Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

Modeling and predicting extreme cyber attack rates via marked point processes

Chen Peng, Maochao Xu, Shouhuai Xu & Taizhong Hu

To cite this article: Chen Peng, Maochao Xu, Shouhuai Xu & Taizhong Hu (2016): Modeling and predicting extreme cyber attack rates via marked point processes, Journal of Applied Statistics, DOI: 10.1080/02664763.2016.1257590

To link to this article: http://dx.doi.org/10.1080/02664763.2016.1257590

Published online: 17 Nov 2016.

JOURNAL OF APPLIED STATISTICS, 2016
http://dx.doi.org/10.1080/02664763.2016.1257590

Taylor & Francis
Taylor & Francis Group

Modeling and predicting extreme cyber attack rates via marked point processes

Chen Peng[a], Maochao Xu[b], Shouhuai Xu[c] and Taizhong Hu[a]

[a]Department of Statistics and Finance, University of Science and Technology of China, Anhui, People's Republic of China; [b]Department of Mathematics, Illinois State University, Normal, IL, USA; [c]Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

ABSTRACT

Cyber attacks have become a problem that is threatening the economy, human privacy, and even national security. Before we can adequately address the problem, we need to have a crystal clear understanding about cyber attacks from various perspectives. This is a challenge because the Internet is a large-scale complex system with humans in the loop. In this paper, we investigate a particular perspective of the problem, namely the extreme value phenomenon that is exhibited by cyber attack rates, which are the numbers of attacks against a system of interest per time unit. It is important to explore this perspective because understanding the statistical properties of extreme cyber attack rates will pave the way for cost-effective, if not optimal, allocation of resources in real-life cyber defense operations. Specifically, we propose modeling and predicting extreme cyber attack rates via *marked point processes*, while using the Value-at-Risk as a natural measure of intense cyber attacks. The point processes are then applied to analyze some real data sets. Our analysis shows that the point processes can describe and predict extreme cyber attack rates at a very satisfactory accuracy.

ARTICLE HISTORY
Received 4 January 2016
Accepted 31 October 2016

KEYWORDS
Autoregressive conditional duration; extreme value theory; generalized Pareto distribution; intensity

1. Introduction

Analyzing cyber attack data, or cybersecurity data analytics, is an important field in cybersecurity research. In order to take full advantage of cyber attack data, we need to characterize the statistical properties that are exhibited by the data and investigate their cybersecurity implications. In this paper, we investigate the *extreme value* phenomenon that is exhibited by the time series of cyber attack rates, which are the numbers of cyber attacks against some targets per time unit. The term 'target' refers to any cyber system of interest (e.g. a set of computers or Internet Protocol (IP) addresses, or a set of services offered by some computers), the term 'time unit' refers to any resolution of interest (e.g. minute, hour or day), and the term 'attack rate' during a time unit reflects the intensity of cyber attacks against the target in question.

The importance of studying the extreme value phenomenon can be justified as follows. In order to effectively defend a target against cyber attacks, the defender needs to allocate

CONTACT Maochao Xu mxu2@ilstu.edu

adequate defense resources. However, the defender cannot always overprovision defense resources because they are too expensive. In order to see this, we mention that defense mechanisms, such as Deep Packet Inspection [17] (which, roughly speaking, examines the individual IP packets that come to, and possibly from, the target so as to block cyber attacks), demand substantial computer resources. When the attack (or packet) rates are above the capacity that can be processed by the allocated defense resource, the defense may be forced to leave some attack packets unexamined and therefore unblocked (because it is not acceptable to simply drop the unexamined packets). Putting another way, the restriction is that the defender cannot constantly deploy an unnecessarily large amount of defense resources, because attack rates are low for most of the time. Therefore, it is important to enable the defender to dynamically allocate extra defense resources *on demand* to accommodate the abnormally large cyber attack rates, when the need arises. To make this possible, we must be able to accurately predict the extreme cyber attack rates ahead of time, the longer the period ahead of time the better. This also justifies the importance of predicting extreme cyber attack rates.

In this paper, we make the following contributions. We propose a novel application of *marked point processes* to fit and predict extreme cyber attack rates, while using the Value-at-Risk (VaR) as a natural measure of intense attacks. In particular, we use the point-over-threshold method to model the magnitudes of extreme attack rates, and use the Autoregressive Conditional Duration approach to describe the arrival of extreme attack rates. The approach is featured by its capability of simultaneously accommodating the magnitudes of extreme values (i.e. extreme attack rates in the context of the present paper), the inter-exceedance times between extreme values, and the dependence between the inter-arrival times of extreme values. Our empirical analysis based on two real-world data sets, which were collected by two widely used cyber instruments known as *network telescope* (or *telescope* for short) and *honeypot*, shows that the approach can accurately describe and predict extreme cyber attack rates. The approach is interesting on its own from a theoretical perspective, and useful in practice because it enables the defender to dynamically allocate defense resources based on the predicted extreme attack rates.

The kinds of data, which we analyze, have been studied in the literature. However, prior studies are for different purposes and use different modeling techniques. On one hand, *telescope* data has been studied for the purpose of understanding Internet background radiation and denial-of-service activities (e.g. [34]), for the purpose of characterizing the behavior of activities such as scanning, peer-to-peer applications, unreachable services, misconfigurations, worms, or one-way traffic (e.g. [14]). On the other hand, *honeypot* data has been analyzed for the purpose of characterizing the traffic of certain kinds of attacks (e.g. known vs. unknown attacks, [2,39]), for the purpose of analyzing probe and scan activities (e.g. [30]), for the purpose of understanding denial-of-service attacks [22] or worm/botnet activities (e.g. [19,31]), and for the purpose of characterizing the statistical properties exhibited by the data (e.g. Long-Range Dependence, [43]). Recently, the extreme value phenomenon exhibited by cyber attack data has been touched in [44] by using a time series approach. Specifically, they used the Fractional AutoRegressive Integrated Moving Average (FARIMA) model to capture the mean part of attack rates, while using the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model to accommodate the high volatilities of attack rates. Although this approach can offer accurate predictions for some cyber attack data, the GARCH model, in general, is not

known to be able to provide a solid theoretical explanation for the clustering behavior of extreme values [33], which is however widely observed in the cybersecurity domain (cf. [43,44]).

The rest of the paper is organized as follows. In Section 2, we describe the data and an exploratory analysis of the data. In Section 3, we introduce the marked point process model and discuss the VaR measure of extreme cyber attack rates. In Section 4, we report a simulation study to examine the performance of the proposed models. In Section 5, we present the results of using marked point processes to model and predict extreme cyber attack rates. We conclude the paper and discuss future work in Section 6.

2. Data and exploratory analysis

The data sets we analyze were collected by the two cyber instruments mentioned above, namely *network telescope* and *honeypot*. In this section we briefly describe these instruments as well as the data sets, and then report on our exploratory analysis of the data sets. For the time series of attack rates, the time unit we use is hour, which is reasonable because 1-hour ahead predictions would be sufficient for the defender to dynamically allocate (deallocate) defense resources to cope with the predicted large (small) attack rates during the next hour. While other time units could be used as well, it would waste defense resources if the time unit is too coarse-grained (e.g. 48 h) because large attack rates may occur occasionally (e.g. during 3 h of the 48 h span). Nevertheless, how to determine the optimal time unit is an interesting problem for future research.

2.1. Cyber instruments and data sets

Network telescope and data set. A *network telescope*, or *telescope*, passively monitors a large chunk of IP addresses, which have no Internet services but are exclusively set up to collect unsolicited network traffic. In other words, a telescope does not interact with any incoming Internet service request. As a result, when a remote attacker tries to connect to a web server with an IP address that belongs to a telescope (e.g. for the purpose of figuring out whether or not there is indeed a server, and if so, further attack steps being taken), the telescope will not respond to the request but simply records the requests. This also explains why telescopes are sometimes called *darknets*, *sinks*, or *blackholes* (e.g. [4,42]). Because there are no Internet services that are associated to those IP addresses, the network traffic coming to a telescope is unsolicited, and is most likely caused by cyber attacks (e.g. computer malwares searching for vulnerable computers in cyberspace). Still, the data collected by a telescope could contain some *non-malicious traffic*, which may be caused by system misconfigurations and background radiations (e.g. [23,41]). There is a widely used preprocessing procedure for filtering such non-malicious traffic. At a high level, the procedure works as follows. First, the collected IP packets (i.e. raw data) are reassembled into *flows* according to a standard procedure described in [12], where the term 'flow' is well-defined in field of computer networks. Each flow represents an attack. Second, the assembled flows are then classified into three categories: *backscatter* flows, *ICMP request* flows, and *other* flows. Since (i) analysis of backscatter flows has been conducted elsewhere (e.g. [25,34]), and (ii) ICMP has been mainly used to launch DOS attacks (e.g. [29,34,40]), we focus on analyzing the other kinds of cyber attacks (i.e. the *other* flows). By counting the number of

attacks that arrive at a telescope per time unit (e.g. hour), we obtain a time series of cyber attack rates.

The telescope data set was collected between 1 March 2013 and 31 March 2013 by CAIDA's network telescope, which monitors a globally routeable /8 network (i.e. approximately 0.4% of Internet's IP version 4 address space, or 2^{24} IP addresses). Because (i) the size of a network telescope is specified by the chunk of IP addresses it monitors, and (ii) we want to see whether or not a difference in telescope size can incur a significant change in statistical properties exhibited by the data, we consider three variant telescopes of CAIDA's. Specifically, we analyze three data sets that correspond to three variant telescopes of sizes $S_1 = 2^{24} - 2^{20}$, $S_2 = 2^{24} - 2^{18}$, and $S_3 = 2^{24} - 2^{16}$. We will refer to both these variant telescopes and the data sets S_1, S_2, and S_3, respectively. This leads to three telescope data sets of time series of cyber attack rates.

Honeypot and data set. Similar to a network telescope, a honeypot is exclusively set up for collecting unsolicited Internet traffic, because legitimate users should not make any request to any IP address within the honeypot. Unlike a network telescope, the IP addresses monitored by a honeypot are indeed associated with some Internet services (e.g. web server). However, these services are only partly 'emulated' to accommodate the remote requests somewhat (for economic reasons). This design allows the defender to use a small number of computers to monitor a certain number of IP addresses, which explains why a honeypot is often much smaller than a network telescope. Although the network traffic coming to a honeypot is most likely cyber attacks, we still need to filter the non-malicious traffic.

Similar to the pre-processing procedure for telescope data, there is a standard procedure for pre-processing honeypot data [1] to extract the time series of cyber attack rates (i.e. the numbers of cyber attacks against the honeypot per time unit).

The honeypot data set we analyze corresponds to what was called 'Period V' in [44], which however cannot be accurately predicted by the models studied there. The data set was collected between 22 June 2011 and 27 August 2011 (i.e. 67 days or 1608 hours) by a honeypot monitoring 166 IP addresses via four popular honeypot programs: Amun, Dionaea, Mwcollector and Nepenthes. In the honeypot, a computer ran multiple honeypot programs, each of which was associated to one of the 166 IP addresses. Since the honeypot is small (i.e. 166 IP addresses), we do not consider variants of it. The resulting data set is the time series of attack rates as observed by the honeypot.

