π_1 that a negligible vulnerability is never discovered, a probability π_2 that a discovered vulnerability is never disclosed, a probability π_3 that it is never exploited, and so on, so

that the data useful for analysis result from a drastic thinning process. It may, therefore, be assumed that the data available for analysis are, to a certain extent, tied to extreme consequences, and that it is reasonable to consider that heavy-tailed models are useful in vulnerability risks management. On the other hand, as some vendors tend to postpone disclosure until a patch for the vulnerability can soon be expected, the empirical data for the time from disclosure to patch availability may eventually be reasonably modeled by an extreme value distribution of minima, such as the Weibull.

After a thorough study of the discovery date vs. the disclosure date, of the exploit availability date vs. the disclosure date (showing that the number of zero-day exploits is increasing drastically, and that exploits may occur at any time, before or after disclosure), the patch availability date vs. the disclosure date, and the gap between the exploit availability and the patch availability, Frei et al. [2] split their data at the disclosure date, and limited their model choice to fit the exploit availability and the patch availability in the years 2001–2005:

- Pareto (type I), in the domain of attraction of the Fréchet model for maxima;
- Weibull, with the parametrization used as the extreme values law for minima;
- Exponential, which is the Weibull(1, δ) of minima.

They concluded that Pareto I (with a shape parameter of less than 1; thus, with no expectation, and indicating a very heavy tail) provided the best fit for exploit availability, both before and after disclosure, and to patch availability before disclosure, while, after disclosure, the best fit to patch availability was Weibull.

However, there is no apparent reason to limit model choice to these three models, and the above comments on the availability of data shed some doubt on the adequacy of the inference that Pareto and Weibull were the best fits. Namely, with regard to the Weibull fitting to the patch availability after disclosure, the estimated shape parameter ranged from 0.199 in 2001 to 14.04 in 2004; this is strong reason to consider that alternative models should be investigated.

Frühwirth and Männistö [29] proposed a change in the selection of options in the temporal metrics, the exploit code maturity, and the remediation level. It is curious that of the three temporal metrics, changes were only proposed to the first two, since only 12 years later Boechat et al. [33] raised the issue of the redundancy of the CVSS temporal metrics, discussing how one can infer attribution of the report confidence from the exploit code maturity and the remediation level.

With regard to the exploit code maturity metric, Frühwirth and Männistö [29] consider that the probability of having malicious intrusion should be estimated by the overall Pareto fit proposed by Frei et al. [2] $F_P(x) = 1 - \left(\frac{k}{x}\right)^\alpha$, $\alpha = 0.26$, $k = 0.00161$, where x is the time elapsed after vulnerability discovery. This value is compared to a generated pseudorandom uniform number *rand*. If $rand \leq F_P(x)$, set the Exploit Code Maturity \leftrightarrow High, if $rand > F_P(x)$, set the Exploit Code Maturity \leftrightarrow Unproven.

With respect to the remediation level metric, the patch availability is calculated by comparing a pseudorandom uniform number *rand* with the overall Weibull fit of Frei et al. [2], $F_W(x) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$, $\lambda = 0.209$, $k = 4.04$. If $rand \leq F_W(x)$, set Remediation Level \leftrightarrow Official Patch, if $rand > F_W(x)$, set Remediation Level \leftrightarrow Unavailable.

The amendments proposed by Frühwirth and Männistö [29] are controversial. The decision rule of comparing the value of the distribution functions with a pseudorandom uniform number does not seem rational. With the passage of time x , $F_P(x)$, (respectively, $F_W(x)$) will take values close to 1, and, therefore, with high probability, the value of $rand \leq F_P(x)$, (respectively, $rand \leq F_W(x)$) and the decision to change the classification will be effective. However, even if a long time has passed, there is a probability $1 - F(x)$ that no change occurs. Furthermore, it is not explained why the decision considers only the extreme possibilities high/unproven (respectively, or unavailable/official fix), completely disregarding the intermediate options of functional/proof-of-concept (respectively, or workaround/temporary fix).

Moreover, if the objective is to have a dynamic CVSS, it would be necessary to compare daily the above distribution functions with a *rand* generated that day, which could lead to contradictory decisions over time, or, if one does not proceed in that way, avoiding paradoxical situations, one would fall into an unalterable alternative, also static. It is one of those situations in which chance does not seem like a reliable ally.

3.2. Time Required to Compromise a Computer System: gamma, log-normal, Pareto, and Weibull Models

Research on the topic of dependable computing is often based on assumptions regarding properties of the failures in systems, and knowledge concerning how system failures behave is, as such, critical. Choosing appropriate models for describing the number of failures in a system, and the time between failures, is of particular importance as employment of an inappropriate statistical model can result in inappropriate research conclusions.

Holm [3]

Holm [3] analyzed “5,602,097 malware alarms corresponding to 203,025 intrusions that have occurred across 261,757 computer systems of a large international enterprise between October 2009 and August 2012”.

At the time (access on 14 September 2012), the Symantec Online Encyclopedia reported that 71.32% of malware incidents caused medium damage, 86.55% were easy to remove, and 90.41% had a low propagation speed, while only 0.43% caused significant damage, with 0.14% being difficult to remove, and 5.52% having a high propagation speed.

With respect to the number of intrusions of a computer system, Holm [3] tested Poisson, normal, and log-normal models, concluding that the log-normal model provided the best fit, without presenting a rationale for considering the two continuous alternatives, instead of using other discrete models, such as a heavy-tailed Zipf or a Zipf–Mandelbrot model, or an overdispersed negative binomial model (see Appendix D).

Further, Holm’s [3] goal was to model the time to first compromise (TTFC), the time between compromises (TBC), and whether the TTFC was correlated (Spearman correlation) to the number of intrusions of a computer system.

To investigate the two former questions, he tested the exponential, gamma, log-normal, Pareto, and Weibull models, using the Kolmogorov–Smirnov, Cramér–von Mises, Anderson–Darling, Shapiro–Wilks, and chi-square tests, as well as QQ-plots (Wilk and Gnanadesikan [34]). Moreover, seeking not only goodness-of-fit but also simplicity, the AIC has been used to select the best fit (Akaike [9]; Burnham and Anderson [35]). Holm’s overall conclusion was that the Pareto distribution provided the best fit for the TTFC data, and that the log-normal model provided the best fit for the TBC data.

Extreme value distributions and their domain of attraction are discussed in Section 4, but a minor remark on the topic is necessary at this point: Although the log-normal distribution is on the domain of attraction of the Gumbel distribution, it exhibits Fréchet type pre-asymptotic behavior, and this is the reason why it appears frequently as a match for the Pareto distribution in the context of heavy-tail modeling.

It is interesting to observe that although Holm [3] reports that Nurmi et al. [36] studied the model fit for three datasets describing the time between failures, concluding that the hyperexponential distribution provided better fit than the Weibull or the Pareto distributions, he did not consider the hyperexponential model to be “more complex due to its additional parameters”, that, therefore, would be penalized by the AIC. Note, however, that the authoritative paper by Feldmann and Whitt [7] discusses in detail the hyperexponential fit to long-tail distributions in network performance models.

3.3. Frequency of Vulnerabilities Exploitation: Power Laws and Paretian Heavy-Tailed Models

The commonly used, industry standard definitions of vulnerability risk based on a single number (e.g., scores assigned by security-testing tools) may be incapable

of describing the distribution of attacks; a point estimate (a score or an average) is not sufficient to describe the phenomena and may lead to substantial overspending/misallocation of resources, as most events may be orders of magnitude away from the point estimate.

Allodi [4]

Allodi [4] collected data from Symantec's data sharing program, the Worldwide Intelligence Network Environment [37], using Symantec's security response, but his analysis comprised only 374 vulnerabilities occurring in 2009–2012 related to Symantec's consumer security products, and, therefore, he recognized that he may have had a self-selection problem.

Allodi [4] used Lorentz curves (Aaberge [38]) to support his claim that *"the distribution of attacks follows a heavily tail distribution"*.

The main objective of Allodi [4] was to substantiate that vulnerabilities exploitation follows a power law, in the loose sense that it has a linear signature to the right of an estimated $\delta = x_{\min}$. He closely followed the methodology proposed in Clauset et al. [39], whereby the estimate of x_{\min} is obtained *"by selecting the cutoff that minimizes the distance between the fitted Power Law distribution and the distribution of the data [...] calculated as the Kolmogorov–Smirnov (KS) statistic."* note, however, that estimating the parameters alters the critical values of the KS statistic. To estimate the scaling parameter, the maximum likelihood estimator $\hat{\alpha}$ was used. Using Vuong's test, described in Section 5, Allodi's conclusion was that the choice of fitting a power law vs. log-normal *"is only inconclusively supported by our evidence"*.

Although the methodology described in Allodi [4] regarding parameter estimation, hypothesis testing, and comparison with other models, in particular log-normal, is sound, there are important limitations, which arise from the data gathering process. In fact, using limited empirical data to choose a power law fit may cause overfitting (Clauset et al. [39]).

Further analysis of the issues relating to time variables in the lifecycle of vulnerabilities is provided in Ruohonen [5] and the references therein.

4. Methodological Issues: The Temptation of Power Laws and Other Heavy-Tailed Models

I can resist anything except temptation

Oscar Wilde, *Lady Windermere's Fan*

The fascinating books by Schroeder [40] and Sornette [41], and papers such as those by Newman [42], have contributed to popularizing power laws, functional relationships of the type $f(x) = \left(\frac{x}{\delta}\right)^{-\alpha}$ that seem to govern many natural phenomena and man-made activities. Since $f(cx) = c^{-\alpha}f(x)$, all power laws with the same scaling exponent α are equivalent up to a constant factor.

As $\ln(f(x)) = -\alpha \ln x + \alpha \ln \delta$, the straight line in a log-log plot is called the "signature" of the power law. Note, however, that straightness of the log-log plot is not a sufficient condition for fitting a power law to empirical data, since, for instance, the log-normal law can generate finite amounts of data displaying the signature behavior (cf. Belevitch [43], Mitzenmacher [44], and Bee [45]).

In Statistics, the Pareto probability density function (PDF), or Pareto type I, with the shape parameter $\alpha - 1$ is

$$f(x) = \frac{\alpha - 1}{\delta} \left(\frac{x}{\delta}\right)^{-\alpha} \mathbb{I}_{[\delta, \infty)}(x), \quad \alpha > 1, \quad \delta > 0,$$

and the Zipf probability mass function,

$$p_k = \frac{k^{-\alpha}}{H_{N, \alpha}}, \quad k \in \{1, 2, \dots, N\}, \quad \alpha > 1,$$

where $H_{N,\alpha} = \sum_{j=1}^N j^{-\alpha}$ is the N th generalized harmonic number, are widely used strict power laws, with their popularity being possibly due to both their straight signature and the very uneven equilibrium they model, which is radically different from uniformity. Note that the Pareto I law was introduced to model the distribution of wealth, reflecting the idea that a small percentage of owners possess a large percentage of wealth. The Zipf law models the frequency vs. the rank of words, on the assumption that the effort to communicate generates an uneven equilibrium of social vs. personal vocabularies, so that the frequency of word use declines steadily. With respect to the lifecycle of vulnerabilities, the evolution of hackers' abilities and the progress of counteracting measures has generated a dynamic equilibrium that justifies the traditional choice of modeling with power laws, which is radically different from uniformity. With respect to vulnerabilities and the risk they represent, this power law equilibrium arguably results from the progress of developers in avoiding design and implementation weaknesses being counterbalanced by the ability of hackers to attack or misuse operating systems.

In a broader sense, and in Statistics generally, a power law (or more precisely, a distribution with a Paretian upper tail) is a distribution whose upper tail $1 - F(x) = x^{-\alpha}L(x)$, with L a slowly varying function, i.e., $\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$, for all $t > 0$, forcing asymptotic scale invariance, whose form controls the shape and finite extent of the lower tail. Note that F is used here to generally denote the cumulative distribution function (CDF) of a random variable. Note also that if the function L is constant, assuming a lower bound δ for the support, the Pareto I strict power law is obtained.

In other words, the upper tail of the distribution is a regularly varying function in the sense of Karamata [46]. (See Feller [47], chapters 8 and 9, pp. 275–284, for the basics of regular variation and its use for the characterization of the domain of attraction in extreme value theory, and Kevei [48]). Detailed information is provided in Bingham et al. [49], and, with respect to the estimation of the exponent of regular variation, see Hall [50], Davis and Resnick [51], or Fedotenkov [52].

The apparent simplicity of power laws has resulted in their widespread use in a huge variety of fields, often assuming a shape parameter $\alpha < 3$, implying heavy tails.

But this widespread use of Paretian tail models can be abused, reflected in criticism by Stumpf and Porter [53] who stated, “A striking feature that has attracted considerable attention is the apparent ubiquity of power-law relationships in empirical data. However [...] the data are typically insufficient and the mechanistic insights are almost always too limited for the identification of power-law behavior to be scientifically useful”. In the same line of criticism, see also The Econophysics Blog [54] and Shalizi [55].

In Section 3, to summarize the state of the art in modeling several time variables in the lifecycle of vulnerabilities, as highlighted by Frei et al. [2], Holm [3], and Allodi [4], the statistical analysis of heavy-tailed empirical data from variables arising in the management of vulnerabilities and the associated risk and remediation will be thoroughly discussed. Overall, the Pareto I model

$$X \sim \text{Pareto I}(\alpha, \delta), \quad \alpha, \delta > 0 \Leftrightarrow F(x) = \begin{cases} 0 & , x < \delta \\ 1 - \left(\frac{x}{\delta}\right)^{-\alpha} & , x \geq \delta \end{cases} \quad (1)$$

was considered the best fit, with some support for the log-normal

$$X \sim \text{log-normal}(\mu, \sigma), \quad \mu \in \mathbb{R}, \sigma > 0 \Leftrightarrow F(x) = \begin{cases} 0 & , x < 0 \\ \frac{1}{2} \left[1 + \text{erf}\left(\frac{\ln x - \mu}{\sqrt{2}\sigma}\right) \right] & , x \geq 0 \end{cases}$$

where $\text{erf}(z) = \int_0^z e^{-t^2} dt$ is the error function.

However, these authors recognized problems in the data they used, such as censoring, the self-selection of data, and possible overfitting. On the other hand, the class of alternative models they considered was very restrictive.

