A Queueing Solution to Reduce Delay in Processing of Disclosed Vulnerabilities

Andrew Feutrill
School of Mathematical Sciences
CSIRO's Data61 and University of Adelaide
Adelaide, Australia
andrew.feutrill@data61.csiro.au

Matthew Roughan
School of Mathematical Sciences
University of Adelaide
Adelaide, Australia
matthew.roughan@adelaide.edu.au

Joshua Ross
School of Mathematical Sciences
University of Adelaide
Adelaide, Australia
joshua.ross@adelaide.edu.au

Yuval Yarom
School of Computer Science
University of Adelaide and CSIRO's Data61
Adelaide, Australia
yval@cs.adelaide.edu.au

***Abstract*—The rate of discovery of vulnerabilities keeps increasing, creating a problem for first responders who need to triage vulnerabilities quickly to decide where to focus their defensive efforts. One of the bottlenecks in this triaging process is the assessment of severity of vulnerabilities and the assignment of the Common Vulnerability Scoring System (CVSS) scores.**

In this work we study the statistical properties of the vulnerability disclosure process and make two important observations. First, we find that the time series of the number of vulnerability disclosures exhibits a long range dependence, meaning that strong correlations persist over long time periods. Such time series have high variation, high burstiness and slow convergence towards conventional estimators, such as the mean.

Our second observation is that the burstiness of the vulnerability disclosure process causes delays in the analysis of vulnerabilities and as a result triaging over 40% of the vulnerabilities takes longer than the median exploit time. Hence, by the time they are analysed and assigned a CVSS score, many vulnerabilities are already being exploited.

We propose techniques for modelling and analysing the vulnerability disclosure time series. We further propose reversing the order of triaging vulnerabilities and show, via simulation, that this significantly increases timely triaging of vulnerabilities, reducing the percentage of delayed assessments to 4%.

***Index Terms*—long-range dependence, time series analysis, queueing theory**

I. Introduction

Security vulnerabilities are currently disclosed at the rate of around 15,000 per year [46], the time series of vulnerability disclosures is shown in Figure 1. This creates a difficult problem for those trying to manage the risk of exploitation of vulnerabilities. However, processing of vulnerabilities takes time for NIST analysts and our simulation shows since 2017 there is an over 40% probability of not being processed before the median exploit time.

This raises the question *can we improve this process?*, particularly over the short term until more resources can be assigned to this task. We answer this question in the affirmative, showing how the median time until vulnerabilities are processed is improved through a very simple, easy to implement mechanism.

We study the statistical properties of the Common Vulnerability and Exposures (CVE) disclosure process and show that the arrival process of disclosures is highly-variable and long-range dependent. These two (often associated) properties lead to undesirable workloads. This causes the queue of disclosures awaiting analysis to back up and increases the time until technical information about vulnerabilities is available for use by managers to prioritise their work.

Long-range dependence (LRD) was originally discovered by Hurst in the context of Nile river flooding [19], [39]. He discovered that the Nile had long periods of larger floods, followed by similar periods of drought. Processes with this behaviour are unintuitive compared to short-range dependent (SRD) processes. In particular, the mean of an LRD process is not as useful a characterisation of such workloads. Commonly, an incoming LRD workload will result in heavy-tailed queueing behaviour, resulting in periods where the queue will fill to an extreme extent.

We show the arrival process can be modelled as a heterogeneous superposition of individual vendor arrival processes, each modelled as a fractal renewal process. Such processes and their related models have been heavily used in modelling internet traffic workloads over the last two decades [26], [47], [48]. The model used here requires an additional dimension of heterogeneity of the individual input processes in order to match the non-Gaussian marginal distribution.

Finally we show the impact of such an arrival process on the management of vulnerabilities. We show, through simulation, that the median delay for a vulnerability to be processed by NIST can be extremely long, adding additional resources for this task during a burst is suggested in the long term. In the short term, however, an alternative strategy can yield significant improvement. Typically, workloads are treated using the FIFO (First In - First Out) queue discipline. This is

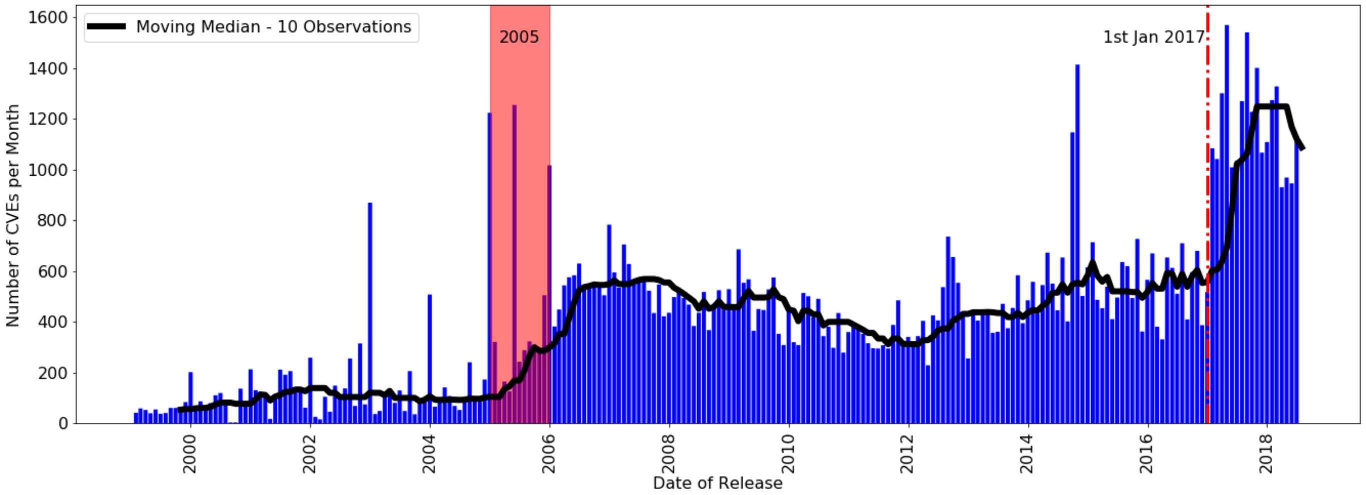


Fig. 1: Number of CVE Entries Disclosed per Month from 1999 to 2018. The CVE arrivals can be split into three distinct eras: (1) from 1999 to 2005, (2) from 2005 to 2017 and (3) from 2017 to date. The highlighted section shows the build up from establishment of the CVE process to the standardisation within the security community in 2005.

simple and is widely considered fair. However, in this problem, fairness is not the criteria of interest. A LIFO (Last In - First Out) queue has the advantage that most tasks have shorter waiting times. This comes at the cost of higher maximum waiting times but arguably this is a better situation. If all vulnerabilities take a long time to be processed, then all can be exploited. If most are processed quickly, then the majority never become fully-fledged threats. We show that the median delay in processing vulnerabilities is reduced from 2 days to a few hours by changing the queueing discipline.

II. Background and Related Work

A. CVEs

Common Vulnerability and Exposures (CVE) is a dictionary of publicly known security vulnerabilities [46] launched in September 1999, to improve operability between different vulnerability databases, and centralise all publicly known information. The dictionary gained prominence as the primary information source for vulnerabilities, after recommendation of CVE for use within government agencies in 2002 by NIST [30]. It was standardised in 2005 after the National Vulnerability Database was introduced and adopted the CVE dictionary as the source of truth for the existence of a vulnerability [46].

Time series modelling of vulnerability disclosures has been performed with the intention of discovering properties and providing accurate predictions of future volume [18], [21], [22], [33], [37], [41]–[43]. Researchers have developed time series models to model various trends [18], [22], [33], [37], [41], [43] and cycles [18], [21], [37] across the entire life of the CVE process. However, none of these studies have considered the strong correlations or identified LRD in the data.