2.2. Exploratory analysis

Figure 1 plots the time series of cyber attack rates and the time series of the extreme cyber attack rates corresponding to the aforementioned variant telescope data sets S_1, S_2 and S_3 and the honeypot data set (time unit: hour). We observe that the telescope time series share some similarity, but are quite different from the honeypot one. Nevertheless, they all exhibit many large attack rates (i.e. extreme values), which represent intense attacks. In order to characterize the extreme values, we need to define some *thresholds* such that attack rates exceeding the thresholds are deemed as extreme attack rates. For the three variant telescopes, we set the threshold to be the 88% quantile of the attack rates during the lifetime of observation; for the honeypot time series, we set the threshold to be the 90% quantile of the attack rates during the lifetime of observation. These thresholds are selected based on the following criteria. From a practical point of view, network defenders

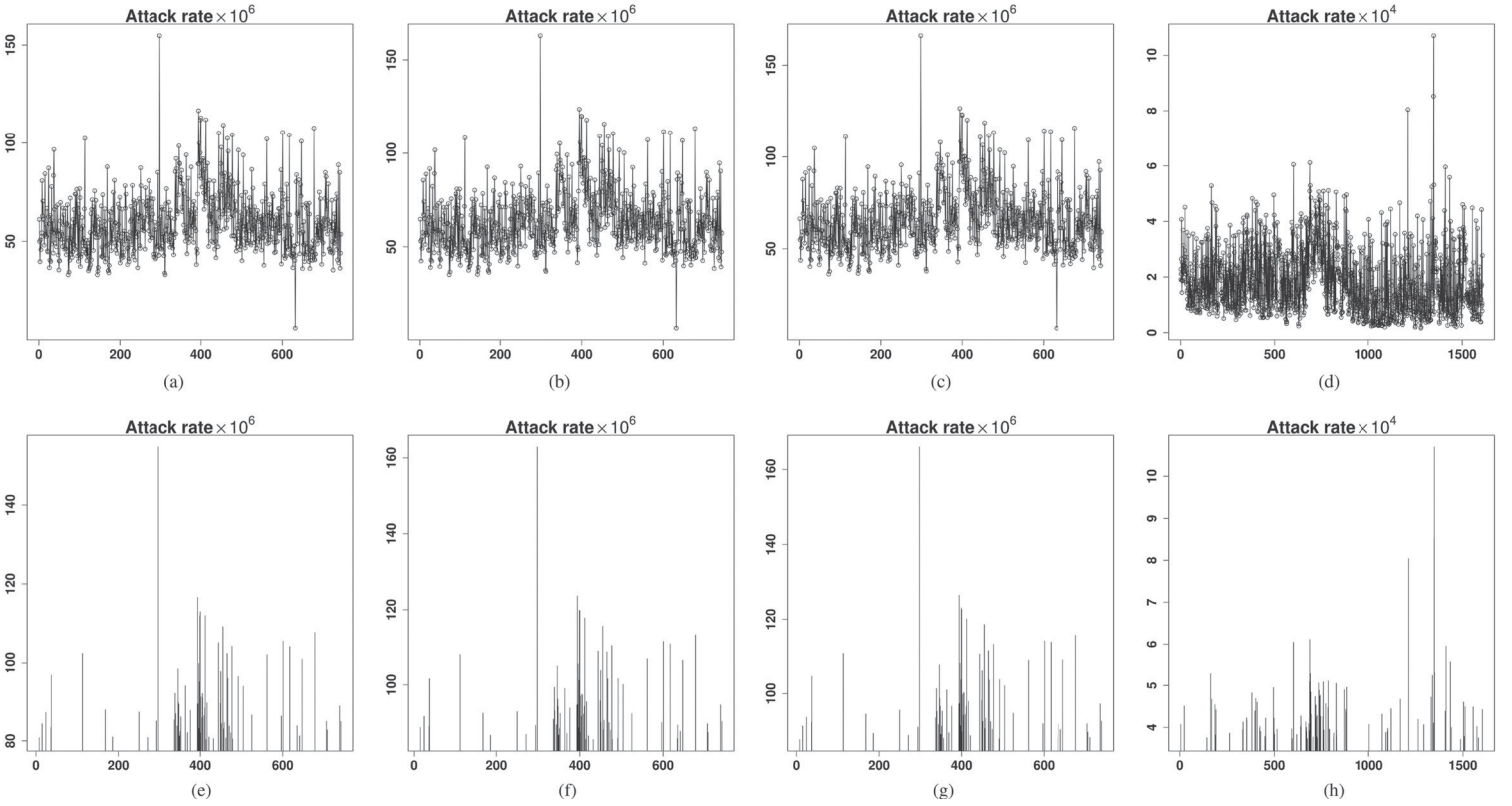

Figure 1. Plots of times series of cyber attack rates (top row) and time series of extreme cyber attack rates (bottom row), where the *x*-axis represents time (unit: hour) and the *y*-axis indicates the attack rate (i.e. the hourly number of attacks observed by the cyber instrument in question). Extreme attack rates are those which are above certain thresholds: the 88% quantile for the telescope data sets and the 90% quantile for the honeypot data set. (a) Attack rates: telescope data set S_1, (b) attack rates: telescope data set S_2, (c) attack rates: telescope data set S_3, (d) attack rates: honeypot data set, (e) extreme attack rates: telescope data set S_1, (f) extreme attack rates: telescope data set S_2, (g) extreme attack rates: telescope data set S_3 and (h) extreme attack rates: honeypot data set.

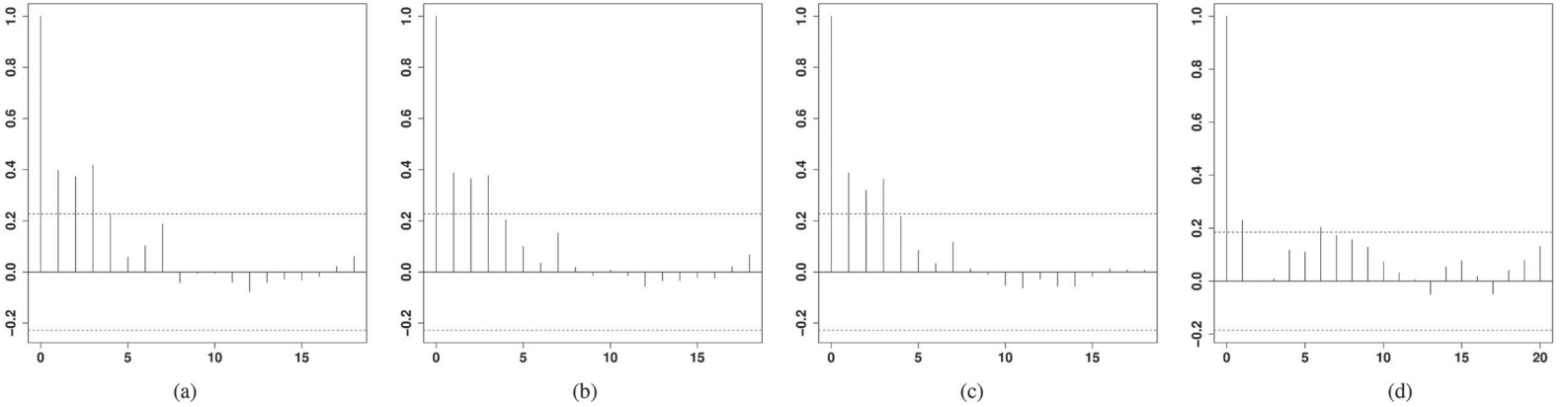

Figure 2. ACF plots of inter-exceedance times (i.e. time intervals between extreme values). (a) Telescope data set S_1, (b) telescope data set S_2, (c) telescope data set S_3 and (d) honeypot data set.

Table 1. Results of the KS test for the exponential distribution of inter-exceedances times: for a data set, D is the value of the test statistic, and p is the corresponding p-value.

	Telescope data set S_1	Telescope data set S_2	Telescope data set S_3	Honeypot data set
D	.2631	.2886	.2496	.3002
p	7.1e−05	8.898e−06	1.98e−04	3.437e−09

are concerned with high quantiles of attack rates that exceed what can be processed by the defense resource. From a theoretical point of view, statistical models require proper selection of threshold values, because too few data points (corresponding to large thresholds) may result in a large bias, while too many data points (corresponding to small thresholds) may lead to difficulties in finding a proper distribution. It is shown in the Appendix that the selected thresholds are robust in terms of both the ground process and the mark distribution. Figure 1(e)–(h) plot the resulting time series of extreme cyber attack rates, which exhibit the cluster phenomenon (i.e. intense attacks sustain for a period of time).

Recall that the literature often assumed that occurrences of extreme values follow the Poisson process, meaning that the time intervals between extreme values, called *inter-exceedance times*, are independent of each other and follow the exponential distribution. We want to know whether or not this hypothesis holds for our data sets. Figure 2 plots the autocorrelations of the time intervals between consecutive extreme attack rates. We observe significant correlations between the inter-exceedance times in all of the cases. This means that the occurrences of extreme attack rates are not independent of each other. Table 1 further describes the one-sample Kolmogorov–Smirnov (KS) test that is performed on the inter-exceedances. We observe that the p-values are small, meaning that the exponential distribution hypothesis for the inter-exceedances is rejected for all of the time series. Hence, the Poisson process may not be suitable for modeling the extreme cyber attack rates.

The need for new models. The above discussion motivates us to characterize the extent at which the present extreme attack rates indicate the future extreme attack rates and how we can predict the future extreme attack rates. For these purposes, one may suggest to use the Extreme Value Theory (EVT). However, the classical point process based on EVT (e.g. the point-over-threshold or POT method, [20,37]) emphasizes on modeling the magnitudes of exceedances, but *without* considering the dependence between the inter-exceedance times. The classical EVT models are not appropriate for our purposes because we already observed the correlation between the inter-exceedance times. In order to incorporate the inter-exceedance times into the POT model, we propose using *marked point processes* to accommodate both the arrivals of extreme attack rates and the magnitudes of their exceedances. Specifically, we use the POT method to model the magnitudes of their exceedances, and use the Autoregressive Conditional Duration (ACD) model to describe the arrivals of extreme attack rates because ACD can accommodate the slow decay of autocorrelation as well as bursts of extreme value clusters (which are exhibited by the data sets as shown in Figures 1 and 2).

3. Marked point process and dynamic cyber risk measure

3.1. Classical EVT

Suppose $X_0, \ldots, X_n$ are the hourly cyber attack rates against a target. Let $u > 0$ be a threshold. Then, any $X_i > u$ is called an extreme value. The extreme values formulate a point

process

$$N(A) = \sum_{i=1}^{n} \mathrm{I}((t_i, X_i) \in A)$$

over state space $\chi = (0, t_n] \times (u, \infty)$. According to the classical EVT (cf. [32,37]), the point process can be treated as a non-homogeneous poisson process with the following intensity function:

$$\lambda(t, x) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi-1}, \tag{1}$$

where $(x)_+ = \max\{x, 0\}$, $\sigma > 0$, and μ and ξ are, respectively, the scale, location and shape parameters. For a set $A = (t_1, t_2) \times (y, \infty)$, the intensity measure is

$$\Lambda(A) = \int_{t_1}^{t_2}\int_{y}^{\infty} \lambda(t, x)\,\mathrm{d}x\,\mathrm{d}t = -(t_2 - t_1)\ln H_{\xi,\mu,\sigma}(x),$$

where

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\{-(1 + \xi(x-\mu)/\sigma)\}, & \xi \neq 0 \\ \exp\{-\exp(-(x-\mu)/\sigma)\}, & \xi = 0 \end{cases}$$

is the generalized extreme value distribution. The intensity function (1) can be rewritten as

$$\lambda(t, x) = \lambda_g(t) f(x),$$

where $\lambda_g(t) = -\ln H_{\xi,\mu,\sigma}(u)$ is the rate of the one-dimensional Poisson process of the exceedances of level u, and

$$f(x) = \frac{1}{\beta}\left(1 + \xi\frac{x-\mu}{\beta}\right)_+^{-1/\xi-1}$$

is the density function of the generalized Pareto distribution (GPD) with $\beta = \sigma + \xi(u - \mu) > 0$ being the redefined scale parameter.