Holm [3] introduced many improvements, namely, consideration of a wider set of candidates for model fitting, and the use of AIC and QQ plots. But it is questionable whether his choice of best fits—Pareto or log-normal—was restricted to a section of the range of observations where he could find a straight signature.

Allodi [4] considered a much wider class of model candidates—power laws, in a broader sense, that he expressed as “ $p(x) \approx x^{-\alpha}$, where x is the measured quantity and α is a scaling factor of the distribution”—hence, a regularly varying Paretian tail

$$X \curvearrowright \text{Power Law}(\alpha) \Leftrightarrow 1 - F(x) = x^{-\alpha} L(x),$$

where L is a slowly varying function. This is a broad class of heavy-tailed models. In fact, for each $\alpha > 0$, this is what characterizes the domain of attraction of a Fréchet- α distribution, as observed below.

Alongside this broadening of scope in model building, Allodi [4] brought in many methodological improvements, such as the use of Lorentz curves and of Vuong’s test for selecting the best among alternative fits. Allodi’s statistical analysis closely follows the Clauset et al. [39] methodology for modeling empirical data using power laws. It is interesting to observe, however, that Allodi recognized that there is inconclusive evidence that a Paretian model is the most adequate, possibly because the sample size ($n = 374$) he used falls in the category described by Stumpf and Porter [53] as “data are typically insufficient [...] for the identification of power-law behavior to be scientifically useful”.

As always, Box’s [56] wise warning “All models are wrong but some are useful” must be kept in mind, so the next step is to have some rationale basis for the choice of useful models, namely, a rationale for using extreme value models, or at least laws in their domains of attraction.

4.1. More on Power Laws, Regular Variation, and Extreme Value Models of Maxima

There are several generalizations of the Pareto I distribution (cf. Appendix D and Arnold [57]). Considering the relationship between regular variation and extreme value theory, which has increasing importance in finance (see Embrechts et al. [58], and Andriani and McKelvey [59]), the generalized Pareto distribution should be considered as well.

The CDF of a generalized Pareto random variable, $X_{\xi, \lambda, \delta} \curvearrowright \text{GPareto}(\xi, \lambda, \delta)$, with shape parameter $\xi \in \mathbb{R}$, location parameter $\lambda \in \mathbb{R}$, and scale parameter $\delta > 0$, has the form

$$GP_{\xi, \lambda, \delta}(x) = \begin{cases} 1 - \exp\left(-\frac{x-\lambda}{\delta}\right) & , \xi = 0 \\ 1 - \left(1 + \xi \frac{x-\lambda}{\delta}\right)^{-\frac{1}{\xi}} & , \xi \neq 0 \end{cases}, \quad \xi, \lambda \in \mathbb{R}, \delta > 0,$$

whose support has a left-endpoint λ if $\xi > 0$, and is $\left[\lambda, \lambda - \frac{\delta}{\xi}\right]$ if $\xi < 0$. For detailed information, consult the thorough monograph of Arnold [57] or Johnson et al. [60].

The relationship between the generalized Pareto and GEV laws, in the strict sense of laws of maxima of independent and identically distributed (IID) random variables, is simple: $GEV_{\xi, \lambda, \delta}(y) = \exp(GP_{\xi, \lambda, \delta}(y) - 1)$, with parameters $\xi, \lambda \in \mathbb{R}, \delta > 0$. Therefore,

$$GEV_{\xi, \lambda, \delta}(y) = \begin{cases} \exp\left(-\exp\left(-\frac{y-\lambda}{\delta}\right)\right) & , \xi = 0 \\ \exp\left(-\left(1 + \xi \frac{y-\lambda}{\delta}\right)^{-\frac{1}{\xi}}\right) & , \xi \neq 0, \xi \frac{y-\lambda}{\delta} > -1 \end{cases},$$

with a left-endpoint $\lambda - \frac{\delta}{\xi}$ if $\xi > 0$ or a right endpoint $\lambda - \frac{\delta}{\xi}$ if $\xi < 0$, is the CDF of GEV random variables of maxima, $Y_{\xi, \lambda, \delta} \curvearrowright GEV(\xi, \lambda, \delta)$, i.e., stable for maxima. For general information, consult Johnson et al. [60], Beirlant et al. [61], or see the overview in Gomes and Guillou [62].

The aim here is to fit a heavy-tail law to empirical data, concentrating on the case $\xi > 0$, and without loss of generality, dealing with the standardized form $\lambda = 0$ and $\delta = 1$. Rewriting $\xi = \alpha^{-1}$ and $1 + \xi y = z$, the standard CDF expression Φ_α ,

$$\Phi_\alpha(z) = \begin{cases} 0, & z < 0 \\ e^{-z^{-\alpha}}, & z \geq 0 \end{cases},$$

of the Fréchet extremal stable law, sometimes called a type II value distribution, is obtained. But this is an inadequate terminology in the context of the convergence of classes in the asymptotic theory of extremes that uses the concept of Khinchine's types (see [47], p. 137), i.e., location-scale families $F(\lambda + \delta x)$, $\lambda \in \mathbb{R}$, $\delta > 0$. Therefore, for each $\alpha > 0$, a different Fréchet type is obtained.

Observe also that the Pareto- α distribution results from truncation of the series expansion of the Fréchet- α distribution:

$$e^{-z^{-\alpha}} = \sum_{k=0}^{\infty} \frac{(-z^{-\alpha})^k}{k!} = \underbrace{1 - z^{-\alpha}}_{\text{Pareto-}\alpha} + \frac{z^{-2\alpha}}{2} - \dots$$

A random variable X with CDF F_X is said to be in the domain of attraction of a Fréchet- α \mathcal{Y}_α random variable if, for all $n \in \mathbb{N}$, there exist attraction coefficients $A_n > 0$ and $B_n \in \mathbb{R}$ such that, denoting by $X_{n:n} = \max(X_1, \dots, X_n)$ of n independent copies of X ,

$$F_{\frac{X_{n:n} - B_n}{A_n}}(x) = F_X^n(A_n x + B_n) \xrightarrow{n \rightarrow \infty} \Phi_\alpha(x).$$

In other words, the CDF F_X is in the domain of attraction of a Fréchet- α distribution, which can be denoted as $X \in \mathcal{D}(\mathcal{Y}_\alpha)$ or $F_X \in \mathcal{D}(\Phi_\alpha)$.

Recall that $F_X \in \mathcal{D}(\Phi_\alpha)$ if, and only if, $\bar{F}(x) = 1 - F(x) = x^{-\alpha} L(x)$, with $L(\cdot)$ a slowly varying function. Thus, random variables $X \in \mathcal{D}(\mathcal{Y}_\alpha)$ are those with an α -regularly varying upper tail, i.e., an upper tail with α -paretian behavior.

Therefore, most of Allodi's [4] results on goodness-of-fit are valid for distributions in the domain of attraction of some Fréchet- α , or power laws in a broader sense, where the upper values have a linear signature but that are not necessarily a Pareto power law in the strict sense.

On the other hand, only a few vulnerabilities have complete records on their lifecycle and ensuing malware. Therefore, it is reasonable to assume that those remaining are the result of a harsh geometric thinning process.

If $X_{N:N}$ is the maximum of a geometric- p random number N of independent copies of X , its CDF is

$$F_{X_{N:N}}(x) = \mathbb{P}[X_{N:N} \leq x] = \sum_{k=0}^{\infty} p(1-p)^{k-1} F_X^k(x) = \frac{p F_X(x)}{1 - (1-p) F_X(x)}.$$

Rachev and Resnick [6] investigated the stable limits of geometric thinned sequences of IID random variables, concluding that they were of the form $\text{geo-GEV}(x) = \frac{1}{1 - \ln(\text{GEV}(x))}$. So, the standard geo-stable models for geometric thinned maxima are the logistic, associated with the Gumbel model, the log-logistic- α , associated with the Fréchet- α ,

$$\text{Log-logistic}(\alpha), \alpha > 0 \Leftrightarrow {}^s\Phi_\alpha(x) = \frac{1}{1 - \ln \Phi_\alpha(x)} = \begin{cases} 0 & , x < 0 \\ \frac{1}{1+x^{-\alpha}} & , x \geq 0 \end{cases}, \quad (2)$$

and the symmetrized log-logistic- α , associated with the Weibull- α model for maxima.

The characterization of their domains of attraction coincides with the characterization of the domains of attraction of the classical stable laws of maxima (Rachev and Resnick [6], cf. also Cline [63]).

As the interest is in heavy-tailed models, the primarily focus will be on assessing the goodness-of-fit of the heavy-tailed models already mentioned. Aside from these, the goodness-of-fit for the Cauchy model with location λ and scale $\delta > 0$,

$$X \sim \text{Cauchy}(\lambda, \delta) \Leftrightarrow F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x - \lambda}{\delta}\right),$$

and for hyperexponential models with CDF

$$F(x) = \begin{cases} 0 & , x \leq 0 \\ \sum_{k=1}^m \pi_k (1 - e^{-\theta_k x}) & , x > 0 \end{cases},$$

will also be investigated. The heavy-tailed Cauchy distribution is the additive symmetric stable law with shape parameter $\alpha = 1$ (note that the Cauchy model with $\lambda = 0$ and $\delta = 1$ is the t -Student model with one degree of freedom), and the hyperexponential models are convex mixtures of exponential distributions with mixing proportions π_1, \dots, π_m ($\pi_k > 0$ and $\sum_{k=1}^m \pi_k = 1$) and rate parameters $\theta_1, \dots, \theta_m$ ($\theta_k > 0$), i.e., scale parameters $\delta_k = \frac{1}{\theta_k}$. Feldmann and Whitt [7] showed that the hyperexponential model can be an alternative fit when the right-tail provides an abundance of values that are far from the mode, median, and mean.

4.2. Some Trivia on Parameter Estimation and Model Selection

In what concerns the Pareto I model (1) (henceforth, simply referred to as the Pareto model), the maximum likelihood (m.l.) estimates of δ and α are $\hat{\delta} = x_{1:n}$ and $\hat{\alpha} = \frac{n}{\sum_{k=1}^n \ln x_k - n \ln \hat{\delta}}$ (as usual, $x_{k:n}$ denotes the k th ascending order statistic of a sample), with the corresponding m.l. estimators being independent random variables $\hat{\delta} \sim \text{Pareto}(\delta, n\alpha)$ and $\hat{\alpha} \sim \text{InverseGamma}(n-1, n\alpha)$ (Malik [64]). Observe that the expectation $\frac{\alpha\delta}{\alpha-1}$ exists for $\alpha > 1$, and the variance $\frac{\alpha\delta^2}{(\alpha-1)^2(\alpha-2)}$ exists for $\alpha > 2$, and, therefore, the Pareto distribution has a very heavy tail when $\alpha \in (0, 3)$ (especially so if $\alpha \in (0, 2)$).

There is some evidence that many phenomena exhibit this type of equilibrium, and in many cases, there are arguments in favor of fitting a power law with shape in $[2, 3)$, thus implying that the expectation does exist but with no finite variance (Clauset et al. [39]).

More generally, with respect to power laws, Bauke [65] addressed m.l. estimation in general. Fedotenkov [52] provided an updated overview of the estimation of tail weights, giving code for implementation of the estimators he investigated. Diaz [66] and Arnold and Brockett [67] provided alternative concepts and frameworks dealing with quantiles.

5. Materials and Methods

The samples used in the analyzes below were selected from the vulnerabilities registered in [8], using the options Vulnerability by Date and CVE Number Ascending, for the years 2021 and 2022.

In the exploratory data phase, a systematic sample of 2% of the $N = 20,171$ vulnerabilities registered in 2021 (sample size $n = 405$) was used. A preliminary analysis of the 2021 sample revealed a marked asymmetry, a large number of outliers, and a linear signature when $0 < \ln x < 5$, that is, when $1 < x < 148.4$. This led to consideration of a subsample of $n = 301$ values smaller than or equal to 148, with standard deviation $s = 23.41$, to fit a power law.

Using the estimated standard deviation of the 2021 sample after cleaning, the sample size n needed to estimate the mean of the time from publication to update in 2022 (population size $N = 25,082$), with an error bound $B = 1$ and confidence level $1 - \alpha = 0.95$, must satisfy $n \geq N \left(1 + \frac{(N-1)B^2}{z_{1-\alpha/2}^2 s^2}\right)^{-1} = 1942.35$, where $z_{1-\alpha/2}$ denotes the $(1 - \alpha/2)$ th quantile for the standard normal distribution. Therefore, a simple random sample without replacement of size $n = 1943$ was selected from the vulnerabilities registered in 2022.

The goodness-of-fit tests used were the Kolmogorov–Smirnov (KS), the Anderson–Darling (AD), and the Cramér–von Mises (CvM) tests. The AD and CvM tests are refinements of the KS test and are generally considered more powerful than the KS test. The results for the three tests are shown since each has unique features that may complement the others. For example, the AD test places more weight on the observations of the tail of the distribution, while the CvM test focuses on the central observations of the distribution, similar to the KS test.

Since the parameters of the models under evaluation are unknown, these are estimated using the maximum likelihood estimation method, with the expectation maximization (EM) method used for the hyperexponential case. In the other exceptional cases, the recommendations in Clauset et al., [39] or in Fedotenkov [52], Hall [50], and Davis and Resnick, [51] were applied.

However, a well-known shortcoming of each goodness-of-fit test considered is that they should not be used in their original form when the theoretical null distribution is not fully specified (which is the case here). Nevertheless, adaptations for the AD and CvM tests do exist when the parameters are estimated, which allows their use in these situations. The same, however, cannot be said of the KS test, unless modifications of the type for the Lilliefors test for normality exist, which do exist but which are very few in number. To get around this obstacle, Monte Carlo simulated p -values for the KS test for each model fit, based on $R = 999$ replicas, and following the guidelines in Davison and Hinkley [68], were obtained.

The goodness-of-fit criterion for model selection is the AIC, with lower values indicating better fits. An alternative goodness-of-fit criterion to the AIC, also quite often used in model selection, is the BIC. It suffices to consider the AIC because the BIC is always greater than the AIC; however, since some researchers prefer to use the BIC instead, both are indicated. Observe, however, that direct comparisons between the power law fit and other fits using the AIC should not be performed, because the power law fit is based on censored data, not on all the data.