Seasonality has been identified and modelled by several authors [18], [21], [37], who discovered weekly [21] and yearly [18], [37] cycles. The trend has been modelled linearly by Haldar and Mishra [18], however the conclusion of Roumani et al. [37] is that a linear trend cannot predict the number of disclosures and concluded the current value of the series was the most significant factor in predicting disclosures. Most of these studies have modelled all vulnerabilities disclosed to NIST, however Roumani et al. models the vulnerability disclosures for separate families of web browsers: Chrome, Internet Explorer, Firefox, Opera and Safari. Many studies have modelled the time series using Auto Regressive Integrated Moving Average models [22], [33], [37], [41], [43] and Exponential Smoothing techniques [18], [37].

Generative models have also been used to model emergence of vulnerabilities using exponential [18], [34], [35], gamma [22] and Weibull distributed [20] interarrival times. Exponential distributions were selected as modelling assumptions for Markov Chain models [34], [35] or an input to a queueing model [18]. The gamma distribution was used to model the time between disclosures as assessed by an individual analyst, as a measure of the likelihood of discovery of zero-day vulnerabilities [22]. This dataset is highly variable with only 20% of the analysts discovering over 60% of the vulnerabilities and 90% of the analysts discovering less than 10 vulnerabilities. Modelling vulnerability discovery using Weibull distributions [20] only considers 4 common operating systems and uses non heavy-tailed Weibull distributions. We show that modelling interarrival times with heavy tailed distributions generates LRD in the observed vulnerability disclosure time series.

B. Stochastic Arrival Processes

Queues have been studied since Erlang developed models of telephony networks to understand the probability that a telephone call is blocked [15]. An arrival process is a stochastic process $\{X_t\}_{t\in T}$, where X_i denotes the time between the

$(i-1)$th and ith arrivals. For telephony models the arrival process of jobs is modelled using a Poisson process, *i.e.* exponentially distributed interarrival times. Extensive work has been done modelling telecommunication systems with this assumption, with a single source with exponentially distributed interarrivals [15] or a superposition of multiple arrival sources [40].

An important class of arrival process models, for strongly correlated data, are LRD models. A time series, X_t, with an autocorrelation function, $\rho(k)$ for a lag k, has LRD if $\sum_{k=0}^{\infty} \rho(k) = \infty$ [6]. Otherwise, we say the time series is SRD. LRD models were developed for telecommunication data by Berger and Mandelbrot when studying errors within telephone networks [7]. LRD models were developed as existing models with exponentially distributed interarrival times, did not have properties such as self similarity on multiple time scales and high variance in ethernet packet flows [26]. Models of ethernet network data, with strong correlations and extreme values, utilised interarrival times with power-law interarrival times to replicate the large observed variance and self-similarity [5], [11], [26], [38]. It was shown that a superposition of many on/off processes with power-law distributed interarrival times generates LRD from multiple sources [45]. LRD arrival processes for queues have been shown to behave qualitatively different from SRD processes [16], [32], [36]. These models show the influence of heavy tailed distributions in different parts of a queueing system increases the queue length required for finite queues, and the mean time to traverse the system, in response to the bursty traffic.

Queueing models to describe the vulnerability disclosure process were considered by Haldar and Mishra [18], who modelled emergence of vulnerabilities and expected time to patching, using a multiple server queue. They assume vulnerability arrivals come from an infinite pool of sources; and, that patches are developed by a fixed finite amount of servers k and selected from a queue of unmitigated vulnerabilities. Improvements can be made by considering patches to be developed by the vendor only.

III. Arrival Process of Vulnerabilities

A. Data

The observed arrival process for CVEs from 1999-2018 is shown in Figure 1. We split the process into three eras: the establishment era from 1999-2005; the standardisation era from 2005-2016; and the current era from 2017-present. We define a process $Y(n)$, as the number of disclosed vulnerabilities in a week n, this time resolution was chosen due to weekly cycles and low confidence in the accuracy of disclosure times. The marginal distribution of the second era of $Y(n)$ is shown in Figure 2. The mean and variance of the number of disclosed CVEs per week for the marginal distribution of this time series are 113.64 and 3324.54 respectively. We fitted a Normal distribution of the observed marginal distribution with a Kolmogorov-Smirnov test statistic value of 0.10, and p-value of 0.000015, hence we rejected the hypothesis of Gaussian distributed data. A Log-Normal distribution was fitted to the observed marginal distribution, with a Kolmogorov-Smirnov test statistic value of 0.04, and corresponding p-value of 0.32. So we cannot reject the hypothesis that the marginal distribution is Log-Normal.

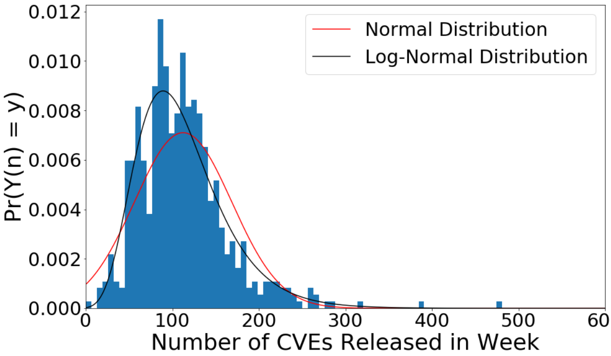


Fig. 2: Marginal distribution of the second era of CVE entry time series, 2005-2017. Normal and Log-Normal distributions are shown fitted to the observed observed data.

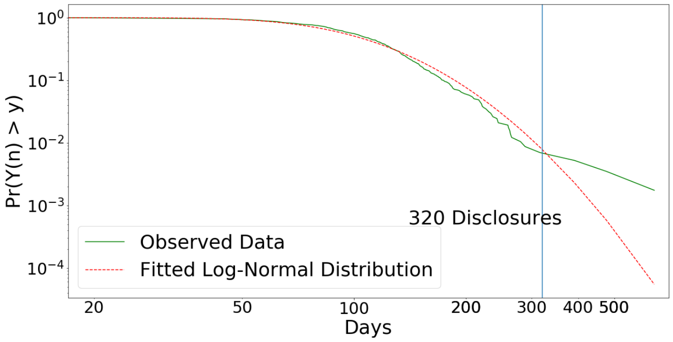


Fig. 3: Complementary Cumulative Distribution Function (CCDF) of the marginal distribution of the second era of the CVE entry time series on a log-log scale. A fitted Log-Normal distribution is shown, with a power-law tail emerging at approximately 320 disclosures.

The tail of the marginal distribution is heavier than a standard Log-normal distribution. Figure 3 shows the Complementary Cumulative Distribution Function (CCDF) on a log-log scale. The tail diverges from Log-Normal at approximately 320 disclosures in a week. The large bursts of disclosures that have been observed in Figure 1, are responsible for the heavy tail. Given these observations, we will model the marginal distribution as a Log-Normal distribution with a heavy tail, as it provides a parsimonious description for the simulation in Section V.

