How Should Enterprises Quantify and Analyze (Multi-Party) APT Cyber-Risk in their Industrial IoT Network?

How Should Enterprises Quantify and Analyze (Multi-Party) APT Cyber-Risk Exposure in their Industrial IoT Network?

RANJAN PAL, Massachusetts Institute of Technology, USA
ROHAN XAVIER SEQUEIRA, University of Southern California, USA
XINLONG YIN, Georgia Institute of Technology, USA
SANDER ZEIJLEMAKER, Massachusetts Institute of Technology, USA
VINEETH KOTALA, University of Illinois Urbana-Champaign, USA

Industrial Internet of Things (IIoT) networks (e.g., a smart grid industrial control system) are increasingly on the rise, especially in smart cities around the globe. They contribute to meeting the day-to-day needs (e.g., power, water, manufacturing, transportation) of the civilian society, alongside making societal businesses more efficient, productive, and profitable. However, it is also well known that IoT devices often operate on poorly configured security settings. This increases the chances of occurrence of (nation-sponsored) stealthy spread-based APT malware attacks in IIoT networks that might go undetected over a considerable period of time. Such attacks usually generate a negative *first-party* QoS impact with financial consequences for companies owning such IIoT network infrastructures. This impact spans (i.e., aggregates) *space* (i.e., the entire IIoT network or a sub-network) and *time* (i.e., duration of business disruption), and is a measure of significant interest to managers running their businesses atop such networks. It is of little use to network resilience boosting managers if they have to wait for a cyber-attack to happen to gauge this impact. Consequently, one of the questions that intrigues us is: *can managers estimate this first-party impact prior to APT cyber-attack(s) causing financial damage to companies?*

In this paper, we propose the first computationally efficient and quantitative network theory framework to (a) characterize this first-party impact *apriori* as a statistical distribution over multiple attack configurations in a family of malware-driven APT cyber-attacks specifically launched on businesses running atop IIoT networks, (b) accurately compute the statistical moments (e.g., mean) of the resulting impact distribution, and (c) tightly bound the accuracy of worst-case risk estimate of such a distribution - captured through the tail of the distribution, using the Conditional Value at Risk (CVaR) metric. In relation to (a) above, our methodology extends the seminal Factor Analysis of Information Risk (FAIR) cyber-risk quantification methodology that does not explicitly account for network interconnections among system-risk contributing variables. We validate the effectiveness of our theory using trace-driven Monte Carlo simulations based upon test-bed experiments conducted in the FIT IoT-Lab. We further illustrate quantitatively that even if spread-based APT cyber-attacks induce a statistically light-tailed first-party cyber-loss distribution on an IIoT networked enterprise in the worst case, the aggregate multi-party cyber-risk distribution incurred by the same enterprise in supply-chain ecosystems could be heavy-tailed. This will pose significant market scale-up challenges to cyber-security improving commercial cyber (re-)insurance businesses. We subsequently propose managerial action items to mitigate the first-party cyber-risk exposure emanating from any given IIoT driven enterprise.

Additional Key Words and Phrases: IIoT, APT, network, cyber-risk, security, CVaR, supply chain, insurance

Authors' addresses: Ranjan Pal, Massachusetts Institute of Technology, USA, ranjanp@mit.edu; Rohan Xavier Sequeira, University of Southern California, USA, rsequeir@usc.edu; Xinlong Yin, Georgia Institute of Technology, USA, connory@umich.edu; Sander Zeijlemaker, Massachusetts Institute of Technology, USA, szeijil@mit.edu; Vineeth Kotala, University of Illinois Urbana-Champaign, USA, vkotala2@illinois.edu.

2158-656X/2023/10-ART
https://doi.org/10.1145/3605949

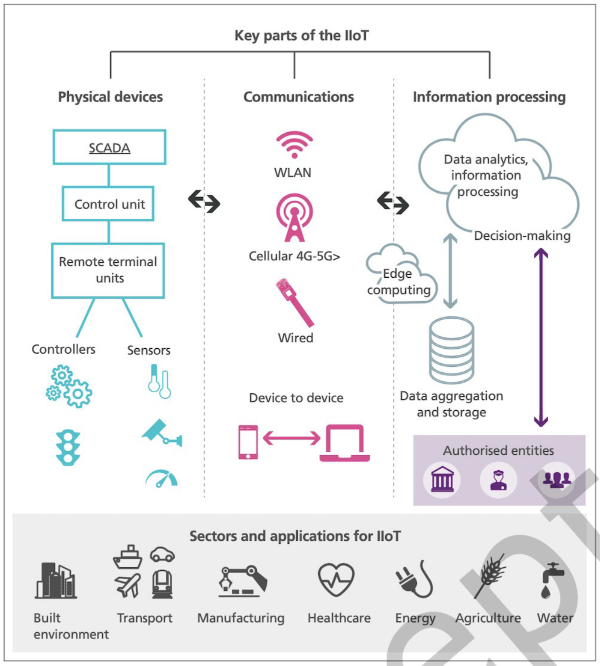


Fig. 1. Illustration of IIoT Features and Application Sectors (*Source:* Google Images)