It is also worth noting that, with respect to the log-logistic model, the usual estimation algorithms assume a zero location. The standard procedure of subtracting the minimum value of the sample $x_{1:n}$, the m.l. estimator of the location parameter λ to the original data may cause the maximum likelihood estimation method to fail to estimate the shape and scale parameters (using the R functions *fitdistr* of library *fitdistrplus* or *mle2* of library *bbmle*). Therefore, a workaround was needed to deal with this issue, as explained below.

The CDF of a standard log-logistic random variable Z with shape parameter $\alpha > 0$ is given by (2); hence, its PDF is $f_Z(x) = \frac{\alpha x^{-\alpha-1}}{(1+x^{-\alpha})^2} \mathbb{I}_{[0,\infty)}(x)$. Therefore, the PDF of the minimum $Z_{1:n}$ of n independent standard log-logistic random variables is $f_{Z_{1:n}}(x) = \frac{\alpha n x^{-\alpha n-1}}{(1+x^{-\alpha})^{n+1}} \mathbb{I}_{[0,\infty)}(x)$, and $\mathbb{E}[Z_{1:n}] = \frac{\Gamma(\frac{1}{\alpha})\Gamma(n-\frac{1}{\alpha})}{\alpha\Gamma(n)}$, where $\Gamma(v) = \int_0^\infty x^{v-1}e^{-x}dx$, $v > 0$ is the gamma function.

Consequently, with respect to the log-logistic with shape parameter $\alpha > 0$, the location parameter $\lambda \in \mathbb{R}$ and the scale parameter $\delta > 0$, the estimator $X_{1:n} = \delta Z_{1:n} + \lambda$ of λ overestimates the location since $\mathbb{E}[X_{1:n}] = \lambda + \frac{\delta}{\alpha} \frac{\Gamma(\frac{1}{\alpha})\Gamma(n-\frac{1}{\alpha})}{\Gamma(n)}$. Thus, an unbiased estimator for λ is the moment estimator $\tilde{\lambda} = X_{1:n} - \text{Bias}(\alpha, \delta, n)$, with $\text{Bias}(\alpha, \delta, n) = \frac{\delta}{\alpha} \frac{\Gamma(\frac{1}{\alpha})\Gamma(n-\frac{1}{\alpha})}{\Gamma(n)}$. Observe, however, that $X_{1:n}$ is asymptotically unbiased because $\lim_{n \rightarrow \infty} \frac{\Gamma(n-\frac{1}{\alpha})}{\Gamma(n)} = 0$, and, therefore, for large samples, $\mathbb{E}[X_{1:n}] \approx \lambda$. Based on these results, the following algorithm was designed and implemented to estimate the parameters of the log-logistic model $(\alpha, \lambda, \delta)$:

1. Start with an initial estimate for the bias value, e.g., $\text{bias}_0 = 0.0001$, and consider as an initial estimate of λ , $\tilde{\lambda}_0 = x_{1:n} - \text{bias}_0$.
2. Use the “almost zero-located” sample $y_k = x_k - \tilde{\lambda}_0$, $k = 1, \dots, n$, to obtain the m.l. estimates $\hat{\alpha}_1$ and $\hat{\delta}_1$ for a log-logistic with zero location.
3. Use values $\hat{\alpha}_1$ and $\hat{\delta}_1$ to obtain a new bias estimate $\text{bias}_1 = \text{Bias}(\hat{\alpha}_1, \hat{\delta}_1, n)$, and update the location estimate to $\tilde{\lambda}_1 = x_{1:n} - \text{bias}_1$.

4. Repeat steps 2 and 3 using the sample $y_k = x_k - \tilde{\lambda}_{i-1}$, $k = 1, \dots, n$, to obtain new estimates $\hat{\alpha}_i$, $\hat{\delta}_i$ and $\tilde{\lambda}_i$, $i = 2, 3, \dots$.
5. Stop when $\max\{|\hat{\alpha}_i - \hat{\alpha}_{i-1}|, |\hat{\delta}_i - \hat{\delta}_{i-1}|, |\tilde{\lambda}_i - \tilde{\lambda}_{i-1}|\} < 0.00001$ or $i = 250$.

On the other hand, the Vuong test for strictly non-nested model selection is also used to supplement arguments in favor of model choice, since the non-parametric tests used do not conclusively support the use of any of the considered models. A brief description of Vuong's test adapted to this context follows:

Let $F_\theta = \{f(\cdot; \theta) : \theta \in \Theta\}$ and $G_\gamma = \{g(\cdot; \gamma) : \gamma \in \Gamma\}$ be two competing non-nested models, that is, $F_\theta \cap G_\gamma = \emptyset$. For an observed sample (x_1, \dots, x_n) of a random sample (X_1, \dots, X_n) , the Vuong test statistic LR_n , based on the likelihood ratio, is defined as

$$LR_n = LR_n(\hat{\theta}, \hat{\gamma}) = L_n^f(\hat{\theta}) - L_n^g(\hat{\gamma}) = \sum_{i=1}^n \ln \frac{f(x_i; \hat{\theta})}{g(x_i; \hat{\gamma})}. \quad (3)$$

Under the null hypothesis of the two models F_θ e G_γ being equivalent,

$$n^{-1/2} LR_n(\hat{\theta}, \hat{\gamma}) / \hat{\omega}_n \xrightarrow[n \rightarrow \infty]{d} N(0, 1),$$

with

$$\hat{\omega}_n = \frac{1}{n} \sum_{i=1}^n \left[\ln \frac{f(x_i; \hat{\theta})}{g(x_i; \hat{\gamma})} \right]^2 - \left[\frac{1}{n} \sum_{i=1}^n \ln \frac{f(x_i; \hat{\theta})}{g(x_i; \hat{\gamma})} \right]^2.$$

Therefore, large positive values of the statistic LR_n provide evidence in favor of F_θ , while large negative values are in favor of G_γ .

The software used to perform the statistical analysis is R, a language and environment for statistical computing (R Core Team, 2022).

Some specific R libraries were used in the analysis: *fitdistrplus* for general fitting of models, *bbmle* for maximum likelihood estimation, *goftest* for goodness-of-fit tests and criteria, *evd* for extreme value models fitting, *sads* for the Pareto fitting, *flexsurv* for the log-logistic fitting, *evmix* for the generalized Pareto fitting, and *mixtools* for the hyperexponential fitting.

6. Results and Discussion

As discussed in Section 3, the evolution of vulnerabilities is relevant for assessing their severity and the risk exposure of the network. Therefore, for each sample of vulnerabilities, the goodness-of-fit of the models discussed in Section 4 to fit the number of days from publication to update of the vulnerabilities was investigated.

6.1. Vulnerabilities Published in 2021, Systematic Sampling

In the exploratory data analysis phase, a sample of 405 vulnerabilities was collected. Two extreme outliers that could bias the statistical analysis ended up being removed from the sample.

In Table 1 some relevant statistics for the 2021 sample ($n = 403$) are indicated (Q_1 , Q_2 , Q_3 , and SD denote the 1st quartile, median (2nd quartile), 3rd quartile, and the standard deviation, respectively). Figure 4 displays the data, selecting $h = 7$ for the bin width of the frequency histogram from periodicity considerations.

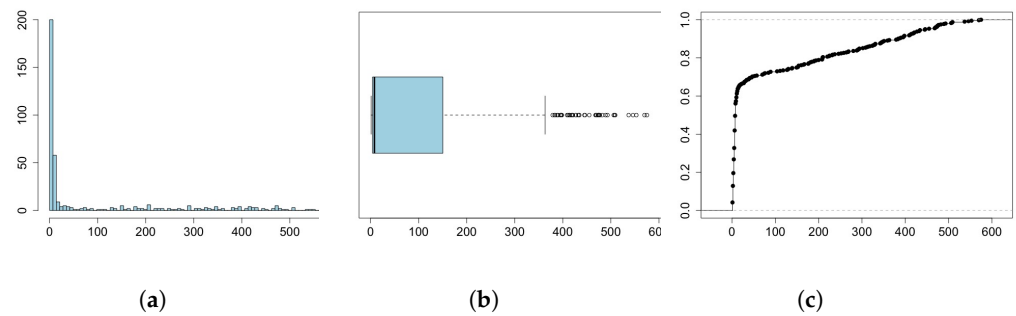


Figure 4. Representations of the 2021 sample: (a) Frequency histogram. (b) Boxplot. (c) Empirical distribution function.

Table 1. Summary statistics for the 2021 sample.

Min	Q_1	Q_2	Mean	Q_3	Max	SD	Mode	Skewness	Kurtosis
1	4	8	94.67	150.5	575	153.92	6	1.55	0.99

6.1.1. Fitting with Log-Normal and Models with Paretian Right-Tail

Following the observations made in Section 4, the fit with the Pareto, log-normal, log-logistic, GEV, generalized Pareto, Cauchy, power law, and hyperexponential distributions was analyzed. With regard to the power law fit, some changes were needed to perform the data analysis, and which justify the censoring of the sample to the right of 148. The procedure is explained below.

The existence of some intervals with null frequency in the histogram of Figure 4a required consideration of intervals with $h = 14$, thus reducing the number of intervals with null frequency to just two, which were later merged with an adjacent interval. In Figure 5, $\ln(f(x))$ is plotted against $\ln x$, with x being the midpoint of an interval and $f(x)$ the ratio between the relative frequency of the x 's interval and the intervals' width, therefore, obtaining a relative frequency histogram with area 1, that is, a density histogram.

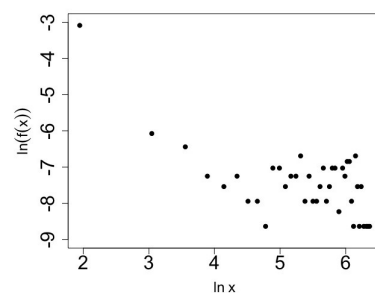


Figure 5. Power law linear signature in a log-log plot (2021 sample).

Observe in Figure 5 that a “linear signature” is noticeable when $0 < \ln x < 5$, that is, when $1 < x < 148.4$, with some points being randomly scattered to the right of $\ln 5$. Hence, a subsample of $n = 301$ values smaller than or equal to 148 was used in the analysis. Table 2 records the recalculated statistics for the censored sample and the observed changes are displayed in Figure 6.

Table 2. Summary statistics for the 2021 censored data.

Min	Q_1	$Q_2 = \text{Mode}$	Mean	Q_3	Max	SD	Skewness	Kurtosis
1	3	6	12.79	8	139	23.41	3.66	13.68

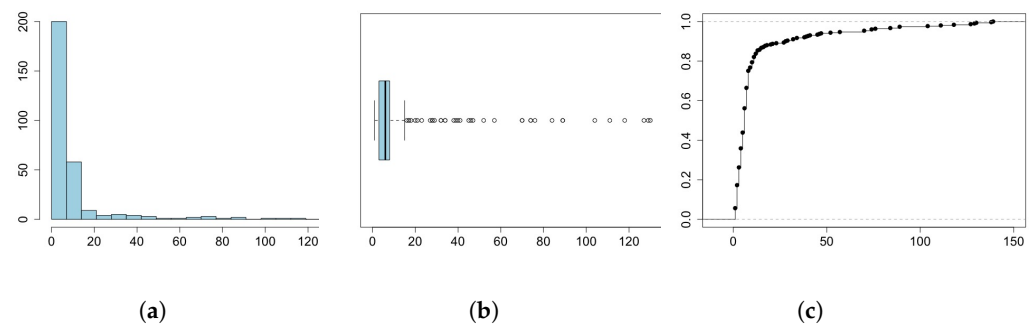


Figure 6. Representations of the 2021 censored sample: (a) Frequency histogram ($h = 7$). (b) Boxplot. (c) Empirical distribution function.

Table 3 exhibits the estimated parameters for each model fit, with standard errors inside parentheses, the p -value for each test, with the test statistic value being indicated underneath the p -value, and the AIC and BIC values.

Table 3. Parameters and goodness-of-fit results for some heavy-tailed models.

	KS	AD	CvM	AIC	BIC
Pareto					
$\hat{\alpha} = 0.353$ (0.018)	0.001	0.001	0.108	3926.3	3930.3
$\hat{\delta} = 1$	($D_n = 0.193$)	($A_n^2 = \infty$)	($\omega_n^2 = 0.826$)		
Log-normal					
$\hat{\lambda} = 2.829$ (0.096) ¹	0.001	0.310	0.024	3958.9	3966.9
$\hat{\delta} = 1.933$ (0.068) ¹	($D_n = 0.212$)	($A_n^2 = 3.360$)	($\omega_n^2 = 1.092$)		
Log-logistic					
$\hat{\alpha} = 0.702$ (0.029)	0.001	0.277	0.686	3849.9	3857.9
$\hat{\lambda} = 1$	($D_n = 0.147$)	($A_n^2 = 3.480$)	($\omega_n^2 = 0.439$)		
$\hat{\delta} = 11.532$ (1.416)					
General Extreme Value					
$\hat{\xi} = 1.700$ (0.089)	0.001	0.155	0.300	3878.9	3890.9
$\hat{\lambda} = 6.198$ (0.526)	($D_n = 0.133^*$)	($A_n^2 = 4.057$)	($\omega_n^2 = 0.632$)		
$\hat{\delta} = 9.540$ (1.037)					
Generalized Pareto					
$\hat{\xi} = 1.756$ (0.132)	0.001	0.001	0.152	<u>3798.0</u>	<u>3810.0</u>
$\hat{\lambda} = 1$	($D_n = 0.165$)	($A_n^2 = \infty$)	($\omega_n^2 = 0.764$)		
$\hat{\delta} = 8.636$ (0.951)					
Cauchy					
$\hat{\lambda} = 5.493$ (0.288)	0.001	0.001	0.002	4390.2	4398.2
$\hat{\delta} = 4.575$ (0.381)	($D_n = 0.271$)	($A_n^2 = 16.652$)	($\omega_n^2 = 1.549$)		
Power Law					
$\hat{\alpha} = 1.544$ (0.031)	0.001	0.001	0.034	2077.5	2081.2
$x_{\min} = 1$	($D_n = 0.277$)	($A_n^2 = \infty$)	($\omega_n^2 = 0.999$)		

¹ Estimates for the parameters of the logarithm of the random variable.