This data is the aggregation of 17,619 vendors that have contributed CVEs since 1999. The distribution of the number of disclosed CVEs per vendor is shown in Figure 4, with the x-axis the rank (*i.e.* vendors ordered by number of CVEs) vs. the normalised frequency of their occurrence. This distribution is highly variable with a small number of vendors contributing a large proportion of the CVEs. For example, the top 10 vendors account for 31% of all CVEs and the top 100 account for

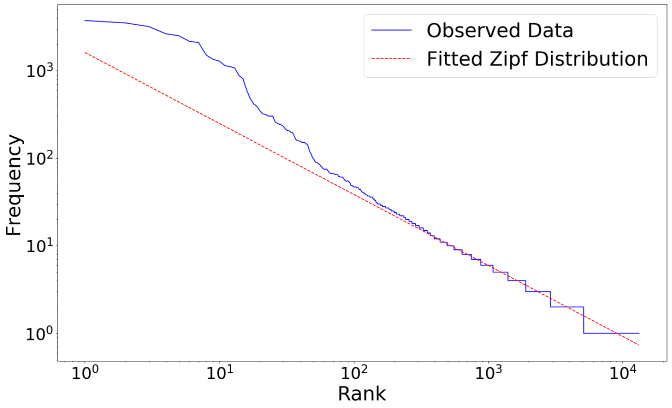


Fig. 4: Zipf distribution fitted to the observed rank of vendors vs. the number of CVEs entries on a log-log scale. Showing a good fit to a power-law, indicating that the vendor distribution is highly heterogeneous.

54% of all CVEs. A Zipf distribution has a probability mass function for the kth ranked vendor,

$$p(k; s, N) \sim \frac{1}{k^s}.$$

As $p(k; s, N)$ decays as a power of s, rank vs. frequency forms a straight line on a log-log plot. Figure 4 shows a fitted Zipf distribution to the volume of disclosed vulnerabilities per vendor. As the rank vs. frequency plot follows a power law, we conclude that the vendors are highly heterogeneous and best modelled by using an aggregation of non-identical sources, in contrast to typical internet traffic models [5], [11], [26], [38].

The arrival process shown in Figure 1, can be decomposed into individual vendor disclosure processes. These arrival processes are not identical, as they have very different rates of disclosure. We model each vendor disclosing a maximum of one batch of disclosures per day.

B. Measurement of Long Range Dependence

We suspect that the process of number of disclosures may be LRD, due to the high variation observed. For a stationary LRD process the autocorrelation function, which is defined as

$$\rho(k) = \frac{E[(X(n+k) - E[X])(X(n) - E[X])]}{\sigma_{n+k}\sigma_n},$$

asymptotically decays as a power-law, *i.e*, $\rho(k) \approx c_\rho |k|^{-a}$ for $a \in (0, 1)$, as $k \to \infty$.

LRD and statistical self-similarity are measured using the Hurst parameter, defined by Hurst to describe the time series of the water level of the Nile River [19]. It takes values between 0 and 1, and if $0.5 < H < 1$ the process exhibits LRD [6]. The Hurst parameter and the exponent parameter of the autocorrelation function are related, a, $H = (1+a)/2$ [6].

The Hurst parameter has been measured in a variety of ways, for instance as the log-linear regression on the autocorrelation function. To estimate the Hurst parameter we use the MATLAB code provided online by Veitch [1], implementing the wavelet-based estimator of Abry and Veitch [3].

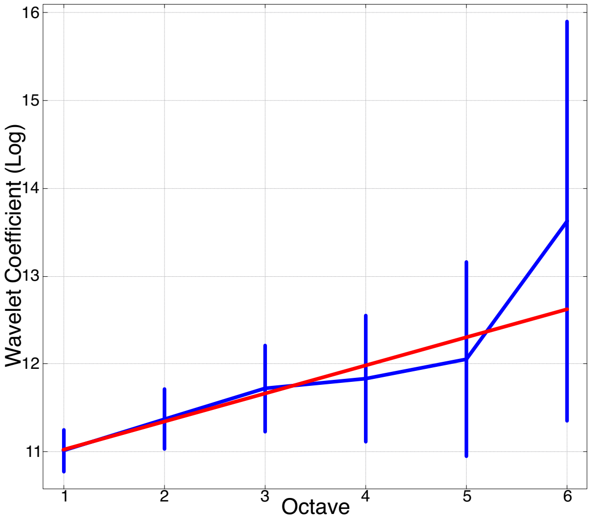


Fig. 5: Log-scale diagram vs. the average of wavelet coefficients at that octave. The blue bars are the 95% confidence intervals of the variance of wavelet coefficients. The linear regression on the average wavelet coefficients is shown in red, and the slope of the red line is $2H$ - 1. From this, the H parameter is estimated to be 0.66, with a confidence interval of [0.58, 0.74].

The code performs a discrete wavelet transformation of the input data, decomposing the sequence in terms of wavelet basis functions on a dyadic grid in the transform space. In this space an octave corresponds to a $\log_2$-scale, and hence when we calculate the variance of *detail* coefficients, $d_x(j, k)$, across a given octave, j, we obtain a measure of burstiness at that scale. Plotting the log-variances across octaves creates a log-scale diagram. LRD and self-similarity are characterised in this diagram by a straight line, which has a slope of $2H$-1.

Thus we can estimate H by performing linear regression on the log-scale plot. 95% Confidence intervals are calculated for each octave in the plot, which enable us to be able to assess the accuracy of the regression. Figure 5 shows the log-scale diagram for the number of disclosed vulnerabilities, with 95% confidence intervals shown as the blue bars at each octave and a linear fit shown in red. The estimate of the Hurst parameter is 0.66 with a confidence interval of [0.58,0.74], and the linear regression is contained within the 95% confidence intervals at each octave, indicating an excellent fit to the data. From this we conclude that the process is LRD. Therefore to construct mathematical models we must incorporate LRD and hence our choice of fractal renewal processes, a common generative model for LRD.

IV. Model Definition

We define an arrival process that is parsimonious while generating the observed properties of the previous section, long range dependence, a heavy-tailed marginal distribution, and a highly heterogeneous vendor arrival processes.

It has been established that heavy-tailed marginal distributions or LRD can be generated by the superposition of simpler arrival processes [27], [45]. However, it is unusual to see both of these phenomena, especially in conjunction with the heterogeneity of the individual processes. We elect to define a renewal process model with heavy-tailed distributions, to induce the observed LRD. This approach was chosen as these provide parsimonious models of LRD data, which are simpler and require fewer parameters. This is in contrast to other possibilities such as Markov Arrival Processes, and in particular Markov Modulated Poisson Processes or Phase type distributions.

We begin by defining a model for the arrival times of an individual vendor. Renewal processes are used to model the time between successive events such as time between phone calls, packet arrivals and component failures. A renewal process is a sequence of arrival times, $(S_n)_{n\in\mathbb{N}}$. Where we define $S_n = \sum_{i=1}^{n} X_i$, as the sum of the interarrival times with $X_1, X_2, X_3, X_4, \ldots$ a sequence of positive, independent and identically distributed random variables, such that $0 < E[X_i] < \infty$ for all i. There are two different random processes of interest, the arrival time process and the counting process. The arrival time process $(S_n)_{n\in\mathbb{N}}$ is a point process on $[0,\infty)$. The counting process counts the number of arrivals up to a particular time, t, and is defined as $N(t) = \sum_{n=1}^{\infty} \mathbb{1}_{S_n \leq t} = \sup\{n : S_n \leq t\}$. Fractal renewal processes are a subset of renewal processes where the interarrival times, X_i, follow a power-law probability distribution [29], which has probability density function

$$f(x) = \frac{\alpha-1}{x_{\text{min}}}\left(\frac{x}{x_{\text{min}}}\right)^{-\alpha}.$$

This distribution exhibits scaling properties since,

$$f(cx) = \frac{\alpha-1}{x_{\text{min}}}\left(\frac{cx}{x_{\text{min}}}\right)^{-\alpha} = c^{-\alpha} f\left(x\right).$$

Fractal renewal processes exhibit self-similarity due to this power-law scaling property [28]. We choose fractal renewal processes to model the individual vendors to capture the large variation, burstiness and to generate LRD in the superposition of many vendors.