Traditionally, a *marked point process* treats the occurrences of extreme attack rates as a Poisson process, with the distribution of marks or extreme values $\tilde{x} = x - u$ (provided that $x > u$) being GPD. However, our exploratory analysis of the inter-exceedance time in Section 2 already showed that Poisson process is not adequate for describing the data. We need new models that can accommodate the phenomena exhibited by the data, including the clustering behavior of extreme values and the correlation of inter-exceedance times.

3.2. Marked point process

Denote by $\{(t, \tilde{x}_t)\}$ the occurrence times and marks (i.e. extreme values) over state space χ, with history

$$\mathfrak{F}_t = (\{t_1, \tilde{x}_1\}, \ldots, \{t_{N(t)}, \tilde{x}_{N(t)}\}),$$

where $N(t)$ represents the last extreme value event before time t. In order to model the historical effect, we let the ground point process depend on the history $\mathfrak{F}_t$, which means that

the conditional intensity function takes the form $\lambda_g(\cdot|\mathfrak{F}_t)$. Suppose the marks' distribution also depends on the history. Then, the conditional intensity of the marked point process can be represented as

$$\lambda(t, \tilde{x}_t|\mathfrak{F}_t) = \lambda_g(t|\mathfrak{F}_t) f(\tilde{x}_t|\mathfrak{F}_t). \tag{2}$$

Various forms of $\lambda_g(t|\mathfrak{F}_t)$ and $f(\tilde{x}_t|\mathfrak{F}_t)$ have been discussed in the literature. For example, the classical Hawkes model has been used for modeling stock market price (e.g. [7,8]); the epidemic-type aftershock sequence (ETAS) model has been used for describing earthquake data (e.g. [28,35]). Specifically, we have

$$\lambda_g(t|\mathfrak{F}_t) = k + \phi \sum_{i:t_i<t} g(t - t_i, \tilde{x}_i),$$

and

$$f(\tilde{x}_t|\mathfrak{F}_t) = \frac{1}{\beta(t, \tilde{x}_t|\mathfrak{F}_t)} \left(1 + \xi \frac{\tilde{x}_t - \mu}{\beta(t, \tilde{x}_t|\mathfrak{F}_t)}\right)^{-1/\xi - 1},$$

where $\beta(t, \tilde{x}_t|\mathfrak{F}_t) = \beta_0 + \eta \sum_{i:t_i<t} g(t - t_i, \tilde{x}_i)$. Note that k and β_0 represent the background rate of events, and $\phi, \eta \geq 0$ represent the magnitudes of self-excitation. The kernel function $g(\cdot, \cdot)$ represents the density at which the self-excitation is triggered. For the Hawkes model, the kernel function is

$$g_h(t - t_i, \tilde{x}_i) = (1 + \delta \tilde{x}_i) \exp(-\gamma(t - t_i)); \tag{3}$$

and for the ETAS model, the kernel function is

$$g_e(t - t_i, \tilde{x}_i) = (1 + \delta \tilde{x}_i) / \left(1 + \frac{t - t_i}{\gamma}\right)^{1+\rho}, \tag{4}$$

where $t - t_i$ represents the time elapsed since the extreme event that occurred at time t. It is seen that for the Hawkes model, the decay function of the ground process is exponential; for the ETAS model, the decay function is hyperbolic.

Because we already observed correlations between the inter-exceedance times, we propose using the ACD model for the ground point process, where the ACD model [21,24] accommodates the correlated arrivals of extreme values. The basic idea is to standardize the inter-exceedance times $\Delta t_i = t_i - t_{i-1}$, where $i = 1, 2, \ldots, n$, by using the history information. Specifically, we define

$$\Delta t_i = \Psi_i \varepsilon_i,$$

where the ε_i's are independent and identically distributed innovations with $\mathrm{E}(\epsilon_i) = 1$ while satisfying

$$\mathrm{E}(\Delta t_i|\mathfrak{F}_\mathrm{i}) = \Psi_i,$$

and the Ψ_i's are functions of the past durations and conditional durations.

We focus on two ACD models:

(a) The standard ACD model (ACD)

$$\Psi_i = \omega + \sum_{j=1}^{p} a_j \Delta t_{i-j} + \sum_{j=1}^{q} b_j \Psi_{i-j},$$

where $\omega, a_j, b_j \geq 0$, and p and q are positive integers indicating the orders of the autoregressive terms.

(b) The Log-ACD model (Log-ACD [5])

$$\Psi_i = \omega + \sum_{j=1}^{p} a_j \log(\Delta t_{i-j}) + \sum_{j=1}^{q} b_j \log(\Psi_{i-j}).$$

In this paper, we restrict our investigation to the case of $p = q = 1$, because a higher order does not necessarily improve the prediction accuracy [6]. For comparison purposes, we also examine the performance of the base model with $a_1 = b_1 = 0$. The distribution of the standardized innovations of the ε_i's is specified as a generalized gamma distribution with a high degree of flexibility, because its hazard rate function is non-monotonic (as recommended for the case of irregular spaced modeling [6,45]). The density function of the generalized gamma distribution is

$$g(x|\lambda, \gamma, k) = \frac{\gamma x^{k\gamma-1}}{\lambda^{k\gamma}\Gamma(k)} \exp\left\{-\left(\frac{x}{\lambda}\right)^{\gamma}\right\}, \tag{5}$$

where $\lambda, \gamma, k > 0$. The generalized gamma distribution includes many well-known distributions as special cases, such as the exponential distribution, the Weibull distribution, the half-normal distribution and the gamma distribution. In particular, the shape properties of the conditional hazard function can be derived from its parameter values. If $k\gamma < 1$, the hazard rate is decreasing for $\gamma \leq 1$ and U-shaped for $\gamma > 1$; if $k\gamma > 1$, the hazard rate is increasing for $\gamma \geq 1$ and inverted U-shaped for $\gamma \leq 1$; if $k\gamma = 1$, the hazard rate is decreasing for $\gamma < 1$, constant for $\gamma = 1$, and increasing for $\gamma > 1$.

Having specified the ground point process, we now parameterize the density of the mark distribution. If the separability is satisfied, this would greatly facilitate model building, fitting, and assessment [38]. In our context, a separable model for extreme cyber attack rates would posit that the magnitude distribution of extreme attack rates does not change over time, that is, not influenced by prior extreme cyber attack rates. However, this is rarely true in practice [43,44]. Therefore, the density of the mark distribution is also parameterized to depend on the history as well. The scale parameter is parameterized as the following form:

$$\beta(t, \tilde{x}_t|\mathfrak{F}_t) = \beta_0 + \beta_1 \tilde{x}_{N(t)} + \beta_2 \lambda_g(t|\mathfrak{F}_t),$$

where $\beta_0, \beta_1, \beta_2 > 0$. This leads to a time-dependent conditional intensity function

$$\lambda(t, \tilde{x}_t|\mathfrak{F}_t) = \frac{\lambda_g(t|\mathfrak{F}_t)}{\beta(t, \tilde{x}_t|\mathfrak{F}_t)} \left(1 + \xi \frac{x - \mu}{\beta(t, \tilde{x}_t|\mathfrak{F}_t)}\right)_{+}^{-1/\xi - 1}, \tag{6}$$

with

$$\lambda_g(t|\mathfrak{F}_t) = \lambda_0\left(\frac{t - t_{N(t)}}{\Psi_{N(t)}}\right)\frac{1}{\Psi_{N(t)}},$$

where $\lambda_0(\cdot)$ is the hazard function of the ϵ_i's. As a result, the log-likelihood function of the marked point process $N \in (0, T] \times (\mu, \infty]$ with history $\mathfrak{F}_t = (\{t_1, \tilde{x}_1\}, \ldots, \{t_{N(T)}, \tilde{x}_{N(T)}\})$ can be expressed as (see Section 7.3 of [15])

$$l = \sum_{i=1}^{N(T)} \log \lambda(t_i, \tilde{x}_{t_i}|\mathfrak{F}_{t_i}) - \int_0^{\mathrm{T}}\int_\mu^\infty \lambda(t, \tilde{x}|\mathfrak{F}_t)\,\mathrm{d}\tilde{x}\,\mathrm{d}t.$$

According to Equation (2), we have

$$l = \sum_{i=1}^{N(T)} \log \lambda_g(t_i|\mathfrak{F}_t) - \int_0^{\mathrm{T}} \lambda_g(u|\mathfrak{F}_u)\,\mathrm{d}u + \sum_{i=1}^{N(T)} \log f(\tilde{x}_i|\mathfrak{F}_i). \tag{7}$$

For a comprehensive discussion of marked point processes, one may refer to some good books [3,9,13,15,16,18,26,27] and some excellent papers [11, 36].

3.3. Dynamic cyber risk measure

We propose using VaR in the financial industry [32] to measure the cyber risk of intense attacks. The VaR at level α is defined as

$$\mathrm{VaR}_\alpha(t) = \inf\{l : P(\tilde{x}_t \le l) \ge \alpha\},$$

where $\tilde{x}_t$ is a mark (i.e. extreme attack rate) at time t and $0 < \alpha < 1$. The conditional VaR at level α based on the marked point process can be computed as

$$\mathrm{VaR}_\alpha(t) = u + \frac{\beta(t, \tilde{x}_t|\mathfrak{F}_t)}{1 - \xi}\left[\left(\frac{1 - \alpha}{\lambda_g(t|\mathfrak{F}_t)}\right)^{-\xi} - 1\right].$$

Intuitively, the VaR risk measure describes the probability of extreme cyber attack rates with a certain confidence level over a period of time. For example, if the VaR on cyber attack rates with 95% confidence level is 86×10^6 per hour, then there is only 5% chance that the hourly attack rate will exceed 86×10^6 attacks. Therefore, the VaR risk measure describes the extreme cyber attack rates that can lead to potentially catastrophic consequences, if the allocated defense resource is not adequate.

4. Simulation study

In this section, we present a simulation study to examine the performance of the proposed models. The experiment data are generated from the following models.

(a) The standard ACD model (ACD)

$$\Psi_i = \omega + a_1\Delta t_{i-1} + b_1\Psi_{i-1},$$

where $\omega, a_1, b_1 \ge 0$,

(b) The Log-ACD model

$$\Psi_i = \omega + a_1 \log(\Delta t_{i-1}) + b_1 \log(\Psi_{i-1}).$$

The time-dependent conditional intensity function is set to be the same as Equation (6).

Note that the proposed ACD/Log-ACD model consists of two components: the ground process describing the occurrences of extreme attack rates, and the conditional mark distribution describing the distribution of the magnitudes of the extreme attack rates. To evaluate the goodness-of-fit of the proposed models, we need to consider both the conditional mark distribution and the conditional intensity of the ground process.

- The conditional mark distribution. The conditional GPD assumption of the marks can be evaluated via the W-statistic, which is defined as

$$W_t = \xi^{-1} \log\left(1 + \xi \frac{\tilde{x}_t}{\beta(t, \tilde{x}_t|\mathfrak{F}_t)}\right). \tag{8}$$

  If the GPD parameters are correctly specified, then $\{W_t : t = 1, 2, \ldots, n\}$ approximately follows the standard exponential distribution (c.f. [7,15,32]). We use the KS test for the exponential distribution.
- The conditional intensity of the ground process. It is known from the theory of point process [7] that

$$\tau_i = \int_0^{t_i} \lambda_g(s|\mathfrak{F}_s)\, \mathrm{d}s, \quad i = 1, \ldots, N(t) \tag{9}$$

  would constitute a homogenous Poisson process of rate 1 on the interval $(0, t]$, which is in fact a part of the transformed time axis. If the conditional intensity of the ground process is correctly specified, then the inter-arrival times of the process $\{\tau_i : i = 1, \ldots, N(t)\}$ follow an independent and identical exponential distribution, which will be evaluated via the KS test as well.