1 INTRODUCTION

Cloud and sensor network driven machine-to-machine (M2M) communication is triggering a paradigm shift in the way various industrial (decision-making) processes are conducted and managed. The aggressive adoption of this M2M communication for service applications in smart grids, and across various (critical) industries and industry verticals such as automotive, utilities, home automation, healthcare, and security (see Figure 1), is expected to rapidly accelerate the industrial Internet of Things (IIoT) market. To drive home our point, the global IIoT market, as of 2021, is worth a USD 100+ billion dollars (projected to reach a trillion USD by 2028), with a steady yearly growth rate, i.e., a CAGR of ≈ 22.8%, according to a recent report by *Grand View Research, Inc.*

The effectiveness of an industrial Internet of Things (IIoT) network (e.g., a smart grid, a smart factory) solely relies upon the reliable and resilient functioning of networked IoT devices that operate collaboratively in collecting, transmitting, relaying, and intelligently processing application information. However, the last half a decade has seen (nation-state sponsored) attacks with increasing frequency both on IoT-driven industrial control systems (ICS) and on the operational technology (OT) side of the IIoT. The prime set of reasons (see [125]) attributed to this trend are (a) the increase in the density and scale of sensor networks associated with an organization's IIoT forming a large cyber-attack surface that has sub-optimal OT/IT air-gaps, (b) hard-to-replace old but Internet exposed OT and control equipment modestly capable of being robust to cyber-attacks, (c) IoT-equipped machines (e.g., HMI computers, SCADAmaster computers, PLCs) from multiple vendors running a patchwork of proprietary, heterogeneous, and non-updatable software, (d) poor or absent (behavioral) organizational cyber-security practices (e.g., poorly configured default security IoT device settings), (e) organizational C-suites allocating insufficient budget to implement cyber-security awareness, monitoring, and prevention technology, (f) a rise in hybrid and in-secure remote work environments post the COVID outbreak, and (g) a significant rise in the number, type, and quality of cyber-attackers.

The above set of reasons increase the chances of occurrence of stealthy but popular advanced persistent threat (APT) attacks in IIoT network systems that (a) spear-phish IT administrators or other employees to gain persistent high level IT network access within the target industrial site via (but not limited to) user accounts, hardware and software assets, (b) consequentially, move laterally inside the site to find and exploit insecure devices and machines (e.g., botnets discovering open RDP ports using the *Shodan* search engine), (c) usually go undetected (due to weak security monitoring) over a considerable period of time (e.g., via a timed logic bomb in a malware that might delay activation of specified maximal adverse impact event(s)) - consequently not allowing the defender(s) to segment and isolate the network, and (d) result in cascading IoT device (and machine) failures that disrupt device and employee networked communication systems, and cause significant physical and/or service quality damage in the long-run.

The quantitative measure of such an adverse impact in an arbitrary IIoT network is usually a time-dependent and non-deterministic (random) variable reflecting a loss (usually converted to tangible economic units) rooted on the quality-of-service (QoS) provided through the network. As an example of an outcome of such a random variable, consider a modern cyber-inspired version of the Northeast power blackout of 2003 that was triggered by a transmission line failure and cascaded into a massive power failure affecting 55 million people in the northeast region of the USA, and resulted in an economic (first and third-party damage) of USD 6 billion [72]. Today, multiple transmission lines within an IIoT-controlled power grid network can be simultaneously compromised by an APT and can potentially cause far greater economic and physical damage to in time and space than one can imagine.

Popular examples of APT attacks on various IIoT systems in multiple application sectors include *NotPetya*, *Stuxnet*, *Ramnit*, *Shamoon*, *BlackEnergy*, *Triton*, and *NightDragon* [125], most of which have accrued significant monetary damage to system managers [126]. Though these examples primarily cover state-sponsored attacks on critical infrastructure, we imply a broader space of APT attacks on IIoT systems beyond critical infrastructure and those that are not state-sponsored. Henceforth, ‘IoT devices’ will refer to the broader class of devices and machines equipped with IoT technology.

1.1 Research Motivation

Two strong practical scenarios motivate our research in this paper.

Scenario 1 - It is often the case in recent years, in the wake of major cyber-attacks in the past decade, that an ICS enterprise management (e.g., CEO, CISO, board) is interested to tangibly estimate *apriori* statistical metrics (e.g., mean, tail-risk) related to the cyber-loss impact post a cyber-attack (e.g., via an APT) event. After all, managers cannot wait for a cyber-attack to cause financial damage to an enterprise to start the cyber-risk management process.