As each vendor is different, we need a sequence of interarrivals for vendor i, is given by process $\{V_j^{(i)}\}_{j\in\mathbb{N}}$, which is sampled from a power law distribution, $V^{(i)}$, with

$$\Pr(V_j^{(i)} \leq x|\alpha_i) = 1 - \left(\frac{x}{x_{\text{min}^{(i)}}}\right)^{-\alpha_i},$$

with arrival times given by,

$$S_j^{(i)} = \sum_{k=0}^{j} V_j^{(i)}.$$

At each arrival time multiple CVEs can be disclosed by a vendor. That is, at time $S_j^{(i)}$ a batch of $B_j^{(i)}$ CVEs is disclosed by vendor j with support on $\mathbb{N}$. Where,

$$\Pr(B_j^{(i)} \leq x|\beta_i) = 1 - \frac{\zeta(\beta_i, x)}{\zeta(\beta_i, x_{\text{min}^{(i)}})},$$

where,

$$\zeta(\beta, x) = \sum_{n=0}^{\infty}(n+x)^{-\beta}.$$

The exponents of the power-law distribution for both the batch size, β_i,and the interarrival time, α_i, are vendor dependent, to model heterogeneous vendors. Due to the heterogeneity in the number of disclosed vendors as shown in Section III, we selected non-identical distributed vendors to generate the observed properties.

We now define, $W_k^{(i)}$ as the number of disclosed vulnerabilities for a vendor i in week k. Let T be the length of one week,

$$W_k^{(i)} = \sum_{j=1}^{\infty} B_j^{(i)} \mathbb{1}_{\{(k-1)T \leq V_j^{(i)} \leq kT\}},$$

and let $\overset{*}{W}_k$ be the superposition of the vendor processes. Then for M vendors,

$$\begin{aligned}\overset{*}{W}_k &= \sum_{i=1}^{M} W_k^{(i)}, \\ &= \sum_{i=1}^{M}\sum_{j=1}^{\infty} B_j^{(i)} \mathbb{1}_{\{(k-1)T \leq V_j^{(i)} \leq kT\}}.\end{aligned}$$

From this generic framework, we define 3 different stationary renewal process models to test the influence of different features of the models. The developed models are:

- Heterogeneous interarrival distribution exponents, α_i,with identical batch size distributions;
- Identical interarrival distributions, with heterogeneous batch distribution exponents, β_i;
- Heterogeneous interarrival distribution $x_{\text{min}^{(i)}}$ parameter, with identical batch size distributions.

A. Heterogeneous interarrival distribution exponents

In this model the parameter of the interarrival distribution, α_i, for a vendor is sampled from a log-normal distribution. This distribution is fitted to the exponents of the observed vendor interarrival times. The form of the log-normal distribution used, the probability density function has the following form,

$$p(x) = \frac{1}{(x-\mu)\sigma\sqrt{2\pi}} e^{-\frac{\log(\frac{x-\mu}{\theta}^2)}{2\sigma^2}},$$

where μ, σ and θ are the location, shape and scale parameters respectively.

The fitted parameters for the log-normal model are a shape parameter of 0.537, a location parameter of 1.159 and a scale parameter of 1.537. A chi-squared test was used to determine whether this is a suitable model for the interarrival exponent distribution. The hypothesis that the distribution was taken from a log-normal distribution could not be dismissed with a chi-squared value of 18.795 and a p-value of 0.470. The x_{min} parameter for the interarrival distribution is fixed at 1 day, based on the observed exploit data. The batch size distribution in this model is the same for each vendor, which has β value of 2.75, chosen to replicate the average value of the fitted exponents from the observed batch sizes, and x_{min} of 1.

The expected number of vulnerabilities disclosed in a time, T, and number of vendors, M, is

$$
\begin{aligned}
E[\overset{*}{W}_k] &= E\left[\sum_{i=1}^{M}\sum_{j=1}^{\infty} B_j^{(i)} \mathbb{1}_{\{(k-1)T\le V_j^{(i)}\le kT\}}\right] \\
&= \frac{TME[B]}{\int_{\alpha_{min}}^{\infty} E[V|\alpha]d\alpha},
\end{aligned}
$$

for the log-normal probability density function $f(\alpha)$ of parameter α.

B. Heterogeneous batch distribution exponents

To test the whether the long range dependence is induced by high variation in the batch size distribution, a model was developed with identical interarrival distributions of the vendors and different batch size distributions. In particular their exponents are sampled from the log-normal model of the parameter space which was fitted to the observed batch size distribution. The interarrival distributions had a fixed exponent, α, for all vendors of 2.5, which was selected to replicate the exponent fitted to interarrivals for all vendors, and x_{min} of 1. This process has an expected number of vulnerabilities in a time period T of

$$
\begin{aligned}
E[\overset{*}{W}_k] &= E\left[\sum_{i=1}^{M}\sum_{j=1}^{\infty} B_j^{(i)} \mathbb{1}_{\{(k-1)T\le V_j^{(i)}\le kT\}}\right] \\
&= \frac{TM\int_{\beta_{min}}^{\infty} E[B|\beta]d\beta}{E[V]},
\end{aligned}
$$

for log-normal probability density function $f(\beta)$ of parameter β.

C. Heterogeneous Interarrival Distribution x_{min}

A model was developed with identical batch size exponents and identical interarrival exponents of the power law distributions, however the x_{min} parameter of the interarrival distribution is distributed with a log-normal distribution. From the observed data, the hypothesis of a log-normal distribution of the x_{min}'s could not be rejected, with a chi-squared value of 15.8 and a p-value of 0.67. In this case, the interarrival distribution exponent was fixed at 2.5 for all vendor processes, with the x_{min} parameter sampled from the log-normal distribution and the batch size distribution had an exponent of 3.25 and an x_{min} of 1.

This process has an expected number of disclosures in a time period T of

$$
\begin{aligned}
E[\overset{*}{W}_k] &= E\left[\sum_{i=1}^{M}\sum_{j=1}^{\infty} B_j^{(i)} \mathbb{1}_{\{(k-1)T\le V_j^{(i)}\le kT\}}\right] \\
&= \frac{TME[B]}{\int_{1}^{\infty} E[V|x_{\text{min}}]f(x_{min})dx_{\text{min}}},
\end{aligned}
$$

for log-normal probability density function $f(x_{min})$ of parameter x_{min}.

V. Simulation

Simulations of the three different models, defined in Section IV, were built to test the LRD of the vendor processes. In addition to the power-law distributed models we defined Poisson process models, with exponentially distributed interarrivals, to compare the influence of SRD models. Each simulation consists of 200 heterogeneous vendors, whose parameters were sampled from the fitted Log-Normal distributions. 200 vendors were selected as the top 200 vendors are responsible for over 60% of all disclosed vulnerabilities and account for most of the variation in the time series. For simplicity, we assumed a constant number of vendors in the simulation. They were simulated as discrete event simulations in continuous time. The continuous process was then quantised and aggregated to weekly intervals.

We use the Hurst parameter, mean and variance to compare the different models to the second era of observed arrival process, as shown in Figure 1. The process is modelled as a delayed renewal processes [12], with the first arrival is distributed as, G, and all subsequent arrivals being distributed as, F, where

$$G(x) = \frac{1}{E[X]}\int_0^x 1 - F(t)dt.$$

Hence, the first arrival distribution for a power-law distributed random variable is

$$G(x) = \frac{x_{\text{min}}^{-\alpha}}{\alpha-1}(1 - x^{-\alpha+2}),\ \text{if } \alpha > 2.$$

To understand the model performance and the influence of the log-normal parameter distributions, we define null models where the parameter is sampled from a uniform distribution on an interval between the maximum and minimum observed values from the dataset.