To evaluate the performances of the ACD and Log-ACD models, for each model we randomly generate two data sets with a small sample size 400 and a large sample size 2000. The generated data sets are split into in-sample and out-of-sample parts with equal sample sizes.

In-sample fitting performance. For the ACD model, the parameters are set to be $\omega = 1.893$, $a_1 = .436$, $b_1 = .356$, $\gamma = .567$, $k = 2.439$ for the ground process, and $\xi = .219$, $\beta_0 = 7.507$, $\beta_1 = .138$, $\beta_2 = 6.873$ for the mark distribution. In what follows, we examine the fitting performance of the ACD model. The estimated parameters based on the 200 samples are $\hat{\omega} = 1.386(.466)$, $\hat{a}_1 = .686(.385)$, $\hat{b}_1 = .231(.687)$, $\hat{\gamma} = 1.333(.427)$, $\hat{k} = .886(.775)$ for the ground process, and $\hat{\xi} = .102(.936)$, $\hat{\beta}_0 = 6.744(.756)$, $\hat{\beta}_1 = .091(.857)$, $\hat{\beta}_2 = 9.826(1.202)$ for the mark distribution, where the values in the parentheses are the corresponding standard deviations from the inverse of the hessian matrix. It is seen that most of the estimated parameters are reasonably close to the real values. The p-values of the KS test for the mark distribution (W) and the conditional intensity of the ground process (τ) are .595 and .097, respectively. The estimated parameters and standard deviations based on 1,000 samples are $\hat{\omega} = 1.561(.457)$, $\hat{a}_1 = .530(.328)$, $\hat{b}_1 = .386(.337)$, $\hat{\gamma} =$

.507(.370), $\hat{k} = 3.836(.722)$ for the ground process, and $\hat{\xi} = .300(.431)$, $\hat{\beta}_0 = 5.116(.325)$, $\hat{\beta}_1 = .236(.490)$, $\hat{\beta}_2 = 8.960(.724)$ for the mark distribution. We observe the estimated parameters are closer to the true values with large sample sizes. The p-values of the KS test for the mark distribution and the conditional intensity of the ground process are .964 and .495, respectively. Therefore, we conclude that the in-sample fitting performance of the proposed ACD model is good for both small and large sample sizes.

Similarly, for the Log-ACD model, we set the parameters as $\omega = .357$, $a_1 = .452$, $b_1 = .460$, $\gamma = .399$, $k = 4.700$ for the ground process, and $\xi = .391$, $\beta_0 = 2.738$, $\beta_1 = .332$, $\beta_2 = 3.966$ for the mark distribution. In the following, we examine the fitting performance of the Log-ACD model. The estimated parameters based on 200 samples are $\hat{\omega} = .301(.155)$, $\hat{a}_1 = .274(.114)$, $\hat{b}_1 = .644(.104)$, $\hat{\gamma} = .505(.328)$, $\hat{k} = 2.527(1.873)$ for the ground process, and $\hat{\xi} = .561(.299)$, $\hat{\beta}_0 = 1.112(.916)$, $\hat{\beta}_1 = .778(.405)$, $\hat{\beta}_2 = 2.711(2.502)$ for the mark distribution. The estimated parameters based on 1,000 samples are $\hat{\omega} = .587(.208)$, $\hat{a}_1 = .326(.087)$, $\hat{b}_1 = .486(.124)$, $\hat{\gamma} = .640(.172)$, $\hat{k} = 2.874(1.429)$ for the ground process, and $\hat{\xi} = .311(.139)$, $\hat{\beta}_0 = 4.206(1.190)$, $\hat{\beta}_1 = .261(.151)$, $\hat{\beta}_2 = 5.144(1.984)$ for the mark distribution. It is seen that the estimates are reasonably close to the true values. The p-values of the KS test for the mark distribution and the conditional intensity of the ground process are respectively .967 and .704 in the case of 200 samples, and respectively .885 and .939 in the case of 1,000 samples. Therefore, we conclude that the in-sample fitting performance of the proposed Log-ACD model is also good for both small and large sample sizes.

Out-of-sample prediction performance. As discussed in Section 3.3, VaR is a good measure of cyber risk in terms of intense cyber attacks. In order to evaluate the prediction performance, we propose considering the violations of the predicted VaR values of extreme cyber attack rates. Specifically, the violation based on VaR_α is defined as

$$\mathrm{I}_\alpha(t) = \begin{cases} 1, & s_t > \mathrm{VaR}_\alpha(t), \\ 0, & o/w, \end{cases}$$

where s_t is the observed extreme attack rate, and $\mathrm{VaR}_\alpha(t)$ is the predicted VaR values of extreme cyber attack rates. For example, $\mathrm{VaR}_{.95}(t)$ describes there is only 5% chance that the observed extreme attack rate s_t would exceed the predicted value of $\mathrm{VaR}_{.95}(t)$. When this happens, we say a violation occurs.

In order to evaluate the prediction performance of the VaR values, we use three widely used tests in the literature [10,24]. The first test is the unconditional coverage test ($\mathrm{LR}_{\mathrm{uc}}$), which evaluates whether or not the fraction of violations is significantly different from the theoretical one. The second test is the independence of violations ($\mathrm{LR}_{\mathrm{ind}}$), where the present violation would have no effect on future violations under the null hypothesis. The third test is the conditional coverage test ($\mathrm{LR}_{\mathrm{cc}}$), which is a combination of the previous two tests. In what follows, we discuss the prediction performance of the proposed models based on those tests.

For the ACD and Log-ACD models, we evaluate the prediction performances based on the out-of-sample parts with 200 samples and 1,000 samples, respectively. We perform recursive rolling predictions, and the evaluations are based on 1-, 4-, 7-, and 10-hour ahead predictions. The testing results for $\alpha = .95$ are reported in Table 2. It is seen that the prediction performances are very satisfactory for both models. The p-values for all the cases

Table 2. Assessing prediction performance based on simulated out-of-sample data.

	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}
	sample size 200					sample size 1000				
ACD model										
1-hour	7	10	.305	.210	.269	47	50	.660	.333	.568
4-hour	8	10	.532	.298	.479	50	50	.983	.249	.514
7-hour	6	10	.191	.147	.148	49	50	.919	.272	.545
10-hour	9	10	.854	.338	.621	47	50	.708	.866	.919
Log-ACD model										
1-hour	10	10	1	.297	.581	53	50	.666	.465	.698
4-hour	10	10	.961	.293	.575	51	50	.868	.383	.674
7-hour	11	10	.675	.242	.461	52	50	.740	.428	.691
10-hour	9	10	.854	.338	.621	56	50	.357	.622	.579

Notes: Obs. represents the observed violations, and Exp. represents the expected violations. The α level is set to be .95.

are very large. It is interesting to observe that even for the 10-hour ahead prediction, the testing results for both models with small sample sizes are very good.

In summary, we conclude that the out-of-sample performances of proposed models are good for both small and large sample sizes.

5. Empirical analysis: extreme attack rates tracking

In this section, we evaluate the proposed models via the real data sets in terms of both in-sample (fitting) and out-of-sample (prediction) performances. For the telescope data sets, we use the first 20 days (i.e. 480 h) for model building and the rest 11 days (i.e. 264 h) for out-of-sample evaluation. For the honeypot data set, we use the first 1,108 hours for model building and the rest 500 hours for out-of-sample evaluation. As mentioned before, deploying an EVT-based method requires a proper threshold that offers the balance between bias and variance, because a small threshold u may result in a failure when approximating the POT model, but a large threshold u can result in few observations and hence a large variability. To be consistent with the exploratory analysis mentioned above, we use the 88% quantile of the in-sample observations in each of the telescope data sets as the threshold, and use the 90% quantile of the in-sample observations in the honeypot data set as the threshold.

5.1. In-sample fitting performance

Table 3 reports the p-values of the KS test for the conditional GPD distribution of the marks and the inter-arrival times τ, and the AICs for the ACD and Log-ACD models. We observe that both the ACD and Log-ACD models pass the test with p-values greater than .05 for the mark distribution (W) and the ground process (τ). The AICs of ACD models for the telescope data sets are smaller than the AICs of Log-ACD models. However, the AIC of Log-ACD model for the honeypot data set is smaller than that of ACD model. Therefore, both models may be used for modeling the dynamics of extreme cyber attack rates.

In order to characterize the effect of the dependence between the inter-exceedance times, we remove the coefficients related to the inter-exceedance times by setting $a_1 = b_1 = 0$. We find that none of the models could pass the test on the conditional intensity of

Table 3. The p-values of the KS test for the conditional intensity function (τ), the mark distribution (W), and the AICs for the ACD and Log-ACD models.

	τ	W	AIC	τ	W	AIC
		S_1			S_2	
ACD	.279	.875	728.548	.055	.970	741.466
Log-ACD	.158	.535	739.755	.311	.213	750.957
		S_3			Honeypot	
ACD	.056	.802	739.775	.055	.373	802.702
Log-ACD	.099	.545	759.698	.070	.919	799.124

the ground process τ at the level of .05. Specifically, for telescope data set S_1, we have the p-values of the τ's for the ACD and Log-ACD models are respectively .015 and .025; for telescope data set S_2, we have the p-values of the τ's for the ACD and Log-ACD models are respectively .021 and .020; for telescope data set S_3, we have the p-values of the τ's for the ACD and Log-ACD models are respectively .013 and .014. For the honeypot data set, the p-values of the τ's for the ACD and Log-ACD models are respectively .007 and .000. This means that the dependence between inter-exceedance times plays an important role in model fitting, which will be further confirmed below.

Table 4 shows the estimated parameters and the corresponding standard deviations for both the ACD and Log-ACD models, where the standard deviations are computed based on the inverse of the hessian matrix. We observe that except for the Log-ACD model of telescope data set S_3, the products of the estimated γ's and k's are all greater than 1, and all of the estimated γ's are smaller than 1. This means that the hazard rates are inverted U-shaped. For the Log-ACD model of telescope data set S_3, we have the estimated $\gamma = 1.009$ and $k = .897$, which means that the hazard rate is U-shaped. We also find that the coefficients a_1's and b_1's are significant in most of the cases. For example, for telescope data set S_3 and the Log-ACD model, we have the estimated parameters $a_1 = .367$ (.077) and $b_1 = .580$ (.068), which confirm the dependence between the inter-exceedance times. It is interesting to observe that the shape parameters, namely the ξ's, in the ACD models are not quite significant, which indicates that the GPDs degenerate to the exponential distributions in these cases.

In concluding the discussion on the in-sample fitting performance of the models, we recommend the Log-ACD model for fitting for two reasons. First, it has a better performance in fitting the underlying ground process, as indicated by the p-values of the τ's. Second, the estimated parameters in the Log-ACD model, such as a_1 and b_1, are more significant than their counterparts in the ACD models. This means that the Log-ACD model can better explain the cluster phenomenon of extreme cyber attack rates.