Observe that, with the exception of the Cauchy model (and the power law, but this case will be considered separately), all models are supported by at least one goodness-of-fit test. The model with the lowest AIC (and BIC) is the generalized Pareto, but it is only supported by the CvM test. The log-logistic and GEV are the models that are supported by both the AD and CvM tests. In fact, the log-logistic has the second lowest AIC, followed by

GEV. Although the KS test does not support any model, it is curious to see that the GEV fit has the smallest value of the test statistic, which measures the maximum distance between the empirical and theoretical distribution functions. It should also be noted that the power law fit, strictly following the procedures recommended by Clauset et al. [39] and Holm [3] for this purpose, is totally inadequate.

To strengthen the arguments in favor of selecting the log-logistic model against the GEV model, since both models are supported by the AD and CvM tests, the Vuong test (3) is used, as described in Section 5 for the selection of strictly “non-nested” models.

Comparing the fitted log-logistic (F_θ) and GEV (G_γ) models with Vuong’s test, $LR_n = 19.091$ ($\hat{\omega}_n = 0.375$), with p -value = 0.006. Therefore, there is some evidence in favor of the log-logistic having a better fit than GEV.

On the other hand, the fact that the extreme index ξ of the fitted GEV is a positive estimate ($\hat{\xi} = 1.700$), and consequently a Fréchet with $\alpha = 1/\xi$ ($\hat{\alpha} = 0.588$), means that a very heavy-tailed law is at play here. Moreover, the log-logistic which best fits the data, with a shape parameter estimate $\hat{\alpha} = 0.702$, is in the domain of attraction of a GEV with $\tilde{\xi} = 1/\hat{\alpha} = 1.425$.

6.1.2. Fitting a Hyperexponential Model

The considerable amount of outliers in the sample may indicate that there is a majority group of vulnerabilities in which the period between disclosure and update is short, in general, less than 30 days. Observe that the censored sample, which is very asymmetrical (skewness = 3.66), used to estimate a power law fit, has median = mode = 6, half the value of the mean (12.79), which, in turn, is half the value of the standard deviation (23.41)), and a less expressive set of vulnerabilities in which the referred period is long. This is clearly visible in the empirical distribution function (see Figure 4c), which indicates that there is indeed a mixture of distributions, and is certainly the fundamental reason why the traditional models used above are inadequate.

The results for the hyperexponential fit for $m = 2, 3$ are shown in Table 4.

Table 4. Parameters and goodness-of-fit results for the hyperexponential model.

	KS	AD	CvM	AIC	BIC
$m = 2$					
$\hat{\pi}_1 = 0.649; \hat{\theta}_1 = 0.163$	0.001	0.772	0.197	3759.6	3771.6
$\hat{\pi}_2 = 0.351; \hat{\theta}_2 = 0.004$	($D_n = 0.141$)	($A_n^2 = 2.209$)	($\omega_n^2 = 0.715$)		
$m = 3$					
$\hat{\pi}_1 = 0.649; \hat{\theta}_1 = 0.163$	0.001	0.569	0.121	3763.6	3783.6
$\hat{\pi}_2 = 0.017; \hat{\theta}_2 = 0.004$	($D_n = 0.141$)	($A_n^2 = 2.665$)	($\omega_n^2 = 0.806$)		
$\hat{\pi}_3 = 0.334; \hat{\theta}_3 = 0.004$					

Observe that both the hyperexponential fits are supported by the AD and CvM tests, and for the case $m = 3$, the approximations to three decimal places of the estimates $\hat{\theta}_2$ and $\hat{\theta}_3$ are indistinguishable. Moreover, the AIC is greater than the corresponding value for $m = 2$, which is to be expected since more parameters are estimated for $m = 3$. Observe also that when comparing the hyperexponential and the log-logistic AIC values, the smallest value is seen with the former. Therefore, the hyperexponential model should be tested against the log-logistic with Vuong’s test to strengthen the rationale behind the choice.

The comparison between the hyperexponential ($m = 2$) and the log-logistic fits for the Vuong’s test statistic yields $LR_n = 40.553$ ($\hat{\omega}_n = 0.453$), with p -value = 4.1×10^{-6} . Hence, there is strong evidence in favor of the hyperexponential model having a better fit.

Figure 7 displays the empirical distribution function (EDF), comparing it with the investigated fits that are supported by at least one goodness-of-fit test. Although the power law is not supported by any test, its fit is compared with the EDF of the censored data (separately) to highlight the lack of fit.

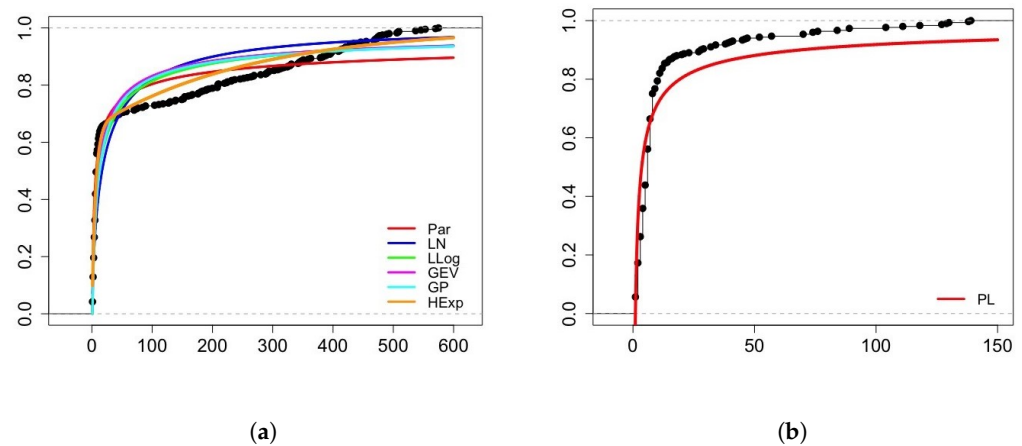


Figure 7. Empirical distribution function (black line) and fitted models: (a) Pareto (Par), log-normal (LN), log-logistic (LLog), general extreme value (GEV), generalized Pareto (GP), and hyperexponential $m = 2$ (HExp). (b) Power law (PL).

6.2. Vulnerabilities Published in 2022, Simple Random Sampling

For each of the 1943 vulnerabilities sampled, the corresponding CVSS base score was also recorded in order to assess the severity of the vulnerabilities registered in 2022. Table 5 gives some statistics for the CVSS base score, with a representation of the data being shown in Figure 8.

Table 5. Some statistics for the CVSS base score.

Min	Q_1	Q_2	Mean	Q_3	Max	SD
2.3	6.1	7.5	7.3	8.8	10	1.7

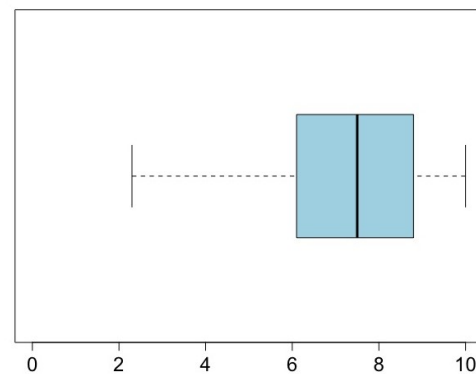


Figure 8. Representation of the CVSS base score sample.

With respect to the distribution of the CVSS base score according to severity risk, observe that 1.2% ([0%,3.6%]) are classified as being low risk, 39.4% ([37.0%,41.9%]) as medium risk, 38.1% ([35.7%,40.5%]) as high risk, and 21.3% ([18.9%,23.8%]) are considered critical (the intervals indicated are the simultaneous 95% confidence intervals for multinomial proportions, cf. Glaz and Sison [69]).

However, the main interest lies in modeling the number of days between publication and update of vulnerabilities. Since 26 vulnerabilities sampled from 2022 have coinciding publication and update dates, these were removed from the sample.

Two modeling approaches seemed worthwhile to explore. One consisted of using all data for the fitting, and the other, using the subsample without the extreme outliers (364 in total), with the aim of making the statistical analysis more robust. The results obtained with each approach will be compared later on with each other, and also with the findings for the 2021 sample, to establish which modeling strategy is the most adequate in this framework.

6.2.1. Modeling the Data

Statistics for the sample (sample size $n = 1917$) are given in Table 6, with some representations of the data being displayed in Figure 9. Note that the bin width of the histogram in Figure 9a is $h = 14$ days (this ensures that all intervals have non-null frequencies). It is also worth mentioning that 72.6% of the vulnerabilities sampled were updated within the first two weeks after publication (50.5% within the first week and 11.6% in exactly 7 days).

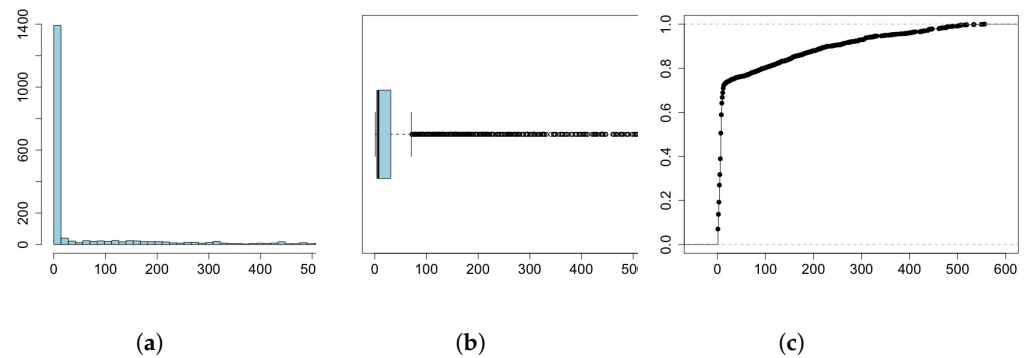


Figure 9. Representations of the 2022 sample: (a) Frequency histogram. (b) Boxplot. (c) Empirical distribution function.

Table 6. Statistics for the 2022 sample.

Min	Q_1	$Q_2 = \text{Mode}$	Mean	Q_3	Max	SD	Skewness	Kurtosis
1	4	7	61.12	31	556	116.88	2.29	4.44

In order to establish the power law signature, the procedure used for the 2021 sample was applied here. In this case, x is the midpoint of an interval of width $h = 14$, and $f(x)$ is the ratio between the relative frequency of the x 's interval and 14. By plotting $\ln(f(x))$ against $\ln x$ (see Figure 10), a linear signature is observed when $0 < \ln x < 4$; that is, if $1 < x < 54.6$. Therefore, to fit a power law model to the data, a subsample of $n = 1465$ values that are smaller than or equal to 54 should be considered.

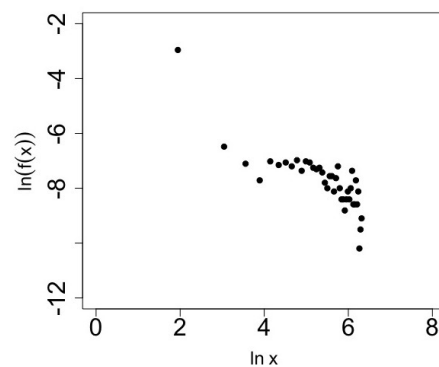


Figure 10. Power law linear signature in a log-log plot (2022 sample).

The recalculated statistics for the censored data are shown in Table 7, and Figure 11 exhibits a plot of the data.

Table 8 shows the results obtained for the heavy-tailed model fits and Table 9 for the hyperexponential fit for $m = 2, 3$.

Table 7. Statistics for the 2022 censored sample.

Min	Q_1	Q_2	Mean	Q_3	Max	SD	Mode	Skewness	Kurtosis
1	3	6	7.03	8	54	6.4	7	3.58	17.50

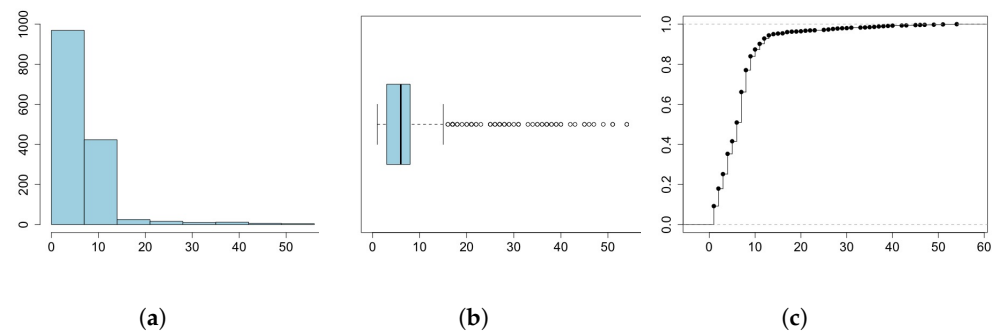


Figure 11. Representations of the 2022 censored sample: (a) Frequency histogram ($h = 7$). (b) Boxplot. (c) Empirical distribution function.

Table 8 shows that the GEV fit is supported by both the AD and CvM tests, and the log-logistic fit just by the CvM test. None of the other model fits are supported by any goodness-of-fit test. For all fits that pass at least one test (GEV and log-logistic), the lowest AIC is obtained with the log-logistic. Although the generalized Pareto has the smallest AIC (note that the power law fit is not under consideration), it should not be considered because it is not supported by any test. Moreover, when comparing the values of the KS test statistic, it is interesting to see again that the smallest value is achieved with GEV, despite the fact that the KS test does not support this model or any other for that matter. As for the power law itself, the results clearly show an inadequate fit.

On the other hand, the results for the hyperexponential model (see Table 9) reveal that there is some evidence in favor of it being a good fit for both $m = 2$ and $m = 3$ (only with the CvM test). There is not much gain when considering the case $m = 3$, because its last two exponential components have rates that are indistinguishable to three decimal places. Additionally, when comparing the AIC for the hyperexponential for $m = 2$ with the AIC of GEV and of the log-logistic, a better fit is achieved with the log-logistic. Note, however, that for all models considered, the hyperexponential is the model that requires more parameters to be estimated, three for $m = 2$ and five for $m = 3$, which may explain, to a certain extent, the bigger AIC value compared with the log-logistic one. It is a well-known fact that the AIC (and the BIC) penalize models with more parameters. Nonetheless, the hyperexponential fit with $m = 2$ has a smaller AIC than the GEV fit.