A. Results

The mean, variance and Hurst parameter of the simulations are shown in Table I. The aggregated arrivals of the heterogeneous interarrival exponent model are shown in Fig 8, all of these processes exhibit high variation with large bursts occurring regularly, with the largest bursts appearing for the heterogeneous batch exponent model. Varying the x_{min} parameter results in a distance of least x_{min} between subsequent arrivals, which results in a low number of vulnerabilities being disclosed in a week. This doesn't reflect the reality of the underlying vendors such as Microsoft, although they eventually developed a monthly disclosure cycle for most vulnerabilities, serious vulnerabilities are disclosed when discovered and the minimum interarrival time is daily [2].

The mean of the heterogeneous interarrival and batch exponent models, are a good fit for the observed mean. Hence, the dynamics of the process cannot be captured by the heterogeneous x_{min} model as the mean is not possible to generate enough CVE disclosures with fixed minimum

Time Series		H	CI	Mean	Variance
Observed CVE Process		0.66	[0.58, 0.74]	113.6	3324.5
Interarrival Exponent	**L-N**	**0.56**	**[0.52, 0.59]**	**119.1**	406.0
	U	0.56	[0.53, 0.6]	159.6	436.5
Batch Exponent	L-N	0.45	[0.42, 0.48]	119.6	65 738.9
	U	0.57	[0.54, 0.61]	158.6	2 403 758.7
Interarrival $x_{\min}$	L-N	0.52	[0.48, 0.55]	15.3	42.1
	U	0.49	[0.45, 0.52]	9.3	39.1
Poisson Interarrivals	L-N	0.50	[0.47, 0.54]	43.4	141.86
	U	0.49	[0.46, 0.53]	15.3	67.5

TABLE I: Simulated values of H with 95% Confidence Intervals (CI), mean and variance for all three models with Log-Normal (L-N) or Uniform (U) distributed parameter distributions. The results show that the heterogeneous interarrival exponent model generates LRD with an exponent in the confidence interval of the observed data Hurst parameter estimate.

durations between arrivals. The variance of the marginal distribution is not a robust measure due to infinite variance of the underlying vendor interarrival models, however the heterogeneous interarrival exponent model is the best fit. The Hurst parameter wavelet estimator regression plots for all models are shown in Figure 9. These show that the heterogeneous batch exponent model can not be reliably be used to estimate the H parameter, as there is high variability for the octaves in the wavelet domain. However, more reliable estimates can be made for the interarrival exponent and $x_{\min}$ models. The estimates of the Hurst parameter with the interarrival $x_{\min}$ model are approximately 0.5, with both of the confidence intervals containing the boundary from short range dependence to long range dependence. Hence for these models we can conclude that the heterogeneous $x_{\min}$ model does not exhibit long range dependence in its time series.

The heterogeneous interarrival exponent model exhibits long range dependence with the 95% confidence intervals of the H parameter estimate of the simulated and observed data overlapping. The null model for the heterogeneous interarrival exponent, has a low likelihood for its regression due to the linear fit going outside of the confidence intervals in the wavelet domain, implying that the log-normally distributed parameter selection is a better fit for the observed data. Hence, to model the emergence of LRD the heterogeneous interarrival exponent model with Log-Normal parameter distribution is a suitable model.

After rejecting all models except for the heterogeneous interarrival exponent model, we examine the marginal distribution of the superposed time series, which is shown in Figures 7 and 6. This shows that at approximately 160 disclosures per week, a power-law tail emerges from this distribution. The Kolmogorov-Smirnov statistic for the log-normal fit is 0.042. The emergence of the heavy tail shows that the infinite variance power-law distributions from the renewal processes have resulted in long range dependence in the superposed time series.

In comparison to the LRD models, the SRD interarrival model is unable to generate the same number of vulnerabilities. These models, as expected did not generate a heavy tail with Hurst parameters estimated close to 0.5 and did not produce bursts of arrivals in the same volume as LRD models. Other SRD models, such as Markov Arrival Processes mentioned in Section IV can accurately model the system, however this comes at the cost of a large number of parameters and a less parsimonious model description.

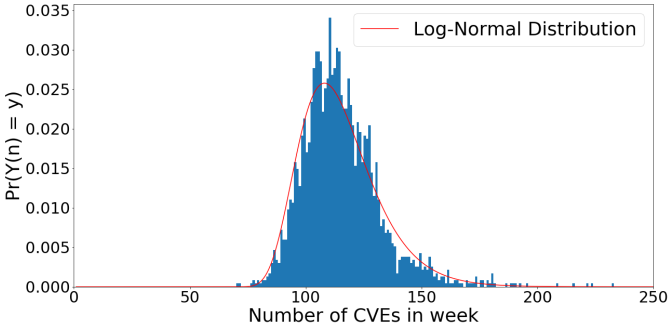


Fig. 6: Marginal distribution of the superposed arrival time series for the heterogeneous interarrival exponent model. A fitted Log-Normal distribution fit is shown, with good agreement to the simulated data. The shape of the resulting marginal distribution is similar to Figure 2, the marginal distribution of the observed data.

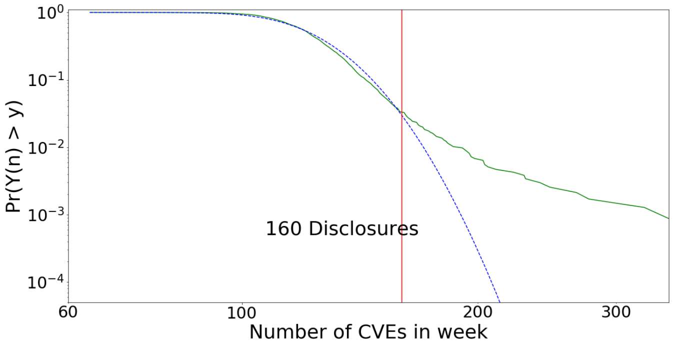


Fig. 7: CCDF for the marginal distribution of the simulated superposed arrival time series on a log-log scale for the heterogeneous interarrival exponent model. A power-law tail emerges at approximately 160. Compare this to Figure 3, with a similar pattern emerging and a reasonable fit to the observed data, on the logarithmic scale of the models.

VI. Queueing Simulation

All disclosed CVEs are analysed by NIST to populate CVSS metrics and release important information about the nature of a vulnerability. When prioritising work to mitigate different vulnerabilities it is important to understand the likely impact due to a vulnerability, and how they can be exploited. For example, an organisation's risk posture may not allow any potential remote exploits, and in the language of CVSS, they would not want to allow any exploit with an *AttackVector* metric value of NETWORK. End users require vulnerabilities to be processed quickly to access information about the risk to a product due to exploit of a vulnerability. However, it was shown in Feutrill et al. [17] that the median time to exploit is less than the median time to process certain classes of vulnerabilities by NIST, as defined by their CVSS metrics.