5.2. Out-of-sample prediction performance

We evaluate the prediction performance in terms of the short-term (1-hour ahead), mid-term (4-hour ahead and 7-hour ahead), and long-term (10-hour ahead) predictions. We use Algorithm 1 in the appendix to perform recursive rolling predictions. For the telescope data sets, 264 observations are used for the prediction evaluation; for the honeypot data set, 500 observations are used for the prediction evaluation. We predict $\mathrm{VaR}_\alpha(t)$'s

Table 4. Estimates of the ACD and Log-ACD models for the telescope and honeypot data sets.

	Ground process $\lambda_g(t\|\mathfrak{F}_t)$					POT $f(\tilde{x}_t\|\mathfrak{F}_t)$			
	ω	a_1	b_1	γ	k	ξ	β_0	β_1	β_2
S_1									
ACD	1.452	.464	.387	.478	3.599	.417	7.173	.028	4.134
	(.343)	(.354)	(.246)	(.368)	(.751)	(.456)	(.286)	(2.508)	(1.444)
Log-ACD	.488	.403	.435	.368	5.730	.759	2.808	.428	−3.051
	(.201)	(.105)	(.125)	(.063)	(1.933)	(.304)	(1.638)	(.215)	(4.481)
S_2									
ACD	1.893	.436	.356	.567	2.439	.219	7.507	.138	6.873
	(.318)	(.364)	(.274)	(.319)	(.657)	(.668)	(.338)	(.831)	(1.368)
Log-ACD	.357	.452	.460	.399	4.700	.391	2.738	.332	3.966
	(.154)	(.105)	(.109)	(.078)	(1.866)	(.171)	(1.507)	(.227)	(4.915)
S_3									
ACD	1.401	.518	.348	.422	4.580	.445	7.567	.115	1.617
	(.365)	(.360)	(.296)	(.399)	(.818)	(.427)	(.285)	(1.094)	(2.503)
Log-ACD	.326	.367	.580	1.009	.897	.272	4.796	.611	−5.209
	(.119)	(.077)	(.068)	(.239)	(.468)	(.165)	(1.785)	(.297)	(1.847)
Honeypot									
ACD	1.193	.259	.664	.078	99.049	.095	.273	.156	1.892
	(.713)	(.421)	(.196)	(.923)	(1.875)	(.662)	(.292)	(.581)	(.311)
Log-ACD	2.677	.363	−.558	.470	2.986	−.212	.531	.134	.919
	(.422)	(.102)	(.148)	(.194)	(2.402)	(.106)	(.127)	(.120)	(.738)

Note: The numbers in parentheses represent standard deviations.

at different α levels, and use the aforementioned statistical tests, namely $\mathrm{LR}_{\mathrm{uc}}$, $\mathrm{LR}_{\mathrm{ind}}$ and $\mathrm{LR}_{\mathrm{cc}}$, for evaluating the prediction performance.

Short-term predictions. Table 5 reports the short-term (i.e. 1-hour ahead) predictions, and the test results at α levels of .91, .93, .95, .97 for the telescope data sets as well as the test results at the α levels of .93, .95, .97 for the honeypot data set. For the predicted VaR_α of extreme attack rates, we observe that all of the p-values are greater than .1, meaning that the prediction models are accurate. In particular, the theoretical number of violations matches the observed number of violations well. Let us consider $\alpha = .95$ as an example. For telescope data set S_1, the number of violations is 13 based on the ACD model prediction or the Log-ACD model prediction, while the observed number of violations is exactly 13. For telescope data set S_2, the ACD model prediction slightly overestimates the number of violations by 3, while the Log-ACD model prediction underestimates the number of violations only by 1. For telescope data set S_3, the ACD model prediction overestimates the number of violations by 2, while the Log-ACD model prediction underestimates the number of violations only by 1. For the honeypot data set, both the ACD and Log-ACD predictions underestimate the number of violations. Since all of the p-values are large, there is no statistical evidence for rejecting the accuracy of the prediction models.

Figure 3 plots the 1-hour ahead predictions of the $\mathrm{VaR}_\alpha(t)$'s of extreme cyber attack rates using the ACD model with α at levels of .93 (purple curve), .95 (green curve), and .97 (red curve). We observe that although the three telescope data sets exhibit similar patterns, the predicted VaR's of extreme cyber attack rates exhibit quite different patterns, which are also different from the pattern exhibited by the predicted VaR's of extreme cyber attack rates for the honeypot data set. Figure 4 shows the 1-hour ahead predictions of the Log-ACD models with α at levels of .93 (purple curve), .95 (green curve), and .97 (red curve).

Table 5. Evaluation of 1-hour ahead predictions, where Obs. represents the observed violations, Exp. represents the violations of predictions, and $\overline{\text{VaR}}_\alpha$ represents the average of the predicted VaR_α values of extreme cyber attack rates where 'average' is in regards to α.

	α	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}	$\overline{\text{VaR}}_\alpha$
Telescope data set S_1							
ACD	.91	20	24	.407	.218	.332	80.308
	.93	15	18	.386	.171	.270	83.052
	.95	13	13	.955	.239	.499	86.824
	.97	6	8	.470	.594	.668	92.780
Log-ACD	.91	19	24	.290	.163	.216	79.789
	.93	16	18	.541	.976	.829	82.963
	.95	13	13	.955	.239	.499	87.313
	.97	5	8	.259	.658	.480	94.211
Telescope data set S_2							
ACD	.91	21	24	.546	.070	.162	84.853
	.93	18	18	.907	.119	.294	87.943
	.95	16	13	.443	.976	.745	92.191
	.97	8	8	.977	.475	.775	98.905
Log-ACD	.91	26	24	.635	.758	.852	84.279
	.93	19	18	.901	.569	.844	87.657
	.95	12	13	.731	.550	.788	92.222
	.97	5	8	.259	.658	.480	99.266
Telescope data set S_3							
ACD	.91	25	24	.791	.251	.500	86.921
	.93	17	18	.718	.383	.640	90.060
	.95	15	13	.619	.867	.871	94.348
	.97	9	8	.703	.420	.672	101.139
Log-ACD	.91	27	24	.494	.406	.560	86.325
	.93	18	18	.907	.811	.965	89.865
	.95	12	13	.731	.550	.788	94.755
	.97	7	8	.735	.533	.778	102.647
Honeypot data set							
ACD	.93	27	35	.145	.643	.311	3.635
	.95	21	25	.399	.267	.378	3.851
	.97	17	15	.608	.113	.249	4.187
Log-ACD	.93	27	35	.145	.663	.315	3.623
	.95	20	25	.288	.820	.555	3.863
	.97	16	15	.795	.522	.788	4.225

We observe the same kind of phenomena that are exhibited by the predictions of the ACD model mentioned above.

Mid-term predictions. For mid-term predictions, we report the results on 4-hour ahead and 7-hour ahead predictions. Table 6 reports 4-hour ahead predictions of the VaR's of extreme attack rates. LR_{uc}, LR_{ind} and LR_{cc} tests show that the p-values are large, meaning that the proposed models are suitable for 4-hour ahead predictions. We observe that most predictions are comparable to the 1-hour ahead predictions mentioned above.

Table 7 reports the 7-hour ahead predictions of the VaR's of extreme attack rates. We also observe that the p-values are large. When compared with the 4-hour ahead predictions mentioned above, we observe that in terms of the number of violations, 7-hour ahead predictions are slightly less accurate than 4-hour ahead predictions.

Long-term predictions. Table 8 reports the results of 10-hour ahead predictions of the VaR's of extreme attack rates based on the ACD and Log-ACD models. We observe that all of the p-values of the tests are greater than .05, meaning that the models can be used for

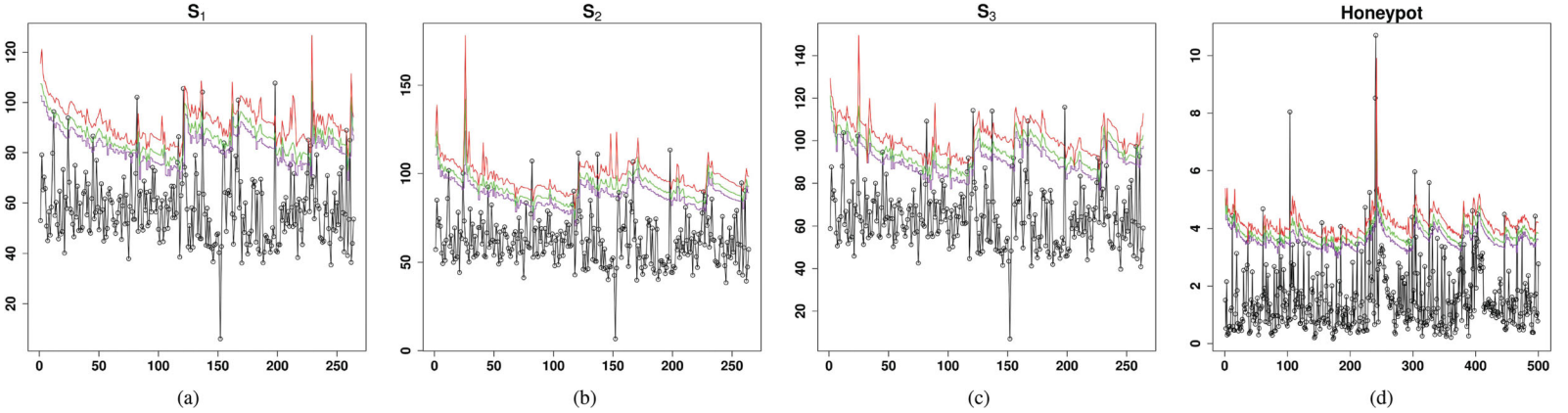

Figure 3. 1-hour ahead predictions of the VaR's of extreme attack rates based on the ACD model with α at levels of .93 (purple curve), .95 (green curve), and .97 (red curve). (a) Telescope data set S_1, (b) telescope data set S_2, (c) telescope data set S_3 and (d) honeypot data set.

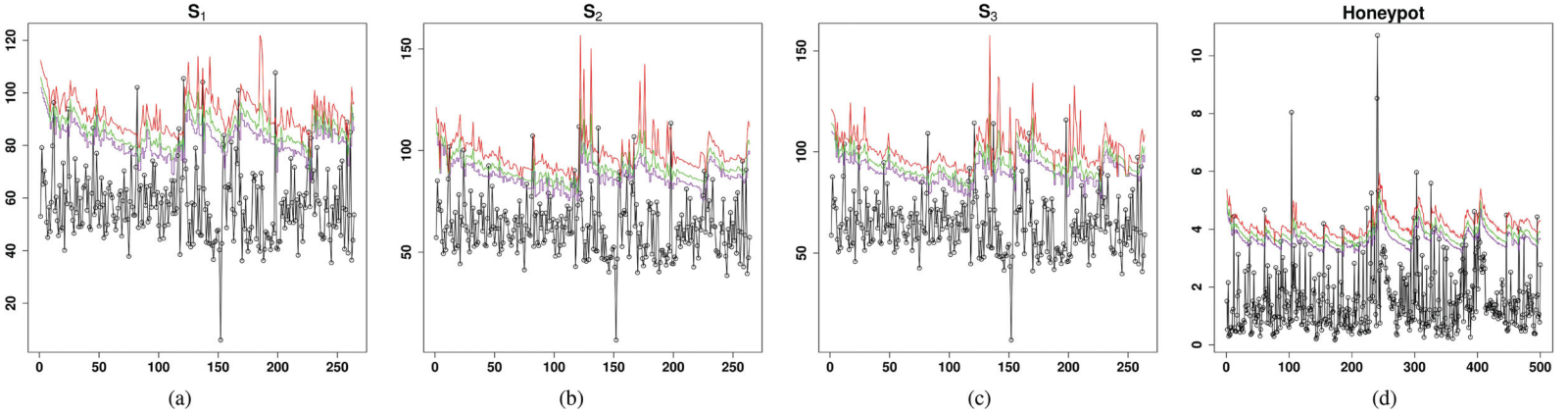

Figure 4. 1-hour ahead predictions of the VaR's of extreme attack rates based on the Log-ACD model with α at levels of .93 (purple curve), .95 (green curve), and .97 (red curve). (a) Telescope data set S_1, (b) telescope data set S_2, (c) telescope data set S_3 and honeypot data set.