Figure 12 shows the EDF, comparing it with the investigated fits that are supported by at least one goodness-of-fit test. The power law case is shown again separately, with the lack of fit to the censored data being quite striking.

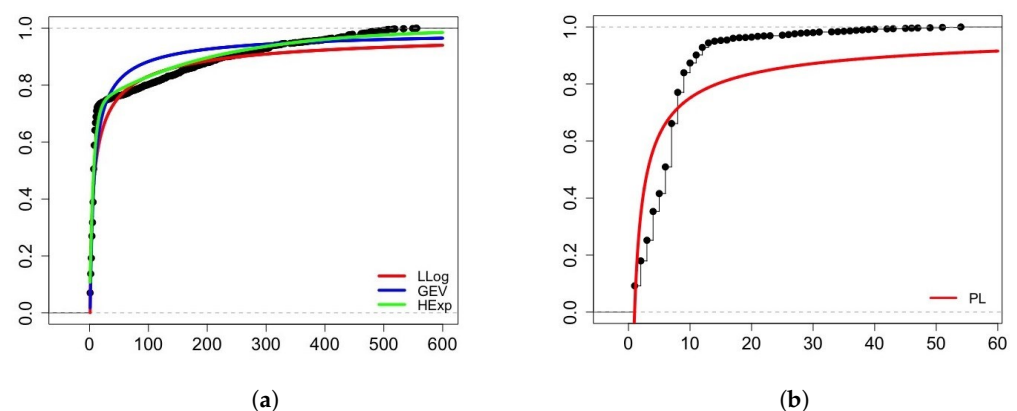


Figure 12. Empirical distribution function (black line) and fitted models: (a) log-logistic (LLog), general extreme value (GEV), and hyperexponential $m = 2$ (HEXp). (b) Power law (PL).

Table 8. Parameters and goodness-of-fit results for some heavy-tailed models.

	KS	AD	CvM	AIC	BIC
Pareto					
$\hat{\alpha} = 0.398$ (0.009) $\hat{\delta} = 1$	0.001 ($D_n = 0.231$)	0.001 ($A_n^2 = \infty$)	0.002 ($\omega_n^2 = 1.686$)	17,009.8	17,015.4
Log-normal					
$\hat{\lambda} = 2.514$ (0.039) ¹ $\hat{\delta} = 1.718$ (0.028) ¹	0.001 ($D_n = 0.217$)	0.026 ($A_n^2 = 6.466$)	0.006 ($\omega_n^2 = 1.498$)	17,157.7	17,168.8
Log-logistic					
$\hat{\alpha} = 0.639$ (0.013) $\hat{\lambda} = 1$ $\hat{\delta} = 8.296$ (0.487)	0.001 ($D_n = 0.164$)	0.001 ($A_n^2 = 14.216$)	0.103 ($\omega_n^2 = 0.986$)	16,182.3	16,193.4
General Extreme Value					
$\hat{\xi} = 1.433$ (0.039) $\hat{\lambda} = 5.348$ (0.187) $\hat{\delta} = 7.207$ (0.317)	0.001 ($D_n = 0.137^*$)	0.156 ($A_n^2 = 4.746$)	0.193 ($\omega_n^2 = 0.865$)	16,792.9	16,809.6
Generalized Pareto					
$\hat{\xi} = 1.358$ (0.051) $\hat{\lambda} = 1$ $\hat{\delta} = 8.686$ (0.399)	0.001 ($D_n = 0.192$)	0.001 ($A_n^2 = \infty$)	0.018 ($\omega_n^2 = 1.311$)	<u>16,114.2</u>	<u>16,130.9</u>
Cauchy					
$\hat{\lambda} = 6.048$ (0.116) $\hat{\delta} = 3.692$ (0.123)	0.001 ($D_n = 0.212$)	0.001 ($A_n^2 = 27.436$)	0.000 ($\omega_n^2 = 2.553$)	18,311.4	18,322.5
Power Law					
$\hat{\alpha} = 1.604$ (0.016) $x_{\min} = 1$	0.001 ($D_n = 0.315$)	0.001 ($A_n^2 = \infty$)	0.000 ($\omega_n^2 = 1.936$)	9262.6	9267.9

¹ Estimates for the parameters of the logarithm of the random variable.**Table 9.** Parameters and goodness-of-fit results for the hyperexponential model.

	KS	AD	CvM	AIC	BIC
$m = 2$					
$\tilde{\pi}_1 = 0.723; \tilde{\theta}_1 = 0.159$ $\tilde{\pi}_2 = 0.277; \tilde{\theta}_2 = 0.005$	0.001 ($D_n = 0.153$)	0.046 ($A_n^2 = 5.921$)	0.290 ($\omega_n^2 = 0.782$)	<u>16,397.5</u>	<u>16,414.2</u>
$m = 3$					
$\tilde{\pi}_1 = 0.723; \tilde{\theta}_1 = 0.159$ $\tilde{\pi}_2 = 0.046; \tilde{\theta}_2 = 0.005$ $\tilde{\pi}_3 = 0.231; \tilde{\theta}_3 = 0.005$	0.001 ($D_n = 0.153$)	0.046 ($A_n^2 = 5.927$)	0.290 ($\omega_n^2 = 0.782$)	16,401.5	16,429.3

From the above results, the log-logistic and hyperexponential ($m = 2$) are the two strongest candidates to model the data. If the choice were to be made solely based on AIC, the log-logistic would be selected. However, Figure 12a seems to reveal that the hyperexponential provides a more adequate fit, since it more closely follows the EDF towards the center to the right of the empirical data distribution. To better justify the choice between these two models, once again, the Vuong test can be helpful.

Comparing the log-logistic and hyperexponential ($m = 2$) models, $LR_n = 98.776$ ($\hat{\omega}_n = 1.134$) and p -value = 0.023, and, therefore, there is some evidence in favor of the log-logistic having a better fit. Note, however, that if the decision is to be made using a significance level of 1%, the conclusion would be that there is some evidence (although weak) in favor of both models being equivalent.

6.2.2. Modeling the Data Without Extreme Outliers

As mentioned earlier, the second modeling approach consists of considering a subsample of the 2022 sample without the extreme outliers (sample size $n = 1553$). Statistics for this subsample are given in Table 10, with some representations of the data being shown in Figure 13 (the bin width of the histogram is $h = 7$ since all intervals have non-null frequencies).

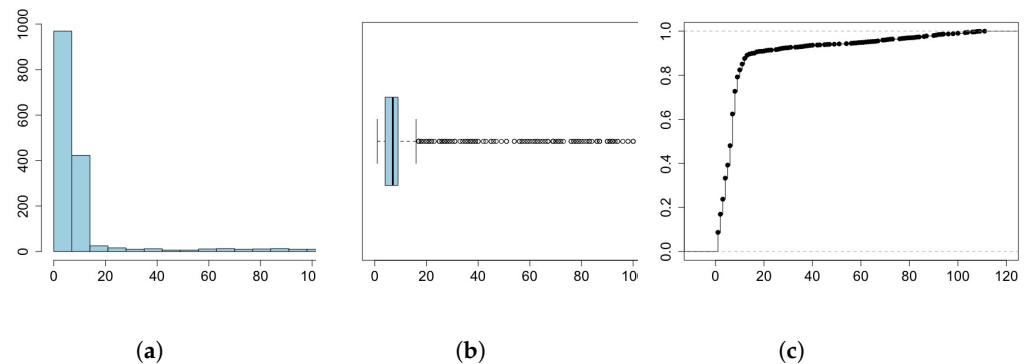


Figure 13. Representations of the 2022 subsample: (a) Frequency histogram. (b) Boxplot. (c) Empirical distribution function.

Table 10. Statistics for the 2022 subsample.

Min	Q_1	$Q_2 = \text{Mode}$	Mean	Q_3	Max	SD	Skewness	Kurtosis
1	4	7	11.3	9	111	18.89	3.53	12.04

The procedure described before to establish the power law signature was applied again to this subsample (see Figure 14). In this case, a linear signature is observed when $0 < \ln x < 4$, or if $1 < x < 54.6$, with exactly the same condition obtained when working with the whole 2022 sample. Therefore, the results for the power law fit are identical in this context to those shown in Table 8, and for this reason will not be replicated here. The results for the other model fits are given in Tables 11 and 12.

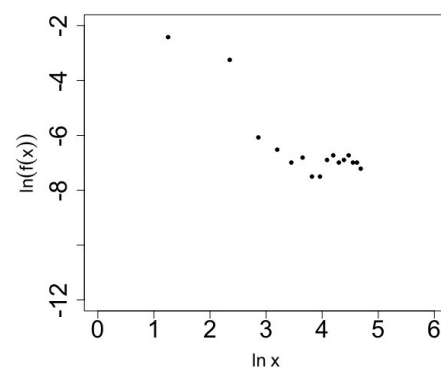


Figure 14. Powerlaw linear signature for the 2022 subsample in a log-log plot.

This modeling approach leads to some different conclusions from those when modeling with all the data. The log-normal, GEV, and Cauchy are potential model candidates, all being supported by both the AD and CvM tests. The hyperexponential is also a potential model candidate, but is only supported by the CvM test. Again there is no gain in using the hyperexponential $m = 3$ model compared with the $m = 2$ model. Curiously, the evidence in favor of the log-logistic model is no longer supported by any goodness-of-fit test, although the p -value of the CvM test is close to the significance level of 5%.

Table 11. Parameters and goodness-of-fit results for some heavy-tailed models.

	KS	AD	CvM	AIC	BIC
Pareto					
$\hat{\alpha} = 0.552$ (0.014) $\hat{\delta} = 1$	0.001 ($D_n = 0.297$)	0.001 ($A_n^2 = \infty$)	0.001 ($\omega_n^2 = 1.811$)	10,578.7	10,584.0
Log-normal					
$\hat{\lambda} = 1.811$ (0.025) ¹ $\hat{\delta} = 0.997$ (0.018) ¹	0.001 ($D_n = 0.141$)	0.469 ($A_n^2 = 3.468$)	0.117 ($\omega_n^2 = 0.939$)	10,026.4	10,037.1
Log-logistic					
$\hat{\alpha} = 0.982$ (0.023) $\hat{\lambda} = 1$ $\hat{\delta} = 4.510$ (0.188)	0.001 ($D_n = 0.170$)	0.001 ($A_n^2 = 9.905$)	0.045 ($\omega_n^2 = 1.116$)	10,016.8	10,027.5
General Extreme Value					
$\hat{\xi} = 0.568$ (0.027) $\hat{\lambda} = 4.337$ (0.107) $\hat{\delta} = 3.639$ (0.106)	0.001 ($D_n = 0.121^*$)	0.377 ($A_n^2 = 3.721$)	0.426 ($\omega_n^2 = 0.676$)	<u>9946.2</u>	<u>9962.2</u>
Generalized Pareto					
$\hat{\xi} = 0.368$ (0.029) $\hat{\lambda} = 1$ $\hat{\delta} = 6.856$ (0.258)	0.001 ($D_n = 0.171$)	0.001 ($A_n^2 = \infty$)	0.029 ($\omega_n^2 = 1.200$)	9353.5	9369.6
Cauchy					
$\hat{\lambda} = 6.102$ (0.100) $\hat{\delta} = 2.484$ (0.082)	0.001 ($D_n = 0.130$)	0.091 ($A_n^2 = 5.159$)	0.509 ($\omega_n^2 = 0.633$)	10,297.1	10,307.8

¹ Estimates for the parameters of the logarithm of the random variable.

Table 12. Parameters and goodness-of-fit results for the hyperexponential model.

	KS	AD	CvM	AIC	BIC
$m = 2$					
$\tilde{\pi}_1 = 0.890; \tilde{\theta}_1 = 0.158$ $\tilde{\pi}_2 = 0.110; \tilde{\theta}_2 = 0.020$	0.001 ($D_n = 0.187$)	0.030 ($A_n^2 = 6.243$)	0.107 ($\omega_n^2 = 0.956$)	<u>10,141.9</u>	<u>10,158.0</u>
$m = 3$					
$\tilde{\pi}_1 = 0.760; \tilde{\theta}_1 = 0.158$ $\tilde{\pi}_2 = 0.130; \tilde{\theta}_2 = 0.158$ $\tilde{\pi}_3 = 0.110; \tilde{\theta}_3 = 0.020$	0.001 ($D_n = 0.187$)	0.030 ($A_n^2 = 6.243$)	0.107 ($\omega_n^2 = 0.956$)	10,145.9	10,172.7

On the other hand, the GEV fit is the one that has the smallest AIC of all the fits, as well as the lowest value for the KS test statistic. Therefore, if the choice of a model is to be made solely on the AIC, the GEV model is clearly the winner, and the log-normal the runner up (it has the second smallest AIC). The different conclusions from the ones obtained before might be somewhat explained by the fact that the subsample used has a skewness of 3.53, as opposed to 2.29 for the whole sample, which introduces some changes in the characteristics of the data and, therefore, of the underlying data distribution.

Figure 15 displays the EDF, comparing it with the log-normal, GEV, Cauchy, and hyperexponential ($m = 2$) fits. Observe that, despite the fact that the hyperexponential has the third lowest AIC of the four models, its fit again follows quite closely the EDF towards the center to the right of the empirical data distribution.

With the GEV and the log-normal models being the two strongest model candidates, it is advisable once again to use the Vuong test to support a choice. Comparing these two models, $LR_n = 41.084$ ($\hat{\omega}_n = 0.025$) and $p\text{-value} = 0$; hence, there is strong evidence in favor of GEV having a better fit here.

On the other hand, and out of curiosity, when comparing the GEV and hyperexponential ($m = 2$) fits with the Vuong test, as $LR_n = 97.879$ ($\hat{\omega}_n = 0.120$), and the $p\text{-value} \approx 0$, when the choice is between GEV and the hyperexponential, the test's results clearly suggest the first model. However, one must remember that the AIC (and the BIC) is an overall measure of goodness-of-fit, which may explain why a better fit is achieved with GEV.