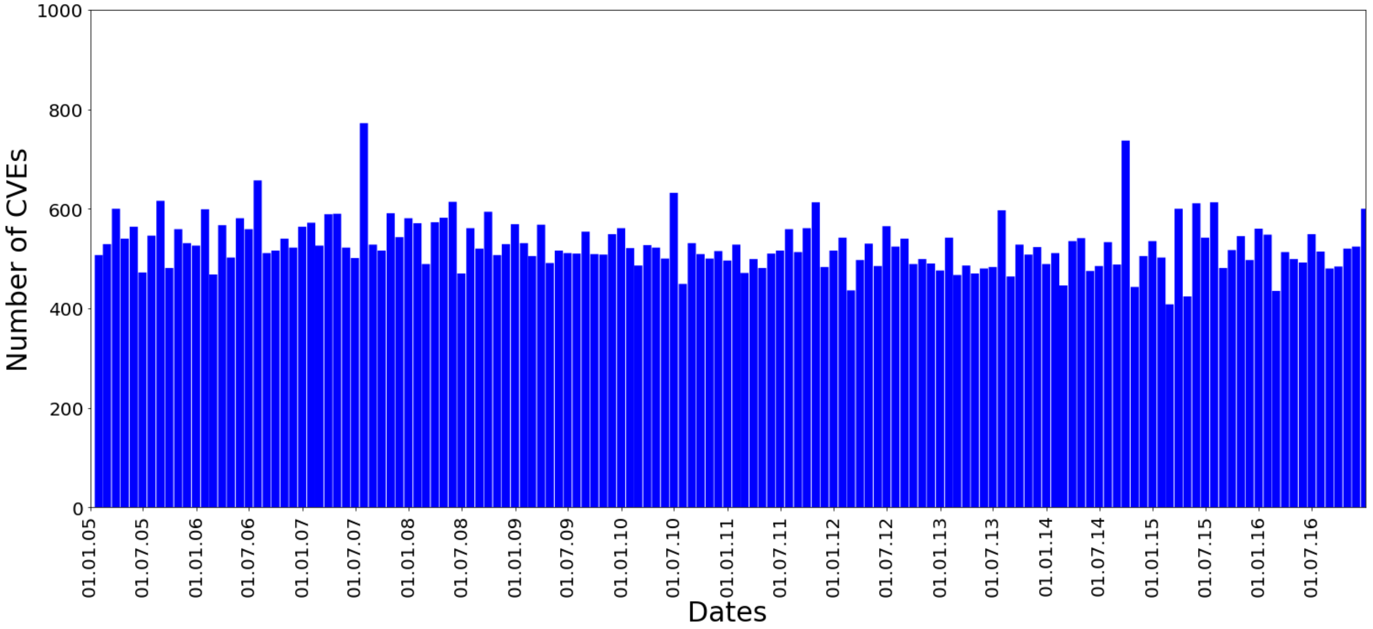


Fig. 8: Simulated time series of the superposed arrival process generated from the heterogeneous interarrival exponent model. This has similar features of LRD and burstiness exhibited by Figure 1, the observed CVE entry time series.

Hence, we would like to be able to find a better strategy to process and disclose CVSS metrics for a vulnerability.

We have built a queueing model to analyse this problem. The arrival process is a superposition of independent but not identical fractal renewal processes with heterogeneous interarrival exponents, as found in Section V. To compare the results of LRD and SRD processes, we have also used a Poisson process model with heterogeneous exponentially distributed interarrival processes and discrete power-law batch sizes. We assume, for simplicity, one server and a deterministic service time, *i.e.*every vulnerability takes the same time to process. In Kendall's notation, this is a $\sum_i GI_i/D/1$ queue. A simulation of this queue was built using the *Powerlaw* library [4] to generate power-law distributed random samples and the *SciPy* [23] package to simulate exponential random variables, with the intention of estimating the probability of a vulnerability being analysed by NIST before the median time to exploit. To ensure we have a stable queue, the arrival rate must be less than the average service rate [9]. From the model definition, the mean number of arrivals in a week is 119.140, so an initial service rate (given a single server model) was assumed of processing 120 vulnerabilities a week, which, in the simulation, fits the observed median of approximately 2 days [17].

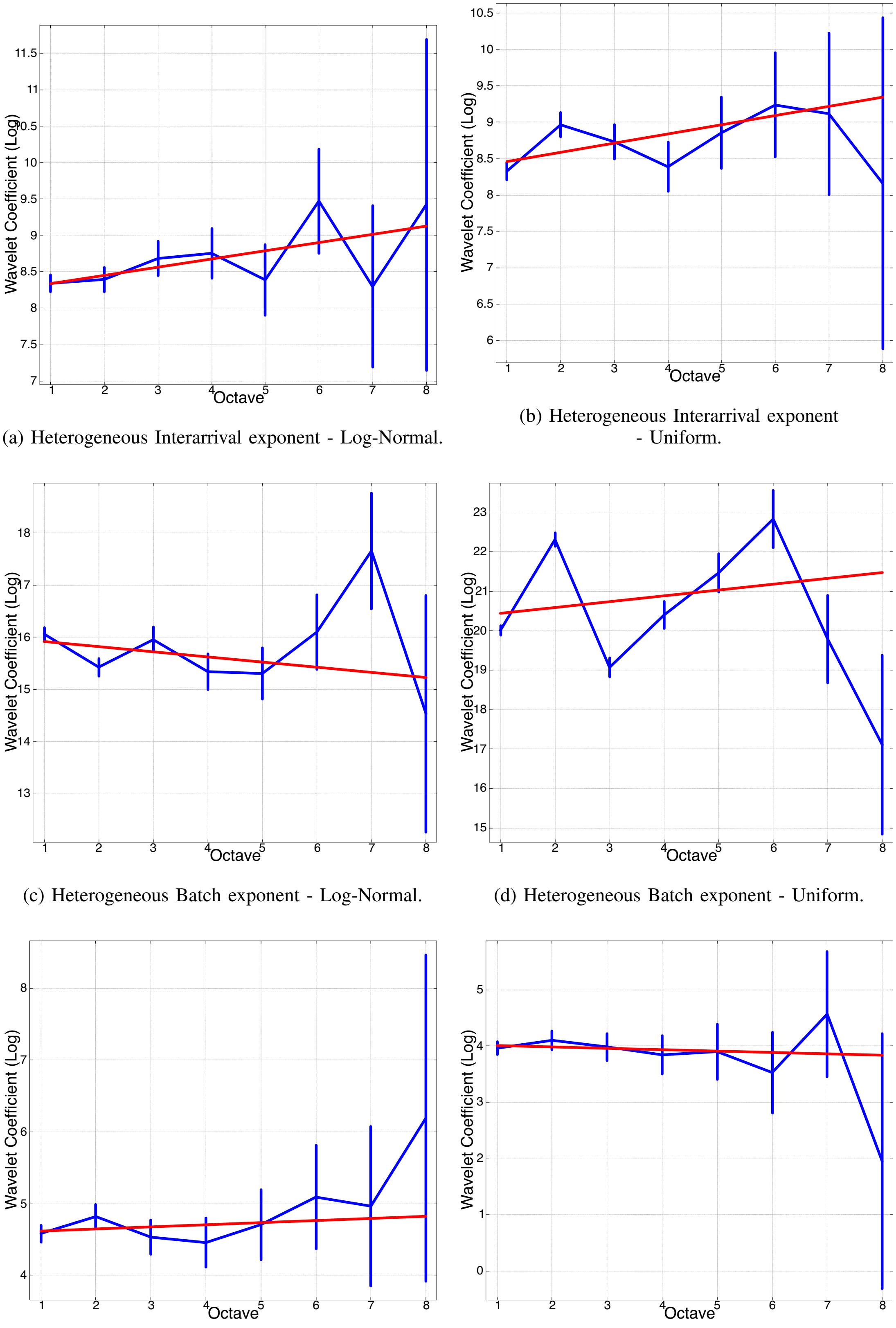
A well established fact from queueing theory is that changing the order of service from First In - First Out to Last In - First Out does not change the mean waiting time, however this can have a large impact on the tail of the distribution [25]. Hence, we evaluate the impact of changing queueing discipline on the median waiting time. This is not a pre-emptive policy, once a job has been commenced from the queue it runs until completion and will not re-enter the queue at any stage.