Table 6. Evaluation of 4-hour ahead predictions, where Obs. represents the observed violations, Exp. represents the violations of predictions, and $\overline{\text{VaR}}_\alpha$ represents the average of the predicted VaR_α values of extreme cyber attack rates where 'average' is in regards to α.

	α	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}	$\overline{\text{VaR}}_\alpha$
Telescope data set S_1							
ACD	.91	20	23	.440	.225	.355	80.096
	.93	16	18	.574	.141	.289	82.982
	.95	12	13	.763	.275	.527	87.151
	.97	6	8	.489	.592	.682	94.436
Log-ACD	.91	19	23	.317	.580	.519	79.787
	.93	14	18	.281	.200	.246	82.763
	.95	12	13	.763	.275	.527	86.800
	.97	7	8	.759	.531	.784	93.064
Telescope data set S_2							
ACD	.91	22	23	.745	.370	.635	85.688
	.93	17	18	.755	.905	.946	88.647
	.95	13	13	.989	.236	.495	92.793
	.97	8	8	.951	.472	.771	99.586
Log-ACD	.91	25	23	.746	.075	.194	84.568
	.93	19	18	.860	.168	.381	87.951
	.95	12	13	.763	.557	.804	92.542
	.97	5	8	.272	.656	.496	99.669
Telescope data set S_3							
ACD	.91	24	23	.912	.198	.433	86.664
	.93	18	18	.948	.123	.303	89.805
	.95	14	13	.790	.767	.924	94.244
	.97	8	8	.951	.472	.771	101.628
Log-ACD	.91	26	23	.593	.107	.236	86.142
	.93	19	18	.860	.168	.381	89.411
	.95	13	13	.989	.131	.319	93.855
	.97	6	8	.489	.592	.682	100.801
Honeypot data set							
ACD	.93	28	35	.218	.593	.405	3.611
	.95	21	25	.416	.902	.713	3.830
	.97	17	15	.591	.597	.753	4.166
Log-ACD	.93	30	35	.389	.881	.683	3.604
	.95	23	25	.700	.379	.630	3.840
	.97	18	15	.431	.145	.254	4.196

long-term predictions. The performance of 10-hour ahead predictions is comparable to the performance of 7-hour ahead predictions in terms of the difference between the observed violations and expected violations for the telescope data sets. For the honeypot data set, it is observed that 10-hour ahead predictions are slightly better than 7-hour ahead predictions, indicating that the models are relatively stable over time.

For the magnitudes of the predicted VaR's of extreme attack rates, we consider the case of $\alpha = .95$ as an example because $\alpha = .95$ is widely used in the literature. For the telescope data sets, we add together all of the predicted $\text{VaR}_{.95}$ values with respect to 1-, 4-, 7- and 10-hour ahead predictions to reveal the prediction patterns of the ACD and Log-ACD models over time. Figure 5(a) plots the sum of the $\text{VaR}_{.95}$'s of extreme cyber attack rates for 1-, 4-, 7- and 10-hour ahead predictions, where the solid line corresponds to the ACD model and the dotted line corresponds to the Log-ACD model. We observe that although the predictions of the ACD model (solid curve) and the predictions of the Log-ACD model (dotted curve) are fairly close to each other, the two curves have different shapes. On the

Table 7. Evaluation of 7-hour ahead predictions, where 'Obs.' represents the observed violations, and 'Exp.' represents the violations of predictions, and $\overline{\text{VaR}}_\alpha$ represents the average of the predicted VaR_α values of extreme cyber attack rates where 'average' is in regards to α.

	α	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}	$\overline{\text{VaR}}_\alpha$
Telescope data set S_1							
ACD	.91	20	23	.474	.232	.378	79.798
	.93	15	18	.443	.166	.285	82.690
	.95	14	13	.756	.197	.415	86.635
	.97	6	8	.509	.590	.695	92.792
Log-ACD	.91	18	23	.239	.127	.155	79.359
	.93	14	18	.304	.776	.566	82.623
	.95	11	13	.578	.316	.518	87.106
	.97	6	8	.509	.590	.695	94.216
Telescope data set S_2							
ACD	.91	21	23	.624	.300	.519	84.700
	.93	18	18	.988	.491	.789	87.746
	.95	14	13	.756	.776	.915	91.976
	.97	9	8	.654	.415	.648	98.811
Log-ACD	.91	24	23	.866	.053	.152	84.173
	.93	15	18	.443	.886	.737	87.640
	.95	13	13	.977	.668	.912	92.330
	.97	7	8	.784	.528	.789	99.582
Telescope data set S_3							
ACD	.91	22	23	.789	.381	.657	87.350
	.93	15	18	.443	.886	.737	90.306
	.95	13	13	.977	.668	.912	94.621
	.97	6	8	.509	.590	.695	102.252
Log-ACD	.91	21	23	.624	.077	.185	86.316
	.93	16	18	.609	.317	.531	89.820
	.95	11	13	.578	.465	.656	94.591
	.97	5	8	.286	.654	.512	102.056
Honeypot data set							
ACD	.93	28	35	.231	.731	.460	3.613
	.95	22	25	.570	.984	.851	3.829
	.97	16	15	.759	.529	.782	4.164
Log-ACD	.93	28	35	.231	.731	.460	3.609
	.95	18	25	.147	.147	.122	3.835
	.97	16	15	.759	.089	.225	4.177

other hand, Figure 5(b) plots the predicted VaR's corresponding to the honeypot data set. We also observe that the predictions of the ACD model (solid curve) and the predictions of the Log-ACD model (dotted curve) are close to each other. Moreover, both curves are decreasing over h, which is the number of hours ahead of time the predictions are conducted.

Comparison of prediction performance. In this section, we compare the prediction performance of proposed models to that based on the Hawkes and ETAS models. The evaluations are based on 1-hour ahead and 2-hour ahead predictions. For the Hawkes and ETAS models, the kernel functions are specified in Equations (3) and (4), respectively.

Table 9 reports the 1-hour ahead predictions with $\alpha = .95$. It is seen that the overall performance of predictions of all these models is good because the p-values are all very large. For the telescope data sets, it is seen from Table 9 that the ACD and Log-ACD models overall outperform the Hawkes and ETAS models in terms of the observed number of violations. However, for the honeypot data, the Hawkes and ETAS models are better. Figure

Table 8. Evaluation of 10-hour ahead predictions, where Obs. represents the observed violations, Exp. represents the violations of predictions, and $\overline{\text{VaR}}_\alpha$ represents the average of the predicted VaR_α values of extreme cyber attack rates where 'average' is in regards to α.

	α	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}	$\overline{\text{VaR}}_\alpha$
Telescope data set S_1							
ACD	.91	20	23	.510	.710	.751	79.868
	.93	15	18	.473	.163	.292	82.749
	.95	12	13	.828	.269	.530	86.650
	.97	7	8	.809	.526	.794	92.678
Log-ACD	.91	21	23	.665	.080	.196	79.411
	.93	14	18	.327	.195	.267	82.660
	.95	13	13	.943	.230	.485	87.082
	.97	5	8	.300	.652	.528	93.988
Telescope data set S_2							
ACD	.91	21	23	.665	.080	.196	85.355
	.93	16	18	.644	.324	.552	88.210
	.95	15	13	.529	.896	.813	92.142
	.97	7	8	.809	.526	.794	98.381
Log-ACD	.91	22	23	.834	.114	.281	84.130
	.93	19	18	.780	.602	.840	87.507
	.95	15	13	.529	.896	.813	92.138
	.97	5	8	.300	.652	.528	99.482
Telescope data set S_3							
ACD	.91	21	23	.665	.823	.888	86.674
	.93	17	18	.834	.407	.694	89.581
	.95	14	13	.723	.785	.905	93.574
	.97	10	8	.409	.360	.468	99.904
Log-ACD	.91	24	23	.819	.056	.157	85.990
	.93	17	18	.834	.884	.968	89.315
	.95	13	13	.943	.676	.914	93.807
	.97	6	8	.530	.588	.708	100.735
Honeypot data set							
ACD	.93	29	34	.330	.816	.606	3.611
	.95	22	25	.591	.329	.537	3.827
	.97	16	15	.740	.090	.225	4.163
Log-ACD	.93	29	34	.330	.816	.606	3.574
	.95	21	25	.451	.276	.416	3.809
	.97	16	15	.740	.090	.225	4.162

6(a), (c), and (e) show the 1-hour predictions of the $\text{VaR}_{.95}$ of extreme attack rates for the telescope data. It is interesting to see that the Log-ACD model always predicts the largest average quantile values, which indicates that this model is more conservative. On the other hand, the ETAS model always predicts the smallest average quantile values, which indicates that this model is more optimistic. Figure 7(a) shows the 1-hour ahead predictions of the $\text{VaR}_{.95}$ of extreme attack rates for the honeypot data. Again, we observe that the overall predictions based on the ACD and Log-ACD models tend to have higher VaRs. It is also interesting to observe that the ETAS and Hawkes models would more likely to predict very high VaRs after seeing an extreme attack rate, which may be due to the triggering mechanisms of those models.

Table 10 reports the 2-hour ahead predictions with $\alpha = .95$. It is found again that all the p-values of all these models are very large, which indicates that a good prediction performance of all models. For the telescope data, it is seen that all the models have the similar prediction performance in terms of the number of violations. For the honeypot data set,

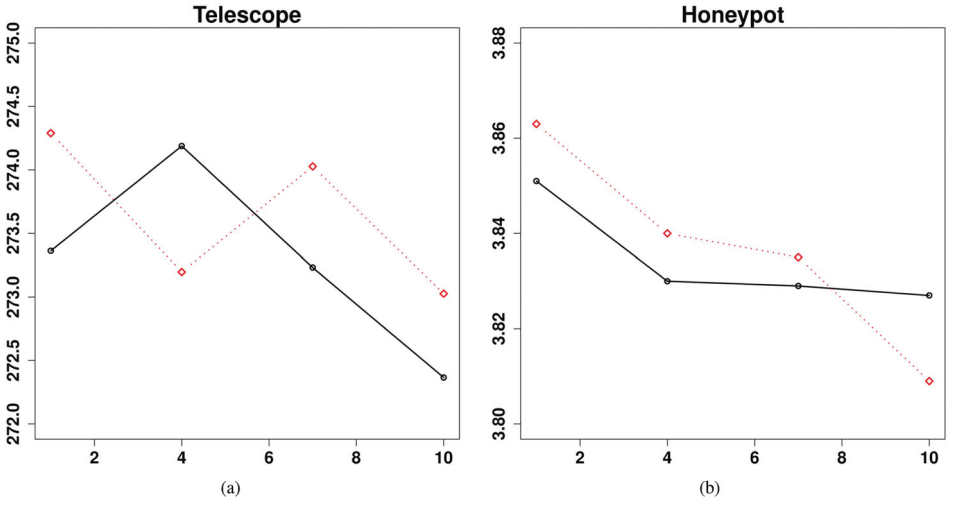

Figure 5. Sums of the short-, mid-, and long-term predicted $\text{VaR}_{.95}$'s of extreme cyber attack rates over h (the x-axis), which is the number of hours ahead of time the predictions are conducted. The solid line corresponds to the ACD model and the dotted line corresponds to the Log-ACD model. (a) $\text{VaR}_{.95}$ predictions for the telescope data sets and (b) $\text{VaR}_{.95}$ predictions for the honeypot data set.