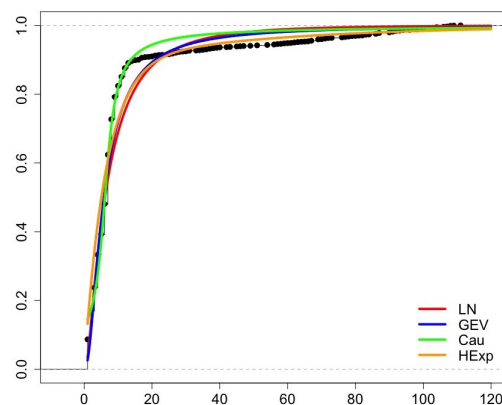


Figure 15. Empirical distribution function (black line) and fitted models: log-normal (LN), general extreme value (GEV), Cauchy (Cau), and hyperexponential $m = 2$ (HExp).

The analysis of the number of days from publication to update of vulnerabilities using a systematic sample ($n = 403$) from 2021 and a simple random sample ($n = 1943$) from 2022 led to consistent results, both with respect to the basic sample statistics (Tables 1, 2, 6, 7, and 10) and with respect to the choice of statistical models, with a clear indication that an extreme value (either GEV, in the classical setting of the maxima of IID random variables, or log-logistic in the context of geometrically thinned sequences of IID random variables, or a hyperexponential mixture of two exponential variables) provided the best fit.

Observe that to be on the safe side, the analysis was performed for the sets of all data and for the censored sets cleaned of extreme outliers, as indicated in the boxplot diagrams. But the adequacy of the hyperexponential fit seems to indicate that the data considered as outliers in the exploratory data analysis phase are data coming from the second component of the hyperexponential population.

There is some evidence, in view of the hyperexponential fit and of the sample statistics seen in Tables 1, 2, 6, 7, and 10, that it is reasonable to split the data into two subsets at the three weeks threshold. This provides a naive indication of the weights of the first and second components of the hyperexponential fit, as shown in Table 13.

Table 13. Weights for the hyperexponential fit (based on the 2022 sample).

Risk Level	Update-Publication ≤ 21	Update-Publication > 21
Low	20 (86.96%)	3 (13.04%) $\approx \frac{1}{8}$
Medium	594 (77.55%)	172 (22.45%) $\approx \frac{1}{5}$
High	499 (67.43%)	241 (32.57%) $\approx \frac{1}{3}$
Critical	330 (79.71%)	84 (20.29%) $\approx \frac{1}{5}$
All	1443 (74.27%)	500 (25.73%) $\approx \frac{1}{4}$

A sensible conjecture is that hackers maintain interest in one-third of high-risk vulnerabilities during a longer period, while for medium-risk and critical-risk vulnerabilities (in this last case possibly because vendors prioritize patching investment and discourage hackers) only one-fifth have a long update period. With respect to low-risk vulnerabilities, the one-eighth of those taking a long update period is even smaller. Observe that there was a subsample of only 23 low-risk vulnerabilities, which seems negligible in a sample of size $n = 1943$ (1.18%). The split sample statistics are indicated in Table 14.

Table 14. Split sample statistics.

	Min	Q ₁	Q ₂	Q ₃	Max	Mean	SD
Medium Risk							
Update-Publication ≤ 21	0	3	6	8	21	5.80	3.58
Update-Publication > 21	26	98.5	188	309	554	220.4	143.08
High Risk							
Update-Publication ≤ 21	0	3	6	8	21	5.96	3.60
Update-Publication > 21	22	108	190	303	556	214.50	134.83
Critical							
Update-Publication ≤ 21	0	4	6	8	19	6.27	3.39
Update-Publication > 21	27	92	180.5	331	505	217.3	144.19
All							
Update-Publication ≤ 21	0	3	6	8	21	5.97	3.55
Update-Publication > 21	22	103.8	186	309	556	217.1	139.41

The close coincidence of values is striking. Concerning the quick update component, the median time is about one week, for the 3rd quartile, 8 days, and (although artificial) this slowly increases to the maximum time to update of 3 weeks.

With regard to the long updating vulnerabilities, around 25% of the vulnerabilities are updated up to 3 months, the median time to update is approximately 6 months, and the 3rd quartile is approximately 10 months.

This may inspire some ways of forecasting a dynamic CVSS score. Instead of the temporal metrics equation $TS = RUp(BS \times ECM \times RL \times RC)$ (see Appendix A and Section 2), a multiplier $R(t)$ can be used to obtain a time-dependent dynamic base score as

$$BS(t) = RUp(BS \times R(t)),$$

where $R(t)$ is computed twice: for the fast updating component and for the slow updating component of the hyperexponential fit. This will necessarily produce two lines showing the probable evolution of the severity score of the vulnerability, assuming that the multiplier $R(t)$ appropriately reflects the behavior of hackers and vendors.

Adequate choices of the multiplier $R(t)$ depend on a fuller understanding of hackers' behavior, and empirical evaluation must take into account its effect on the number of changes over time from medium to high, from high to critical, and vice versa, and, concomitantly, the time intervals in which these changes occur. These are informal indicators of the effectiveness of the R multipliers in boosting the CVSS.

An example, assuming that the interest of hackers in the vulnerability increases steadily until the 1st quartile and also between the 1st quartile and the median (Q_2), and that, in most cases, they will loose interest in the vulnerability (eventually, since the vendors invest in patching), is

$$R(t) = \begin{cases} 0, & t < 0 \\ 1 + F^2(t), & 0 \leq t < Q_1 \\ 1 + 0.25^2 + \sqrt{[F(t) - 0.25]F(t)}, & Q_1 \leq t < Q_2 \\ 1 + 0.25^2 + \sqrt{0.25 \times 0.5 + [F(t) - 0.5][F(t) - 0.75]}, & Q_2 \leq t < Q_3 \\ 1 + 0.25^2 + \sqrt{0.25 \times 0.5 - (1 + \sqrt{3})[F(t) - 0.75]}, & Q_3 \leq t \end{cases}.$$

Figure 16 exhibits the above multiplier function, as well as the evolution of a vulnerability with score 7.4 over time.

Although the initial aim of this research was to change the static status of the CVSS, the development of the work led to some reflections on the past and expected evolution of the measurement concept.

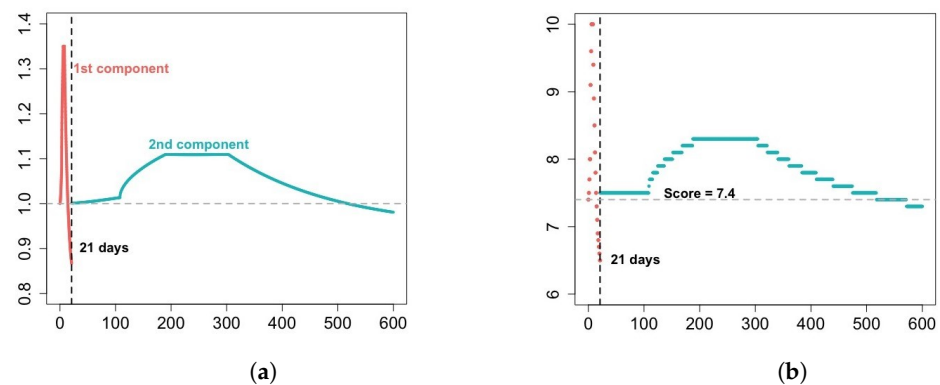


Figure 16. (a) Multiplier $R(t)$. (b) Evolution of score 7.4 over time assuming the multiplier $R(t)$.

Just as the law of large numbers and the central limit theorem became non-physical measurement aids, and regression made it possible to measure one variable to evaluate another (for instance, an instrument to measure the glucose level on a daily basis explains in the technical description of the device that it effectively measures the angle of refraction of light in the blood deposited on the strip, and transforms this measurement into the “measurement” of glucose), the digital transition will eventually lead metrology to evolve towards encompassing other non-physical auxiliary means of measurement, such as metrics and CVSS indices and machine learning algorithms, such as those that evaluate EPSS (the theme of the 2022 World Metrology Day was ‘Metrology in the Digital Era’).

In the case of assessing the severity of vulnerabilities or the likelihood of risk of exploitation, it will imply hard work in the harmonization process and eventually a desirable evolution towards standardization. However, the situation is far from satisfactory. For example, with a file containing the vulnerabilities that in April 2021 to March 2022 were rated ‘Functional’ or ‘High’ with respect to the metric exploit code maturity (which is a collection of data, not a random sample), searching the CVE-ID in other databases to obtain CTI and EPSS, the unpleasant surprise was to find that the empirical correlations were $\text{corr}(\text{CVSS}, \text{CTI}) = -0.12$, $\text{corr}(\text{CTI}, \text{EPSS}) = -0.01$ and $\text{corr}(\text{CVSS}, \text{EPSS}) = 0.14$, indicating that they are uncorrelated scores. Thus, as prioritization guides in the urgency of patching vulnerabilities, while they are often concordant, they do not offer the comfortable sense of security to which such an economically sensitive area aspires.

7. Conclusions

Since the pathbreaking paper of Frei et al. [2], it has been common knowledge that accurate modeling inter-event times in the lifecycle of vulnerabilities is essential to assess the risk of exposure of a network. As Stumpf and Porter [53] highlighted, often the data available do not support the biased option of modeling using power laws. However, on the other hand, there exists sound evidence that heavy-tailed models should be used.

Therefore, the goodness-of-fit of competitive models, including Pareto, log-normal, log-logistic, general extreme value, generalized Pareto, and Cauchy, were investigated using the usual goodness-of-fit Kolmogorov–Smirnov, Anderson–Darling, and Cramér–von Mises tests, and the AIC and BIC. A refinement of statistical choice was obtained using Vuong’s test [11], with the overall conclusion that the GEV and log-logistic models provided the best adjustments. This is not surprising since the log-logistic is also an extreme value law in the context of geometrical thinning.

On the other hand, following the claim of Feldmann and Whitt [7] and Nurmi et al. [36] that hyperexponential models provide a good fit to long-tailed data, this was also investigated. The goodness-of-fit was striking, reflecting the fact that there is good reason to believe that attackers lose interest in most of the vulnerabilities quite soon, but some hackers maintain interest in some vulnerabilities for a long period. Thus the number of outliers in the samples of time from publication to the update of vulnerabilities indicates that very likely a mixture of at least two populations is present.

The number of yearly reported vulnerabilities has sharply increased since 2017, and this is the source of unavoidable limitations due to sampling. Another limitation comes from hackers' diverse preferences for medium-, high-, and critical-risk vulnerabilities. However, the close coincidence of the sample statistics in Table 14 indicates some sort of stability.

As the thinned extreme value log-logistic law and the hyperexponential model provide such good fits to the data, the logical next step is to define a hyperlog-logistic model, combining the features of the two, and to implement the EM algorithm to compare the results.

Author Contributions: Conceptualization, M.d.F.B., D.P., P.P. and M.L.R.; methodology, M.d.F.B. and D.P.; software, M.d.F.B., P.P. and M.L.R.; validation, M.d.F.B., P.P. and M.L.R.; formal analysis, M.d.F.B.; investigation, M.d.F.B., D.P., P.P. and M.L.R.; resources, M.d.F.B. and M.L.R.; data curation, M.d.F.B., D.P., P.P. and M.L.R.; writing—original draft preparation, M.d.F.B. and D.P.; writing—review and editing, M.d.F.B., D.P., P.P. and M.L.R.; visualization, M.d.F.B. and P.P.; supervision, D.P.; project administration, M.d.F.B.; funding acquisition, M.d.F.B. and D.P. All authors have read and agreed to the published version of the manuscript.

Funding: Research partially supported by National Funds through FCT—Fundação para a Ciência e Tecnologia, project UIDB/00006/2020 (CEAUL).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the statistical analysis can be requested from the corresponding author.

Acknowledgments: The authors are grateful to S. Frei, M.A. Porter and L. Allodi for permission to quote from their pathbreaking papers Frei et al. [2], Stumpf and Porter [53], and Allodi [4], respectively. The authors are also grateful to FIRST for permission to reproduce Figures 1, 2, A2, and A3, to Frei for sending us a higher resolution version of Figure 3, and to Dave Dugal for authorizing the reproduction of Figure A1, and sending a higher resolution image.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Anderson–Darling
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
CVE	Common Vulnerabilities and Exposures
CvM	Cramér–von Mises
CVSS	Common Vulnerabilities Scoring System
EDF	Empirical Distribution Function
EM	Expectation Maximization
EPSS	Exploit Prediction Scoring System
GEV	General Extreme Value
IID	Independent and Identically Distributed
KS	Kolmogorov–Smirnov
PDF	Probability Density Function
TBC	Time Between Compromises
TTC	Time To First Compromise

Appendix A. CVSS v3.1 Metrics

Base metrics

$$\text{Base values: } \left\{ \begin{array}{l} \text{AV—Attack Vector} = \left\{ \begin{array}{ll} \text{Network} & 0.85 \\ \text{Adjacent} & 0.62 \\ \text{Local} & 0.55 \\ \text{Physical} & 0.2 \end{array} \right. \\ \text{AC—Attack Complexity} = \left\{ \begin{array}{ll} \text{Low} & 0.77 \\ \text{High} & 0.44 \end{array} \right. \\ \text{PR—Privileges Required} = \left\{ \begin{array}{ll} \text{None} & 0.85 \\ \text{Low} & 0.62 \text{ (0.68)} \\ \text{High} & 0.27 \text{ (0.5)} \end{array} \right. \\ \text{UI—User Interaction} = \left\{ \begin{array}{ll} \text{None} & 0.85 \\ \text{Required} & 0.62 \end{array} \right. \\ \text{S—Scope} = \left\{ \begin{array}{l} \text{Unchanged} \\ \text{Changed} \end{array} \right. \\ \text{C—Confidentiality} = \left\{ \begin{array}{ll} \text{None} & 0 \\ \text{Low} & 0.22 \\ \text{High} & 0.56 \end{array} \right. \\ \text{I—Integrity} = \left\{ \begin{array}{ll} \text{None} & 0 \\ \text{Low} & 0.22 \\ \text{High} & 0.56 \end{array} \right. \\ \text{A—Availability} = \left\{ \begin{array}{ll} \text{None} & 0 \\ \text{Low} & 0.22 \\ \text{High} & 0.56 \end{array} \right. \end{array} \right.$$