Strategy	Mean	Median	95th Percentile	Maximum
FIFO	394	109	1689	5459
LIFO	394	1.4	154	315 282
FIFO - SRD	4.0	0.8	12.6	421.5
LIFO -SRD	4.0	0.4	12.1	646.8

TABLE II: Comparison of mean, median, 95th percentile and maximum waiting time (hours) for First In - First Out (FIFO) and Last In - First Out (LIFO) strategies. There is a large decrease in the median time to process and 95th percentile, at the cost of a larger maximum. The improvement was greater for LRD arrivals than SRD arrivals.

The results of the queueing simulation are shown in Table II. The mean waiting time was not effected by the change of queue discipline however the waiting time distribution changed dramatically. The median of the waiting time distribution was reduced from 109 hours to 1.4 hours, and maximum waiting time increasing two orders of magnitude. However, the LIFO queueing discipline had a lower waiting time than the FIFO queueing discipline until the 98th percentile.

We analysed the probability of an individual vulnerability being processed before the mean waiting time, and the median time to exploit, for the entire dataset, since 2010 and since 2017. These results are shown in Table III. Using a LIFO queueing discipline, the probability of being processed before the median exploit time for vulnerabilities disclosed since 2017 increased from 0.58 to 0.96. For the SRD arrival model, as there were less extreme bursts of arrivals the queue was not overwhelmed to the same extent as in the LRD case. The results show that a Poisson process interarrival model still shows improvement from moving from a FIFO to a LIFO queueing discipline, however the effect is less extreme than



(a) Heterogeneous Interarrival exponent - Log-Normal.

(b) Heterogeneous Interarrival exponent - Uniform.

(c) Heterogeneous Batch exponent - Log-Normal.

(d) Heterogeneous Batch exponent - Uniform.

(e) Heterogeneous Interarrival x_{min} - Log-Normal.

(f) Heterogeneous Interarrival x_{min} - Uniform.

Fig. 9: Log-scale diagrams vs. the average of wavelet coefficients at that octave, compare to Figure 5. No reliable estimation of Hurst parameter is possible for the heterogeneous batch model, as shown in (c) and (d), indicating no long range dependence is present in the times series. No long range dependence is generated in the heterogeneous interarrival x_{min} model, with the estimates of the Hurst parameter being around 0.5. The null model from the heterogeneous interarrival exponent model has a couple of octaves with wavelet coefficient values being outside of the linear regression fit, indicating that this is not a good estimator for the Hurst parameter. Hence, the model which best generates the behaviour of interest is the heterogeneous interarrival exponent model with log-normally distributed parameters.

Strategy	Pr(X < E[X])	Pr(X < m) Exploits All time	Pr(X < m) Exploits Since 2010	Pr(X < m) Exploits Since 2017
FIFO	0.70	0.89	0.77	0.58
LIFO	0.97	0.98	0.98	0.96
FIFO - SRD	0.86	0.99	0.99	0.99
LIFO - SRD	0.88	0.99	0.99	0.99

TABLE III: Comparison of probability of being less than mean, E[X], and median, m, time to exploit. The probability of being processed before median time to exploit decreases using a LIFO queue discipline. There is a dramatic performance increase for LRD processes, and a smaller effect for SRD processes.

for LRD models with only an improvement in probability of being processed before mean time to exploit, shifting from 0.86 to 0.88 and a negligible change in the probability of being processed before the median time to exploit.

VII. Discussion

We have discovered that the time series of vulnerability disclosures exhibits long range dependence. This was shown by the heavy tail of the marginal distribution of the arrival process and an estimated Hurst parameter, from the wavelet estimator [3], of 0.66. In addition, we have shown that the different vendors are well modelled by a Zipf distribution for the total number of disclosed vulnerabilities. Hence, the heterogeneity in the vendors is also a key feature of the arrival process and models must incorporate this variation.

We built a stochastic model of the arrival process to accurately model the observed properties. Instead of modelling the appearance of vulnerabilities as being discovered according to a geometrically distributed discrete model [34], [35] or arriving from a pool of sources with exponentially distributed interarrival times [18] we defined an independent fractal renewal process for each vendor with the power-law exponent sampled from a log-normal distribution. The superposition of these renewal processes with non-identical interarrival distributions was able to generate the observed long range dependence and a log-normal marginal distribution with a heavy tail.

The discovery of long range dependence in this process has implications for how we accurately measure and model vulnerability processes. An implication is that strong positive correlations exist in the time series, hence frequent large bursts are expected i.e. there is a higher chance of a large number of vulnerabilities being disclosed the week after a large number of vulnerabilities. For example, in Figure 1, the months of September and October of 2014, a large surge of vulnerabilities occurred including a serious vulnerability named *Shellshock* [31], which was publicly disclosed and with a patch released 10 days later [31]. The availability of vulnerability information is important for security teams to effectively understand the risk posed by particular vulnerabilities and to prioritise work to mitigate these potential threats. If serious vulnerabilities are disclosed within a large burst, there is a risk that crucial information is not released until several days after a threat is being exploited. The long range dependence discovered in the process suggests an underlying complexity in the security ecosystem, which has been shown in other domains such as finance [49], internet traffic [26], [47], [48], biological systems [24], meteorology [44] and hydrology [19].

The heavy tail of the marginal distribution of the arrival process has large effects on the availability of the CVSS metrics. A queueing model was built to model the impacts from the bursts of vulnerabilities, and to test different queueing disciplines to investigate whether the median time to process vulnerabilities could be reduced. The waiting time distribution for vulnerability processing was shown to have a heavy tail due to the heavy tail distribution of the superposition of the vendor arrival processes. However, by changing the order in which the vulnerabilities are processed we were able to show that improvements could be made by serving the most recently arrived vulnerability first and reduce the median waiting time from 2 days to less than 2 hours. This comes at the cost of the maximum time for a vulnerability being increased, and higher processing time from 98th percentile and above. In contrast to this result, there was a much smaller performance increase for short range dependent processes by using a LIFO discipline, indicating that this strategy has the best performance to handle large bursts in arrivals, such as from the long range dependent models. We believe this strategy would be able to increase the availability of security information by processing most vulnerabilities sooner. Many vulnerability prediction models have been built that utilise the CVSS information [8], [10], [13], [14] and the utility of these models is reduced by not having relevant security information available at the time of prediction.

VIII. Conclusion

This paper shows that the time series of public vulnerability disclosures has the statistical property of long range dependence. This is characterised by behaviour such as burstiness, high variation and slow convergence of estimators, making traditional analysis techniques difficult. We have shown that this time series can be well modelled by a superposition of independent but not identically distributed fractal renewal processes. Due to the long range dependence, delays for processing vulnerabilities by NIST dramatically increase during bursty periods. We have shown that changing the order of service from FIFO to LIFO reduces median waiting time, with the probability of a vulnerability being processed before the median exploit time since 2017 increasing from 0.58 to 0.96.

Acknowledgment

This project was supported by Data61, the Australian Research Council Centre of Excellence in Mathematical and Statistical Frontiers (ACEMS) and Data to Decisions CRC.

References

[1] Code for the estimation of scaling exponents - Darryl Veitch, University of Technology Sydney. [Online]. Available: http://crin.eng.uts.edu.au/~darryl/secondorder_code.html.

[2] Patch Tuesday... Exploit Wednesday - TrendLabs Security Intelligence Blog. [Online]. Available: https://blog.trendmicro.com/trendlabs-security-intelligence/patch-tuesday-exploit-wednesday, Oct. 2006.

[3] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2–15, Jan 1998.

[4] J. Alstott, E. Bullmore, and D. Plenz. Powerlaw: a Python package for analysis of heavy-tailed distributions. *PLoS one*, 9(1):e85777, Jan. 2014. arXiv: 1305.0215.