Table 9. Assessing 1-hour ahead prediction performance based on the out-of-sample data.

Model	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}	$\overline{\text{VaR}}_{\alpha}$
Telescope data set S_1						
ACD	13	13	.955	.239	.499	86.824
Log-ACD	13	13	.955	.239	.499	87.313
Hawkes	15	13	.619	.867	.871	84.432
ETAS	15	13	.619	.234	.436	83.563
Telescope data set S_2						
ACD	16	13	.443	.976	.745	92.191
Log-ACD	12	13	.731	.550	.788	92.222
Hawkes	16	13	.443	.304	.439	89.858
ETAS	15	13	.619	.234	.436	89.344
Telescope data set S_3						
ACD	15	13	.619	.867	.871	94.348
Log-ACD	12	13	.731	.550	.788	94.755
Hawkes	15	13	.619	.234	.436	92.227
ETAS	14	13	.823	.176	.390	91.673
Honeypot data set						
ACD	21	25	.399	.267	.378	3.851
Log-ACD	20	25	.288	.820	.555	3.863
Hawkes	24	25	.836	.436	.722	3.789
ETAS	26	25	.838	.571	.834	3.746

Notes: Obs. represents the observed violations, Exp. represents the expected violations, and $\overline{\text{VaR}}_{\alpha}$ represents the average of the VaR predictions based on the α level.

the Hawkes model is slightly better than the others. Figure 6(b), (d), and (f) show the 2-hour ahead predictions of the $\text{VaR}_{.95}$'s of extreme attack rates for the telescope data set. Figure 7(b) shows the 2-hour ahead predictions of the $\text{VaR}_{.95}$'s of extreme attack rates for the honeypot data set. Similar conclusions can be drawn for both 1-hour ahead predictions and 2-hour ahead predictions.

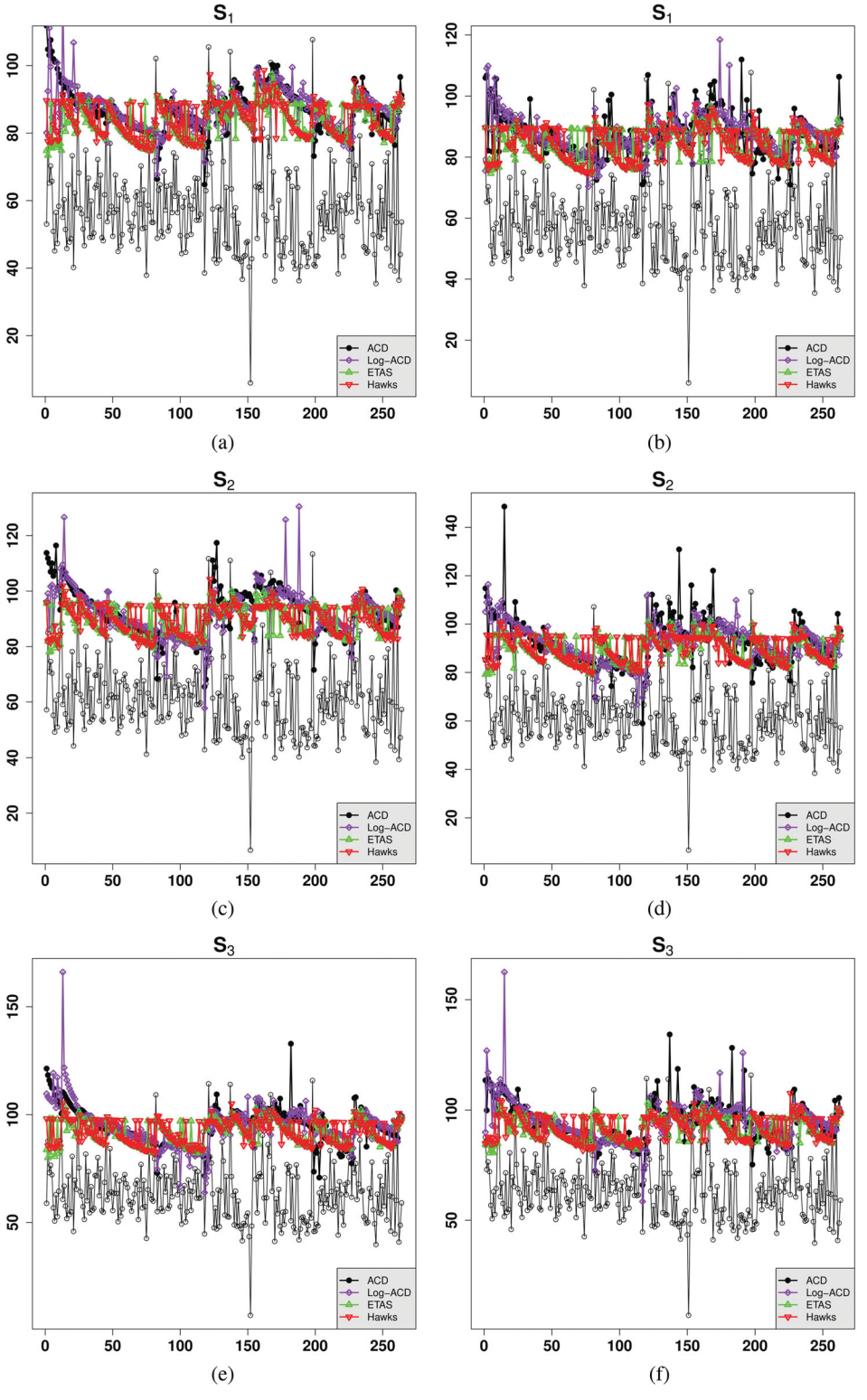

Figure 6. 1-hour and 2-hour ahead predictions of the $VaR_{.95}$'s of extreme attack rates for the telescope data set. (a) 1-hour ahead, (b) 2-hour ahead, (c) 1-hour ahead, (d) 2-hour ahead (e) 1-hour ahead and (f) 2-hour ahead.

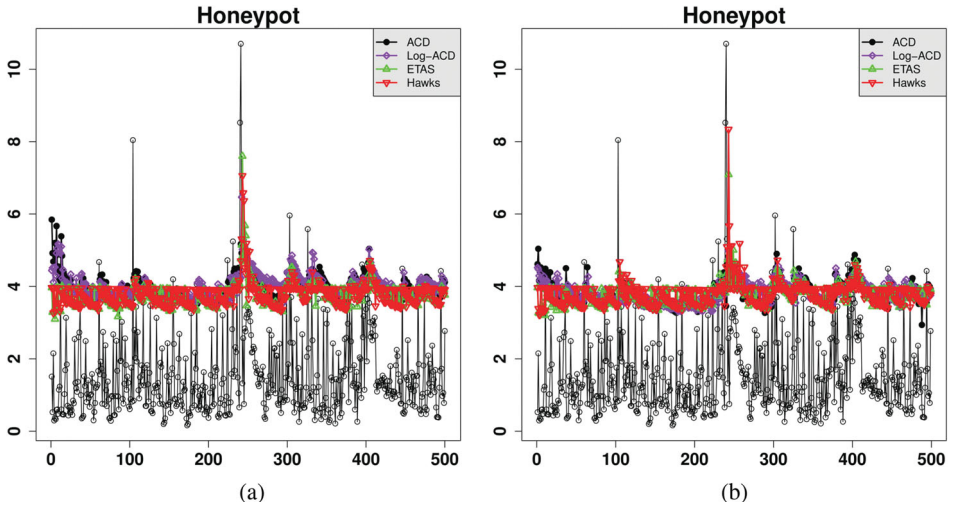

Figure 7. 1-hour and 2-hour ahead predictions of the $\text{VaR}_{.95}$'s of extreme attack rates for the honeypot data set. (a) 1-hour ahead and (b) 2-hour ahead.

Table 10. Assessing 2-hour ahead prediction performance based on the out-of-sample data.

Model	Obs.	Exp.	LR_{uc}	LR_{ind}	LR_{cc}	$\overline{\text{VaR}}_{\alpha}$
Telescope data set S_1						
ACD	14	13	.812	.202	.431	87.502
Log-ACD	12	13	.741	.299	.552	95.940
Hawkes	12	13	.741	.552	.794	84.564
ETAS	11	13	.532	.456	.623	84.364
Telescope data set S_2						
ACD	14	13	.812	.762	.928	92.563
Log-ACD	11	13	.532	.456	.623	91.985
Hawkes	13	13	.966	.129	.315	90.166
ETAS	13	13	.966	.257	.526	90.528
Telescope data set S_3						
ACD	16	13	.435	.980	.737	94.714
Log-ACD	13	13	.966	.655	.904	94.797
Hawkes	15	13	.608	.204	.392	92.099
ETAS	16	13	.435	.268	.400	91.338
Honeypot data set						
ACD	21	25	.405	.268	.382	3.835
Log-ACD	22	25	.537	.319	.503	3.837
Hawkes	25	25	.992	.503	.799	3.793
ETAS	26	25	.830	.572	.833	3.742

Notes: Obs. represents the observed violations, Exp. represents the expected violations, and $\overline{\text{VaR}}_{\alpha}$ represents the average of the VaR predictions based on the α level.

To conclude this section, we note that in terms of the number of violations of extreme cyber attack rates, the ACD model and the Log-ACD model with dynamic adjustment of quantiles can predict the extreme cyber attack rates. Moreover, the prediction models must accommodate the dependence between the inter-arrival times of extreme attack rates. This is because when ignoring the dependence between the inter-arrival times of extreme attack rates (by setting $a_1 = b_1 = 0$), we find that the overall prediction performance is poor because almost all models fail to pass the LR_{uc} test for $\alpha = .91$ at the significant level

of .05 (tables showing the details are available upon request). Compared to the Hawkes model and ETAS models, it is seen from the previous analysis that the proposed models are very competitive for the predictions. Particularly, for the 1-hour ahead prediction, the Log-ACD model overall outperforms the other models for the telescope data. Therefore, the ACD and Log-ACD models are very promising for modeling and predicting cyber attacks in the domain of cybersecurity.

6. Conclusions and future work

We have presented a novel application of marked point processes for modeling extreme cyber attack rates, where the ground process is accommodated by the ACD or Log-ACD model and the extreme values are accommodated by the POT method. The marked point processes can further accommodate the dependence between the inter-exceedance times of extreme attack rates, and hence can adequately describe the cluster behavior of the extreme attack rates. Our empirical analysis based on some real data sets shows that marked point processes offer accurate in-sample fitting performance and out-of-sample prediction performance.

There are many interesting problems that are left for future studies. In particular, we need to establish a systematic statistical framework that is tailored for the purpose of predicting extreme cyber attack rates. Moreover, we need to identify the optimal time resolution that leads to the most accurate predictions, while giving the defender sufficient time to dynamically allocate defense resources (noting that the specific time resolution used in the present paper is *hour*). Yet another research problem is to characterize the dependence between the time series of cyber attack rates, the time series of the number of attackers, and the time series of the number of victims (i.e. computers or IP addresses under attacks). The kind of dependence could be exploited to further improve the prediction accuracy.

Acknowledgements

The authors are very grateful to the two anonymous referees for their insightful and constructive comments which led to this improved version of the paper. We thank CAIDA for providing us the telescope data set that is analyzed in the present paper. We thank Sajad Khorsandroo for preprocessing the telescope data set and Zhenxin Zhan for preparing the honeypot data set.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

M. Xu and S. Xu were supported in part by ARO Grant #W911NF-13-1-0141, and T. Hu was supported by the NNSF of China [No. 11371340].