Base metrics equations

ISS—Impact Sub-Score; Imp—Impact; Expl—Exploitability; BS—BaseScore

$$\text{ISS} = 1 - [(1 - C) \times (1 - I) \times (1 - A)]$$

$$\text{Imp} = \begin{cases} 6.42 \times \text{ISS} & \text{if } S = \text{Unchanged} \\ 7.52 \times (\text{ISS} - 0.029) - 3.25(\text{ISS} - 0.02)^{15} & \text{if } S = \text{Changed} \end{cases}$$

$$\text{Expl} = 8.22 \times \text{AV} \times \text{AC} \times \text{PR} \times \text{UI}$$

ScoreBase, BS (RUp meaning RoundUp):

$$\text{BS} = \begin{cases} 0 & \text{if } \text{Imp} = 0 \\ \text{RUp}(\min(\text{Imp} + \text{Exp}, 10)) & \text{if } S = \text{Unchanged} \\ \text{RUp}(\min(1.08 \times (\text{Imp} + \text{Exp}), 10)) & \text{if } S = \text{Changed} \end{cases}$$

Temporal metrics

$$\text{Temporal values: } \left\{ \begin{array}{l} \text{ECM—Exploit Code Maturity} = \left\{ \begin{array}{ll} \text{Not Defined} & 1 \\ \text{Unproven} & 0.91 \\ \text{Proof-of-Concept} & 0.94 \\ \text{Functional} & 0.97 \\ \text{High} & 1 \end{array} \right. \\ \text{RL—Remediation Level} = \left\{ \begin{array}{ll} \text{Not Defined} & 1 \\ \text{Official Fix} & 0.95 \\ \text{Temporary Fix} & 0.96 \\ \text{Workaround} & 0.97 \\ \text{Unavailable} & 1 \end{array} \right. \\ \text{RC—Report Confidence} = \left\{ \begin{array}{ll} \text{Not Defined} & 1 \\ \text{Unknown} & 0.92 \\ \text{Reasonable} & 0.96 \\ \text{Confirmed} & 1 \end{array} \right. \end{array} \right.$$

Temporal metrics equations—Temporal Score—TS:

$$TS = RU_p(BS \times ECM \times RL \times RC)$$

Environmental metrics

$$\text{Environmental values: } \left\{ \begin{array}{l} \text{CR - Confidentiality Report} = \begin{cases} \text{Not Defined} & 1 \\ \text{Low} & 0.5 \\ \text{Medium} & 1 \\ \text{High} & 1.5 \end{cases} \\ \text{R—Integrity Requirement} = \begin{cases} \text{Not Defined} & 1 \\ \text{Low} & 0.5 \\ \text{Medium} & 1 \\ \text{High} & 1.5 \end{cases} \\ \text{AR—Availability Requirement} = \begin{cases} \text{Not Defined} & 1 \\ \text{Low} & 0.5 \\ \text{Medium} & 1 \\ \text{High} & 1.5 \end{cases} \end{array} \right.$$

Modified base metrics: MAV—Modified Attack Vector; MAC—Modified Attack Complexity; MPR—Modified Privileges Required; MUI—Modified User Interaction; MS—Modified Scope;

MC—Modified Confidentiality; MI—Modified Integrity; MA—Modified Availability.

MISS—Modified Impact Sub-Score; MI—Modified Impact; ME—Modified Exploitability

$$MISS = \min(1 - [(1 - CR \times MC) \times (1 - IR \times MI) \times (1 - AR \times MA)], 0.915)$$

$$MI = \begin{cases} 6.42 \times MISS & \text{if MS = Unchanged} \\ 7.52 \times (MISS - 0.029) - 3.25(MISS \times 0.9731 - 0.02)^{13} & \text{if MS = Changed} \end{cases}$$

$$ME = 8.22 \times MAV \times MAC \times MPR \times MUI$$

Environmental metrics equations — Environmental Score—ES:

$$ES = \begin{cases} 0 & MI = 0 \\ RU_p\{RU_p[\min(MI + ME, 10)] \times ECM \times RL \times RC\} & MS = \text{Unchanged} \\ RU_p\{RU_p[\min(1.08 \times (MI + ME), 10)] \times ECM \times RL \times RC\} & MS = \text{Changed} \end{cases}$$

Appendix B. CVSS Version 4.0 Calculator

Source: <https://www.first.org/cvss/v4.0/specification-document> (accessed on 1 September 2023).

CVSS v4.0, to be released soon, contains many new features, and will have finer granularity (271 possible scores, instead of the actual 101).

For the base metrics, a new base metric, attack requirements (AT) has been added, and the user interaction (UI) has new metric values, active and passive. Scope has been retired from the impact metrics, and there is explicit assessment of impact to vulnerable systems (VC, VI, VA) and subsequent systems (SC, SI, SA).

The temporal metric group has been renamed the threat metric group, with the remediation level (RL) and report confidence (RC) retired and the exploit code maturity renamed exploit maturity (E)

Safety metric values have been added to the environmental metrics. The application of the environmental and threat metrics is the responsibility of the CVSS consumer.

To stress that CVSS is not just a base score, new nomenclature was adopted:

- CVSS-B: CVSS Base Score
- CVSS-BT: CVSS Base + Threat Score
- CVSS-BE: CVSS Base + Environmental Score
- CVSS-BTE: CVSS Base + Threat + Environmental Score

Details in <https://www.first.org/cvss/v4-0/> (accessed on 1 September 2023) and Dugal and Rich [70]. Further information at https://github.com/FIRSTdotorg/cvss-v4-calculator/blob/main/cvss_lookup.js (accessed on 1 September 2023).

Base Metrics ?

Exploitability Metrics

Attack Vector (AV):

Network (N)

Adjacent (A)

Local (L)

Physical (P)

Attack Complexity (AC):

Low (L)

High (H)

Attack Requirements (AT):

None (N)

Present (P)

Privileges Required (PR):

None (N)

Low (L)

High (H)

User Interaction (UI):

None (N)

Passive (P)

Active (A)

Vulnerable System Impact Metrics

Confidentiality (VC):

High (H)

Low (L)

None (N)

Integrity (VI):

High (H)

Low (L)

None (N)

Availability (VA):

High (H)

Low (L)

None (N)

Subsequent System Impact Metrics

Confidentiality (SC):

High (H)

Low (L)

None (N)

Integrity (SI):

High (H)

Low (L)

None (N)

Availability (SA):

High (H)

Low (L)

None (N)

Supplemental Metrics ?

Safety (S):

Not Defined (X)

Negligible (N)

Present (P)

Automatable (AU):

Not Defined (X)

No (N)

Yes (Y)

Recovery (R):

Not Defined (X)

Automatic (A)

User (U)

Irrecoverable (I)

Value Density (V):

Not Defined (X)

Diffuse (D)

Concentrated (C)

Vulnerability Response Effort (RE):

Not Defined (X)

Low (L)

Moderate (M)

High (H)

Provider Urgency (U):

Not Defined (X)

Clear

Green

Amber

Red

Figure A1. CVSS v4.0 Base and supplemental metrics (source: Dugal and Rich [70]). (Calculator v4.0 full description at <https://redhatproductsecurity.github.io/cvss-v4-calculator/> (accessed on 1 September 2023)).

Base Metric Group

Exploitability Metrics

Attack Vector

Attack Complexity

Attack Requirements

Privileges Required

User Interaction

Impact Metrics

Vulnerable System Confidentiality

Vulnerable System Integrity

Vulnerable System Availability

Subsequent System Confidentiality

Subsequent System Integrity

Subsequent System Availability

Threat Metric Group

Exploit Maturity

Figure A2. CVSS v4.0 base and threat metrics (source: <https://www.first.org/cvss/v4-0/> (accessed on 1 September 2023)). Copyright © 2023 Forum of Incident Response and Security Teams, Inc. All Rights Reserved).

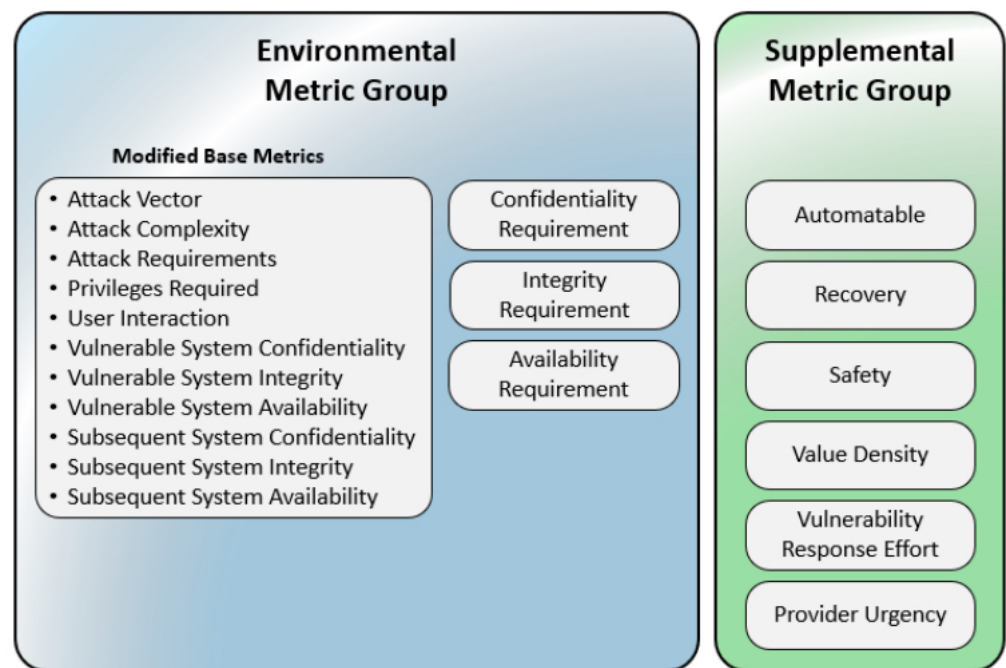


Figure A3. CVSS v4.0 Environmental and supplemental metrics (source: <https://www.first.org/cvss/v4-0/> (accessed on 1 September 2023). Copyright © 2023 Forum of Incident Response and Security Teams, Inc. All Rights Reserved).

Appendix C

In Table A1, vulnerabilities “officially” exploited, data collected [8] from 1999 to 2022, are shown. These apparently record only post-disclosure exploits of the vulnerability, not accounting for “0-day” exploits in the pre-disclosure highest risk period.

Table A1. Vulnerabilities exploited.

Year	No. of Vulnerabilities	No. of Exploits					Total	% Exploits
		1	2	3	4	5		
1999	894	2	0	0	0	0	2	0.22
2000	1020	0	0	0	0	0	0	0
2001	1677	2	0	0	0	0	2	0.12
2002	2156	1	0	0	0	0	1	0.05
2003	1527	5	0	0	0	0	5	0.33
2004	2451	5	0	0	0	0	5	0.2
2005	4935	12	1	0	0	0	13	0.26
2006	6610	26	2	0	0	0	28	0.42
2007	6520	43	1	0	0	0	44	0.67
2008	5632	66	3	0	1	0	70	1.24
2009	5736	430	48	4	0	0	482	8.4
2010	4653	685	390	8	3	0	1086	23.34
2011	4155	312	117	1	2	0	432	10.4
2012	5297	493	58	2	2	0	555	10.48
2013	5191	176	12	2	0	0	190	3.66
2014	7939	336	23	4	1	1	365	4.6
2015	6504	116	6	0	0	0	122	1.88
2016	6454	1	0	0	0	0	1	0.02
2017	14,714	4	1	0	0	0	5	0.03
2018	16,557	4	0	0	0	0	4	0.02
2019	17,344	13	0	0	0	0	13	0.07
2020	18,325	50	7	0	0	0	57	0.31
2021	20,171	0	0	0	0	0	0	0
2022	19,500	0	0	0	0	0	0	0
Total	185,962	2782	669	21	9	1	3482	1.87

Appendix D. Zipf and Pareto Extensions

The Zipf law and the Pareto I law are in a strict sense power laws, respectively, discrete and absolutely continuous, and are closely related. Both have useful generalizations, namely the generalized Pareto described in Section 4.

Zipf's law is a special case $q = 0$ of the Zipf–Mandelbrot family of discrete models with a probability mass function $\mathbb{P}[X = k] = \frac{(k+q)^{-\alpha}}{H_{N,q,\alpha}}$, $k = 1, \dots, N$, $q \geq 0$, $H_{N,q,\alpha} = \sum_{j=1}^N \frac{1}{(j+q)^\alpha}$, where k is the rank of the data and $H_{N,q,\alpha}$ is a generalization of the harmonic number.

The CDF is $F(x) = \frac{H_{\lfloor x \rfloor, q, \alpha}}{H_{N, q, \alpha}}$, where $\lfloor x \rfloor$ is the largest integer not greater than x .

The interested reader may find detailed discussion in Johnson et al. [71], and varied useful information in Sornette [41] and Saichev et al. [72].

The Pareto I family of random variables with CDF (1) has several generalizations with additional parameters providing more flexibility:

- $X \sim \text{Pareto II}(\lambda, \delta, \alpha)$, $\lambda \in \mathbb{R}$, $\delta > 0$, with CDF

$$F(x) = \begin{cases} 0 & , x < \lambda \\ 1 - \left(1 + \frac{x - \lambda}{\delta}\right)^{-\alpha} & , x \geq \lambda \end{cases}$$

with the special case $\lambda = 0$ being the Lomax distribution;

- $X \sim \text{Pareto III}(\lambda, \delta, \gamma)$, $\lambda \in \mathbb{R}$, $\delta, \gamma > 0$, with CDF

$$F(x) = \begin{cases} 0 & , x < \lambda \\ 1 - \left[1 + \left(\frac{x - \lambda}{\delta}\right)^{\frac{1}{\gamma}}\right]^{-1} & , x \geq \lambda \end{cases}$$

- $X \sim \text{Pareto IV}(\lambda, \delta, \gamma, \alpha)$, $\lambda \in \mathbb{R}$, $\delta, \gamma > 0$, with CDF

$$F(x) = \begin{cases} 0 & , x < \lambda \\ 1 - \left[1 + \left(\frac{x - \lambda}{\delta}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} & , x \geq \lambda \end{cases}$$

(Note that $\text{Pareto IV}(\delta, \delta, 1, \alpha)$ reduces to $\text{Pareto I}(\delta, \alpha)$, $\text{Pareto IV}(\lambda, \delta, 1, \alpha)$ reduces to $\text{Pareto II}(\lambda, \delta, \alpha)$, and $\text{Pareto IV}(\lambda, \delta, \gamma, 1)$ reduces to $\text{Pareto III}(\lambda, \delta, \gamma)$.)