[5] M. A. Arfeen, K. Pawlikowski, A. Willig, and D. McNickle. Fractal renewal process based analysis of emerging network traffic in access networks. In *2016 26th International Telecommunication Networks and Applications Conference (ITNAC)*, pages 265–270, Dec. 2016.

[6] J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.

[7] J. M. Berger and B. Mandelbrot. A new model for error clustering in telephone circuits. *IBM Journal of Research and Development*, 7(3):224–236, July 1963.

[8] M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker. Beyond heuristics: learning to classify vulnerabilities and predict exploits. In *Proceedings of the 16th ACM international conference on Knowledge discovery and data mining*, pages 105–114. ACM, 2010.

[9] M. Bramson. *Stability of queueing networks*. Springer, 2008.

[10] B. L. Bullough, A. K. Yanchenko, C. L. Smith, and J. R. Zipkin. Predicting exploitation of disclosed software vulnerabilities using open-source data. In *Proceedings of the 3rd ACM on International Workshop on Security And Privacy Analytics*, IWSPA '17, pages 45–53, New York, NY, USA, 2017. ACM.

[11] A. B. Downey. Evidence for long-tailed distributions in the internet. In *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, IMW '01, pages 229–241, New York, NY, USA, 2001. ACM.

[12] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, New York, NY, USA, 4th edition, 2010.

[13] M. Edkrantz. Predicting Exploit Likelihood for Cyber Vulnerabilities with Machine Learning. Master's thesis, 2015.

[14] M. Edkrantz and A. Said. Predicting cyber vulnerability exploits with machine learning. In *SCAI*, 2015.

[15] A. K. Erlang. The theory of probabilities and telephone conversations. *Nyt. Tidsskr. Mat. Ser. B*, 20:33–39, 1909.

[16] A. Erramilli, O. Narayan, and W. Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking*, 4(2):209–223, April 1996.

[17] A. Feutrill, D. Ranathunga, Y. Yarom, and M. Roughan. The effect of common vulnerability scoring system metrics on vulnerability exploit delay. In *2018 Sixth International Symposium on Computing and Networking (CANDAR)*, pages 1–10, Nov 2018.

[18] K. Haldar and B. K. Mishra. Mathematical model on vulnerability characterization and its impact on network epidemics. *International Journal of System Assurance Engineering and Management*, 8(2):378–392, June 2017.

[19] H. E. Hurst. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–799, 1951.

[20] H. Joh, J. Kim, and Y. K. Malaiya. Vulnerability discovery modeling using weibull distribution. In *2008 19th International Symposium on Software Reliability Engineering (ISSRE)*, pages 299–300, Nov 2008.

[21] H. Joh and Y. K. Malaiya. Periodicity in software vulnerability discovery, patching and exploitation. *Int. J. Inf. Sec.*, 16(6):673–690, 2017.

[22] P. Johnson, D. Gorton, R. Lagerström, and M. Ekstedt. Time between vulnerability disclosures: A measure of software product vulnerability. *Computers & Security*, 62:278–295, 2016.

[23] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[24] Karmeshu and A. Krishnamachari. Sequence Variability and Long-Range Dependence in DNA: An Information Theoretic Perspective. In *Neural Information Processing*, volume 3316, pages 1354–1361. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[25] J. F. C. Kingman. The effect of queue discipline on waiting time variance. *Mathematical Proceedings of the Cambridge Philosophical Society*, 58(1):163–164, 1962.

[26] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic. In *Conference Proceedings on Communications Architectures, Protocols and Applications*, SIGCOMM '93, pages 183–193, New York, NY, USA, 1993. ACM.

[27] J. B. Levy and M. S. Taqqu. Renewal reward processes with heavy-tailed inter-renewal times and heavy-tailed rewards. *Bernoulli*, 6(1):23–44, 02 2000.

[28] S. B. Lowen and M. C. Teich. Fractal renewal processes. *IEEE Transactions on Information Theory*, 39(5):1669–1671, Sept. 1993.

[29] S. B. Lowen and M. C. Teich. *Fractal-Based Point Processes*. Wiley-Interscience, New York, NY, USA, 2005.

[30] P. Mell and T. Grance. Use of the common vulnerabilities and exposures (cve) vulnerability naming scheme. Technical report, National Institute of Standards and Technology, Gaithersburg, MD, Computer Security Div, 2002.

[31] National Institute of Standards and Technology. NVD - CVE-2014-6271. [Online]. Available: https://nvd.nist.gov/vuln/detail/CVE-2014-6271, 2014.

[32] I. Norros. A storage model with self-similar input. *Queueing Systems*, 16(3):387–396, Sep 1994.

[33] N. R. Pokhrel, H. Rodrigo, and C. P. Tsokos. Cybersecurity: Time Series Predictive Modeling of Vulnerabilities of Desktop Operating System Using Linear and Non-Linear Approach. *Journal of Information Security*, 8(4):362–382, Oct. 2017.

[34] S. M. Rajasooriya, C. P. Tsokos, and P. K. Kaluarachchi. Stochastic Modelling of Vulnerability Life Cycle and Security Risk Evaluation. *Journal of Information Security*, 07(04):269–279, 2016.

[35] S. M. Rajasooriya, C. P. Tsokos, and P. K. Kaluarachchi. Cyber security: Nonlinear stochastic models for predicting the exploitability. *Journal of Information Security*, 08(02):125–140, 2017.

[36] M. Roughan, D. Veitch, and M. Rumsewicz. Computing queue-length distributions for power-law queues. In *Proceedings. IEEE INFOCOM '98, the Conference on Computer Communications. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Gateway to the 21st Century (Cat. No.98*, volume 1, pages 356–363 vol.1, March 1998.

[37] Y. Roumani, J. K. Nwankpa, and Y. F. Roumani. Time series modeling of vulnerabilities. *Computers & Security*, 51:32–40, 2015.

[38] B. Ryu and S. Lowen. Fractal traffic models for Internet simulation. In *Proceedings of Fifth IEEE Symposium on Computers and Communications*, pages 200–206, 2000.

[39] G. Samorodnitsky. Long Range Dependence. *Foundations and Trends® in Stochastic Systems*, 1(3):163–257, 2006.

[40] K. Sriram and W. Whitt. Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE Journal on Selected Areas in Communications*, 4(6):833–846, Sep. 1986.

[41] M. Tang, M. Alazab, and Y. Luo. Exploiting vulnerability disclosures: Statistical framework and case study. In *2016 Cybersecurity and Cyberforensics Conference (CCC)*, pages 117–122, Aug 2016.

[42] M. Tang, M. Alazab, and Y. Luo. Big data for cybersecurity: Vulnerability disclosure trends and dependencies. *IEEE Trans. Big Data*, 5(3):317–329, 2019.

[43] M. Tang, M. Alazab, Y. Luo, and M. Donlon. Disclosure of cyber security vulnerabilities: time series modelling. *IJESDF*, 10(3):255–275, 2018.

[44] M. S. Taqqu. Random processes with long-range dependence and high variability. *Journal of Geophysical Research: Atmospheres*, 92(D8):9683–9686, 1987.

[45] M. S. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar traffic modeling. *SIGCOMM Comput. Commun. Rev.*, 27(2):5–23, Apr. 1997.

[46] The MITRE Corporation. Common vulnerabilities and exposures (CVE). [Online]. Available: http://cve.mitre.org/, 2018.

[47] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson. Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Statist. Sci.*, 10(1):67–85, 02 1995.

[48] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: statistical analysis of ethernet lan traffic at the source level. *IEEE/ACM Transactions on Networking*, 5(1):71–86, Feb 1997.

[49] W. Willinger, M. S. Taqqu, and V. Teverovsky. Stock market prices and long-range dependence. *Finance and Stochastics*, 3(1):1–13, Jan. 1999.