References

[1] S. Almotairi, A. Clark, G. Mohay, and J. Zimmermann, *Characterization of attackers' activities in honeypot traffic using principal component analysis*, Proceedings of the 2008 IFIP International Conference on Network and Parallel Computing, 2008, pp. 147–154.

[2] S. Almotairi, A. Clark, G. Mohay, and J. Zimmermann, *A technique for detecting new attacks in low-interaction honeypot traffic*, Proc. International Conference on Internet Monitoring and Protection, 2009, pp. 7–13.
[3] L. Anselin, *Spatial Econometrics: Methods And Models*, Vol. 4, Springer Science & Business Media, New York, 2013.
[4] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha, *Practical darknet measurement*, 2006 40th Annual Conference on Information Sciences and Systems, March 2006, pp. 1496–1501.
[5] L. Bauwens and P. Giot, *The logarithmic acd model: An application to the bid-ask quote process of three NYSE stocks*, Annales d'Economie et de Statistique 60 (2000), pp. 117–149.
[6] L. Bauwens, P. Giot, J. Grammig, and D. Veredas, *A comparison of financial duration models via density forecasts*, Int. J. Forecast. 20 (2004), pp. 589–609.
[7] V. Chavez-Demoulin, A.C. Davison, and A.J. McNeil, *Estimating value-at-risk: A point process approach*, Quant. Financ. 5 (2005), pp. 227–234.
[8] V. Chavez-Demoulin and J.A. McGill, *High-frequency financial data modeling using Hawkes processes*, J. Banking Financ. 36 (2012), pp. 3415–3426.
[9] S.N. Chiu, D. Stoyan, W.S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, John Wiley & Sons, Hoboken, NJ, 2013.
[10] P.F. Christoffersen, *Evaluating interval forecasts*, Internat. Econom. Rev. 39 (1998), pp. 841–862.
[11] A. Christou Micheas, *Hierarchical Bayesian modeling of marked non-homogeneous Poisson processes with finite mixtures and inclusion of covariate information*, J. Appl. Stat. 41 (2014), pp. 2596–2615.
[12] K. C. Claffy, H.-W. Braun, and G. C. Polyzos, *A parameterizable methodology for internet traffic flow profiling*, IEEE J. Sel. Areas Commun. 13 (1995), pp. 1481–1494.
[13] N. Cressie, *Statistics for Spatial Data*, Wiley Series in Probability and Statistics, Vol. 15, Wiley-Interscience, New York, 1993, pp. 105–209.
[14] A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescapè, *Analysis of a '/0' stealth scan from a botnet*, Proceedings of the 2012 ACM Conference on Internet Measurement Conference (IMC'12), 2012, pp. 1–14.
[15] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Vol. 1, 2nd ed., Springer, New York, 2002.
[16] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, Springer Science & Business Media, New York, 2007.
[17] S. Dharmapurikar, P. Krishnamurthy, T. Sproull, and J. Lockwood, *Deep packet inspection using parallel bloom filters*, Proceedings 11th Symposium on High Performance Interconnects, IEEE, 2003, pp. 44–51.
[18] P.J. Diggle, *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, CRC Press, Boca Raton, FL, 2013.
[19] T. Dubendorfer and B. Plattner, *Host behaviour based early detection of worm outbreaks in internet backbones*, Proc. IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise, 2005, pp. 166–171.
[20] P. Embrechts, C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997.
[21] R.F. Engle and J.R. Russell, *Autoregressive conditional duration: A new model for irregularly spaced transaction data*, Econometrica (1998), pp. 1127–1162.
[22] Y. Gao, Z. Li, and Y. Chen, *A dos resilient flow-level intrusion detection approach for high-speed networks*, Proc. IEEE International Conference on Distributed Computing Systems (ICDCS'06), 2006, p. 39.
[23] E. Glatz and X. Dimitropoulos, *Classifying internet one-way traffic*, Proceedings of the 2012 ACM Conference on Internet Measurement Conference (IMC'12), 2012, pp. 37–50.
[24] R. Herrera and B. Schipp, *Value at risk forecasts by extreme value models in a conditional duration framework*, J. Empir. Financ. 23 (2013), pp. 33–47.

[25] A. Hussain, J. Heidemann, and C. Papadopoulos, *A framework for classifying denial of service attacks*, Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM '03, ACM, New York, NY, USA, 2003, pp. 99–110.
[26] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns*, Vol. 70, John Wiley & Sons, Hoboken, NJ, 2008.
[27] A. Karr, *Point Processes and their Statistical Inference*, Vol. 7, CRC press, New York, 1991.
[28] T. Kumazawa and Y. Ogata, *et al. Nonstationary ETAS models for nonstandard earthquakes*, Ann. Appl. Stat. 8 (2014), pp. 1825–1852.
[29] F. Lau, S.H. Rubin, M.H. Smith, and L. Trajkovic, *Distributed denial of service attacks*, IEEE International Conference on Systems, Man, and Cybernetics, 2000, Vol. 3, 2000, pp. 2275–2280.
[30] Z. Li, A. Goyal, Y. Chen, and V. Paxson, *Towards situational awareness of large-scale Botnet probing events*, IEEE Trans. Inf. Forensics Secur. 6 (2011), pp. 175–188.
[31] C. Livadas, R. Walsh, D. Lapsley, and W. Timothy Strayer, *Using machine learning techniques to identify botnet traffic*, Proc. IEEE LCN Workshop on Network Security (WoNS'2006), 2006, pp. 967–974.
[32] A.J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press, Princeton, NJ, 2010.
[33] T. Mikosch, *Modeling dependence and tails of financial time series*, Extreme Values in Finance, Telecommunications, and the Environment, 2003, pp. 185–286.
[34] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, *Inferring internet denial-of-service activity*, ACM Trans. Comput. Syst. 24 (2006), pp. 115–139.
[35] Y. Ogata, *Statistical models for earthquake occurrences and residual analysis for point processes*, J. Amer. Statist. Assoc. 83 (1988), pp. 9–27.
[36] Y. Ogata, *Space-time point-process models for earthquake occurrences*, Ann. Inst. Stat. Math. 50 (1998), pp. 379–402.
[37] S. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, Ithaca, NY, 2007.
[38] F.P. Schoenberg, *Multidimensional residual analysis of point process models for earthquake occurrences*, J. Amer. Statist. Assoc. 98 (2003), pp. 789–795.
[39] O. Thonnard and M. Dacier, *A framework for attack patterns' discovery in honeynet data*, Digit. Investigation 5 (2008), pp. S128–S139.
[40] N. Weiler, *Honeypots for distributed denial-of-service attacks*, Proceedings. Eleventh IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2002. WET ICE 2002, 2002, pp. 109–114.
[41] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston, *Internet background radiation revisited*, Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, IMC '10, ACM, New York, NY, USA, 2010, pp. 62–74.
[42] V. Yegneswaran, P. Barford, and D. Plonka, *On the design and use of internet sinks for network abuse monitoring*, Recent Advances in Intrusion Detection, Springer, 2004, pp. 146–165.
[43] Z. Zhan, M. Xu, and S. Xu, *Characterizing honeypot-captured cyber attacks: Statistical framework and case study*, IEEE Trans. Inf. Forensics Secur. 8 (2013), pp. 1775–1789.
[44] Z. Zhan, M. Xu, and S. Xu, *Predicting cyber attack rates with extreme values*, IEEE Trans. Inf. Forensics Sec. (2015), pp. 1666–1677.
[45] M.Y. Zhang, J.R. Russell, and R.S. Tsay, *A nonlinear autoregressive conditional duration model with applications to financial transaction data*, J. Econometrics 104 (2001), pp. 179–207.

Appendix

(a) *Threshold selection for marked point process.* It is known in the extreme value theory that if a high threshold is selected, then the bias may be reduced but the estimates may not be stable. On the other hand, if a lower threshold is selected, the extreme value theory may not be applicable [32,33,37]. Since extreme cyber attack rates tend to be dependent upon each other, the standard method of the extreme value theory for the threshold selection cannot used. In this

paper, we propose a practical method for selecting a proper threshold, which would produce stable results in terms of the conditional ground process and the mark distribution. Specifically, for the telescope data sets and the honeypot data sets, we fix in advance possible threshold quantiles $q \in [.8, .95]$ with an increment .1 based on the training data sets. For each threshold value, we fit the proposed models, and then calculate the W-statistics in Equation (8) and τ in Equation (9). For the telescope data sets, we only report the results based on S_1 because the other testing results are similar. Table A1 presents the p-values of the KS test for the conditional intensity function and the mark distribution for the ACD and Log-ACD models. It is seen that for the telescope data, the testing results are very stable above the 87% quantile, except that the p-values for W and τ become a little bit smaller for the 89% quantile (which is still acceptable at level .05). Therefore, for the telescope data sets, we select the 88% quantile for modeling purpose. For the honeypot data, it is seen that above the 90% quantile, the testing results become stable for both models, although the p-value of τ for the Log-ACD model at the 91% quantile is slightly smaller than .05. Therefore, for the honeypot data, the 90% quantile seems to be a reasonable threshold for modeling purpose.

(b) *Algorithm for recursive rolling prediction.* The following algorithm is used for the out-of-sample prediction in the empirical study.

Algorithm 1 Recursive rolling prediction of the VaR_α's of extreme attack rates

INPUT: extreme attack rates time series $\{(t, \tilde{x}_t)\}$; α (level); $l_1 = l_2 = l_3 = 480$ (480 samples in the telescope data sets for model building), $l_4 = 1108$ (1108 samples in the honeypot data set for model building); h (# of hours ahead prediction; $h = 1, 4, 7, 10$)
OUTPUT: predicted VaR_α

for $i = 1$ **to** 4 **do**
  $m = l_i - h + 1$
  while $m + h \leq n$ **do**
    use $\{(t, \tilde{x}_t), t \leq m\}$ in the telescope data set S_i $(1 \leq i \leq 3)$ or the honeypot data set $(i = 4)$ to fit model M_i
    use the fitted model M_i to predict $\text{VaR}_\alpha(m + h)$
    $m = m + 1$
    return $\text{VaR}_\alpha(m + h)$
  end while
end for

Table A1. The *p*-values of the KS test for the conditional intensity function (τ) and the mark distribution (W) based on the ACD model and the Log-ACD model.

	Telescope data				Honeypot			
	ACD		Log-ACD		ACD		Log-ACD	
μ	W	τ	W	τ	W	τ	W	τ
.80	.227	.003	.871	.302	.139	0	.595	0
.81	.630	.002	.934	.041	.413	0	.816	0
.82	.701	.002	.705	.401	.188	0	.629	0
.83	.339	.003	.947	.398	.131	.001	.803	0
.84	.378	.026	.961	.525	.180	0	.950	0
.85	.412	.038	.928	.105	.122	0	.586	0
.86	.836	.007	.962	.002	.209	.001	.967	0
.87	.575	.048	.808	.318	.344	.190	.701	.004
.88	.875	.279	.535	.158	.308	.002	.805	.003
.89	.903	.094	.957	.056	.325	.004	.812	.018
.90	.967	.392	.856	.199	.373	.055	.919	.070
.91	.836	.327	.958	.496	.410	.064	.662	.049
.92	.987	.395	.955	.211	.947	.085	.954	.053
.93	.729	.493	.959	.292	.929	.129	.995	.073
.94	.910	.349	.958	.283	.925	.094	.485	.086
.95	.919	.701	.925	.088	.648	.307	.647	.211