For more information, consult Johnson et al. [60] and the monograph of Arnold [57].

The PDF of $Z = \frac{X}{Y}$, where X and Y are independent random variables, $X \sim \text{Exponential}(1)$ and $Y \sim \text{Gamma}(\alpha, 1)$, is

$$f_Z(z) = \int_0^\infty e^{-zy} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} y \, dy = \frac{1}{\Gamma(\alpha)} \int_0^\infty \left(\frac{t}{(1+z)}\right)^\alpha e^{-t} \frac{dt}{1+z} = \frac{\alpha}{(1+z)^{\alpha+1}}, \quad z > 0,$$

hence, $W = \delta(1 + Z)$ has PDF $f_W(w) = \frac{\alpha}{\delta} \left(\frac{w}{\delta}\right)^{-\alpha-1} \mathbb{I}_{[0, \infty)}(w)$, i.e., $W \sim \text{Pareto I}(\delta, \alpha)$.

Observing this, Feller [47] (p. 50) defined a family of Feller–Pareto random variables

$W = \lambda + \delta \left(\frac{X}{Y}\right)^\gamma \sim \text{FellerPareto}(\lambda, \delta, \alpha_1, \alpha_2, \gamma)$, where X and Y are independent random variables $X \sim \text{Gamma}(\alpha_1, 1)$ and $Y \sim \text{Gamma}(\alpha_2, 1)$, with special cases:

- $\text{FellerPareto}(\delta, \delta, 1, \alpha, 1) \equiv \text{Pareto I}(\delta, \alpha)$;
- $\text{FellerPareto}(\lambda, \delta, 1, \alpha, 1) \equiv \text{Pareto II}(\lambda, \delta, \alpha)$;

- $\text{FellerPareto}(\lambda, \delta, 1, 1, \gamma) \equiv \text{Pareto III}(\lambda, \delta, \gamma);$
- $\text{FellerPareto}(\lambda, \delta, 1, \alpha, \gamma) \equiv \text{Pareto IV}(\lambda, \delta, \gamma, \alpha).$

References

1. CVSS. Common Vulnerability Scoring System, Version 3.1. Available online: <https://first.org/cvss/calculator/3.1> (accessed on 7 August 2023).
2. Frei, S.; May, M.; Fiedler, U.; Plattner, B. Large-scale vulnerability analysis. In Proceedings of the 2006 SIGCOMM Workshop on Large-Scale Attack Defense, ACM, Pisa, Italy, 11–15 September 2006; pp. 131–138.
3. Holm, H. A large-scale study of the time required to compromise a computer system. *IEEE Trans. Dependable Secur. Comput.* **2013**, *11*, 2–15. [CrossRef]
4. Allodi, L. The heavy tails of vulnerability exploitation. In *Engineering Secure Software and Systems, ESSoS 2015. Lecture Notes in Computer Science 8978*; Piessens, F., Caballero, J., Bielova, N., Eds.; Springer: Berlin, Germany, 2015.
5. Ruohonen, J. A look at the time delays in CVSS vulnerability scoring. *Appl. Comput. Inform.* **2019**, *15*, 129–135. [CrossRef]
6. Rachev, S.T.; Resnick, S. Max-geometric infinite divisibility and stability. *Stoch. Model.* **1991**, *7*, 191–218. [CrossRef]
7. Feldmann, A.; Whitt, W. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Perform. Eval.* **1998**, *31*, 245–279. [CrossRef]
8. CVEdetails. Available online: <https://www.cvedetails.com/> (accessed on 1 August 2023).
9. Akaike, H. Factor analysis and AIC. *Psychometrika* **1987**, *52*, 317–332. [CrossRef]
10. Schwarz, G.E. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
11. Vuong, Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **1989**, *57*, 307–333. [CrossRef]
12. Tunggal, A.T. What is a Vulnerability? Definition + Examples, 2022. Available online: <https://www.upguard.com/blog/vulnerability> (accessed on 1 August 2023).
13. Eling, M.; Elvedi, M.; Falco, G. The economic impact of extreme cyber risk scenarios. *N. Am. Actuar. J.* **2022**, *27*, 429–443. [CrossRef]
14. Vulnerability Database. Available online: <https://vulldb.com/> (accessed on 7 August 2023).
15. National Vulnerability Database. Available online: <https://nvd.nist.gov/general/nvd-dashboard> (accessed on 7 August 2023).
16. Exploit Database. Available online: <https://www.exploit-db.com> (accessed on 7 August 2023).
17. Anderson, R. Why information security is hard — an economic perspective. In Proceedings of 17th Annual Computer Security Applications Conference (ACSAC), New Orleans, USA, 10–14 December 2001.
18. Bollinger, J. Economies of disclosure. *SIGCAS Comput. Soc.* **2004**, *34*, 1. [CrossRef]
19. Dubendorfer, T.; Wagner, A.; Plattner, B. An economic damage model for large-scale internet attacks. In Proceedings of the 13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'04), Modena, Italy, June 14–16, 2004; pp. 223–228.
20. Kannan, K.; Telang, R. An economic analysis of market for software vulnerabilities. In Proceedings of the Third Workshop on the Economics of Information Security, Minneapolis, MN, USA, 13–14 May 2004.
21. Qualys Research Report Laws of Vulnerabilities 2005. Available online: <http://www.qualys.com/docs/Laws-Report.pdf> (accessed on 7 August 2023).
22. Anderson, R.; Moore, T. The economics of information security. *Science* **2006**, *314*, 610–613. [CrossRef]
23. Frei, S. Security Econometrics — The Dynamics of (In)security. Doctoral Dissertation, ETH ZURICH, Zurich, Switzerland, 2009.
24. EPSS—Exploit Prediction Scoring System. Available online: <https://www.first.org/epss/model> (accessed on 7 August 2023).
25. Jacobs, J.; Romanosky, S.; Edwards, B.; Roytman, M.; Adjerid, I. Exploit Prediction Scoring System (EPSS). *Digit. Threat. Res. Pract.* **2021**, *2*, 1–17. [CrossRef]
26. CTI—Cyber Threat Intelligence. Available online: <https://fidelissecurity.com/resources/edu/cyber-threat-intelligence/> (accessed on 7 August 2023).
27. CWSS—Common Weakness Scoring System. Available online: <https://cwe.mitre.org/index.html> (accessed on 7 August 2023).
28. SSVc—Stakeholder-Specific Vulnerability Categorization system. Available online: <https://www.cisa.gov/stakeholder-specific-vulnerability-categorization-ssvc> (accessed on 7 August 2023).
29. Frühwirth, C.; Männistö, T. Improving CVSS-based vulnerability prioritization and response with context information. In Proceedings of 3rd International Symposium on Empirical Software Engineering and Measurement, Florida, USA, 15–16 October 2009; pp. 535–544. Available online: <https://ieeexplore.ieee.org/abstract/document/5314230> (accessed on 1 August 2023). [CrossRef]
30. Frei, S. Vulnerability Management. Master's Thesis, Department of Management Technology and Economics, ETH ZURICH, Zurich, Switzerland, 2007.
31. Shahzad, M.; Shafiq, M.Z.; Liu, A.X. A large scale exploratory analysis of software vulnerability life cycles. In Proceedings of the 34th International Conference on Software Engineering, Zurich, Switzerland, 2–9 June 2012; pp. 771–781.
32. CVE—Common Vulnerabilities and Exposures. Available online: <http://cve.mitre.org/cve/> (accessed on 7 August 2023).
33. Boechat, F.; Ribas, G.; Senos, L.; Bicudo, M.; Nogueira, M.; Aguiar, L.; Menasché, D. Is vulnerability report confidence redundant? Pitfalls using temporal risk scores. *IEEE Secur. Priv.* **2021**, *19*, 44–53. [CrossRef]
34. Wilk, M.B.; Gnanadesikan, R. Probability plotting methods for the analysis of data. *Biometrika* **1968**, *55*, 1–17. [CrossRef] [PubMed]

35. Burnham, K.; Anderson, D. *Model Selection and Multimodel Inference: A Practical Information—Theoretic Approach*; Springer: New York, NY, USA, 2002.
36. Nurmi, D.; Brevik, J.; Wolski, R. Modeling machine availability in enterprise and wide-area distributed computing environments. In *Euro-Par 2005 Parallel Processing, Lecture Notes in Computer Science*; Cunha, J., Medeiros, P., Eds.; Springer: Berlin, Germany, 2005; Volume 3648, pp. 432–441.
37. Symantec. Available online: <https://www.symantec.com/> (accessed on 7 August 2023).
38. Aaberge, R. Gini's nuclear family. *J. Econ. Inequal.* **2007**, *5*, 305–322. [[CrossRef](#)]
39. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-law distributions in empirical data. *SIAM Rev.* **2009**, *51*, 661–703. [[CrossRef](#)]
40. Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*; Reedition Dover Publications: Mineola, New York, USA, 2009; Freeman: New York, NY, USA, 1991.
41. Sornette, D. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-organization and Disorder: Concepts and Tools*, 2nd ed.; Springer Series in Synergetics; Springer: Heidelberg, Germany, 2006.
42. Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351. [[CrossRef](#)]
43. Belevitch, V. On the statistical laws of linguistic distributions. *Ann. Soc. Sci. Bruxelles, Ser. I* **1959**, *73*, 310–326.
44. Mitzenmacher, M. A brief history of generative models for power law and lognormal distributions. *Internet Math.* **2004**, *1*, 226–251. [[CrossRef](#)]
45. Bee, M. On discriminating between lognormal and Pareto tail: An unsupervised mixture-based approach. *Adv. Data Anal. Classif.* **2022**, *1*–9. [[CrossRef](#)]
46. Karamata, J. Sur un mode de croissance régulière des fonctions. *Mathematica* **1930**, *4*, 38–53.
47. Feller, W. *An Introduction to Probability Theory and its Applications*, 2nd ed.; Wiley: New York, NY, USA, 1971; Volume 2.
48. Kevei, P. Regularly Varying Functions. 2019. Available online: https://www.math.u-szeged.hu/~kevei/tanitas/1819regvar/RegVar_notes.pdf (accessed on 1 August 2023).
49. Bingham, N.H.; Goldie, C.M.; Teugels, J.L. *Regular Variation*; Cambridge University Press: Cambridge, UK, 1989.
50. Hall, P. On some simple estimates of an exponent of regular variation. *J. R. Stat. Soc. Ser. B* **1982**, *44*, 37–42. [[CrossRef](#)]
51. Davis, R.; Resnick, S. Tail estimates motivated by extreme value theory. *Ann. Stat.* **1984**, *12*, 1467–1487. [[CrossRef](#)]
52. Fedotenkov, I. A review of more than one hundred Pareto-tail index estimators. *Statistica* **2020**, *80*, 245–299. [[CrossRef](#)]
53. Stumpf, M.P.H.; Porter, M.A. Critical truths about power laws. *Science* **2012**, *335*, 665–666. [[CrossRef](#)] [[PubMed](#)]
54. The Econophysics Blog. Tyranny of the Power Law (and Why We Should Become Eclectic). 2006. Available online: <http://econophysics.blogspot.com/2006/07/tyranny-of-power-law-and-why-we-should.html> (accessed on 1 August 2023).
55. Shalizi, C.R. So You Think You Have a Power Law? Well Isn't that Special? 2007. Available online: bactra.org/weblog/491.html (accessed on 1 August 2023).
56. Box, G.E.P. Robustness in the strategy of scientific model building. In *Robustness in Statistics*; Launer, R.L., Wilkinson, G.N., Eds.; Academic Press: Cambridge, MA, USA, 1979; pp. 201–236.
57. Arnold, B.C. *Pareto Distributions*, 2nd ed.; Chapman and Hall/CRC: New York, NY, USA, 2015.
58. Embrechts, P.; Klüppelberg, C.; Mikosch, T. *Modelling Extremal Events for Insurance and Finance*; Springer Science & Business Media: Berlin, Germany, 1997; Volume 33.
59. Andriani, P.; McKelvey, B. Beyond Gaussian averages: Redirecting international business and management research toward extreme events and power laws. *J. Int. Bus.* **2007**, *38*, 1212–1230. [[CrossRef](#)]
60. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1994; Volume 1.
61. Beirlant, J.; Teugels, J.L.; Vynckier, P. *Practical Analysis of Extreme Values*; Leuven University Press: Leuven, Belgium, 1996.
62. Gomes, M.I.; Guillou, A. Extreme value theory and statistics of univariate extremes: A review. *Int. Stat. Rev.* **2015**, *83*, 263–292. [[CrossRef](#)]
63. Cline, D. Convolution tails, product tails and domains of attraction. *Probab. Theor.* **1986**, *72*, 529–557. [[CrossRef](#)]
64. Malik, H.J. Estimation of the parameters of the Pareto distribution. *Metrika* **1970**, *15*, 126–132. [[CrossRef](#)]
65. Bauke, H. Parameter estimation for power-law distributions by maximum likelihood methods. *Eur. Phys. J. B* **2007**, *58*, 167–173. [[CrossRef](#)]
66. Diaz, F.J. Identifying tail behavior by means of residual quantile functions. *J. Comput. Graph. Stat.* **1999**, *8*, 493–509. [[CrossRef](#)]
67. Arnold, B.C.; Brockett, P.L. When does the β th percentile residual life function determine the distribution? *Oper. Res.* **1983**, *31*, 391–396. [[CrossRef](#)]
68. Davison, A.C.; Hinkley, D.V. *Bootstrap Methods and their Application*; Cambridge University Press: Cambridge, UK, 1997.
69. Glaz, J.; Sison, C.P. Simultaneous confidence intervals for multinomial proportions. *J. Stat. Plan. Inference* **1999**, *82*, 251–262. [[CrossRef](#)]
70. Dugal, D.; Rich, D. Announcing CVSSv4.0. In Proceedings of the 35th Annual FIRST Conference, Montréal, QC, Canada, 4–9 June 2023.
71. Johnson, N.L.; Kotz, S.; Kemp, A.W. *Univariate Discrete Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1992.
72. Saichev, A.; Malevergne, Y.; Sornette, D. Theory of Zipf's law and beyond. In *Lect. Notes Econ. Math. Systems*; Springer: Berlin, Germany, 2010; Volume 632.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.