In addition, note that it is usual in practice for managers to find it difficult to estimate the hypothetical impact of an adverse cyber-incident [24]. This difficulty is aggravated via uncertainties in the knowledge of the cyber-risk terrain, system complexity, lack of cyber-incident data and cyber-loss impact metrics, and the inability to predict future cyber-incidents [47, 55]. Add to this is the role of cognitive biases that prevent even the most experienced of system managers to assess the impact of cyber-risk accurately enough [66, 108, 111].

Hence, managers should prioritize to ‘simulate’ in advance multiple cyber-attack configuration scenarios to apriori derive attack impact statistics showcasing the likelihood of best, average, and worst possible financial impact events rather than work with a faulty perception of exact cyber-risk impact. This it prioritizes to minimize tangible (e.g., monetary, stock value) and non-tangible (e.g., reputation) multi-party losses post an inevitable future cyber-attack event through investing effectively in cyber-protection mechanisms. As management guru Peter Drucker once famously said: *“if you cannot measure it, you cannot manage it".*

Scenario 2 - A (standalone) cyber (re-)insurance market to mitigate adverse financial impacts for the multiple (IIoT driven) societal end-user vertical sectors has grown significantly in the last decade, to a point where it will soon move beyond the US$10 billion annual mark globally. However, the market is severely sparse with a supply-demand gap of approximately hundreds of billions of dollars [16][102]. In other words, the supply is far less than the coverage demand from the IIoT sectors.

One primary reason for this wide gap is the lack of robust quantitative estimates of adverse non-binary impact distributions in IIoT networks post (APT) cyber-attack events. More specifically, cyber-insurers, through internal audit processes, are interested to get accurate-enough estimates of (a) statistical moments (e.g., mean, variance) of the adverse non-binary impact of a cyber-attack for an IIoT network client, and (b) the statistics of worst k-th (usually between 5 and 10) percentile of the associated adverse impact. A non-robust estimate of such statistical metrics might expose cyber-insurers to unwanted tail risks they will be blind about. In risk theory and popular industry parlance, the latter statistics (i.e., that in (b)) is often measured through the Conditional-Value-at-Risk (CVaR) metric [70] (see a concise background in Section 5), which in our current work moulds itself into a measure of the APT risk in an IIoT network. The values of these moments and APT risk are (a) a suitable proxy to a measure of the cyber (in)security in an IIoT network, and (b) essential (if not sufficient) to the design of suitable commercially viable coverage policy parameters (e.g., premiums, deductibles) for the IIoT network client.

To the best of our knowledge, a general event apriori systematic quantification and formal analysis of network and time dependent cyber-risk statistics arising from a family of stealthy spread-based APT malware attacks is absent from the managerial toolbox and the (I)IoT security literature (see more details in Section 7).

Research Goals - Our main research goal in this paper is to develop a cyber-risk quantification methodology serving two purposes. Our first purpose is to accurately estimate, alongside providing rigorous performance bounds, the vital statistics of the first-party time-dependent adverse impact distribution generated by a family of spread-based APT malware attack in an IIoT network. Our second purpose is to tightly bound the APT risk, i.e., the CVaR, of the said distribution. *Our obtained statistics and bounds are event apriori estimates and reflect (among best and average estimates) the worst case cyber-loss impact incurred by an enterprise post the occurrence of an APT cyber-attack.*

As an important side goal, we aim to study how an aggregate of such cyber-risks sourced from multiple interdependent IIoT-driven organizations (enterprises) that are part of networked service supply chain ecosystems affect the market sustainability of coverage solutions provided by residual cyber-risk managers (e.g., cyber-insurers) tasked upon managing aggregate cyber-risk.

Broader Impact of Our Research on Cyber-Security - The inability of a cyber (re-)insurer to derive robust estimates of first-party cyber-risk inside an IIoT network post an APT cyber-breach event will (a) drive the latter to remain conservative in pricing attractive risk coverage policies for its clients, (b) consequently prevent the current cyber insurance market to grow more dense, (c) subsequently disallow the transference of appropriate cyber-hygiene liability upon IIoT network managers, and (d) finally, adversely impact cyber-security in the IIoT network simply because dense cyber-insurance markets necessarily promote cyber-security [6][61][18][84][83][86][107][133][16] [102]). Our proposed research methodology to assess APT (tail) cyber-risk impact will also enable the C-suite of IT/ICS driven organizations (enterprises) invest appropriate amount of time and money on securing ‘central’ adversary targets (e.g., processes, humans, hardware) within the organization affected by APTs to boost cyber-resilience and reduce tail risk. Examples of such investment products include (but not limited to) (a) perimeter security elements (such as firewalls, antivirus, intrusion detection systems [IDS], proxy servers, and Remote Authentication Dial-In User Service [RADIUS] servers) to secure critical hardware and software processes in operation in an IIoT network, (b) internal IT auditing, business continuity and disaster recovery (BCP/DR) processes inside the IIoT network if critical resources are compromised, (c) IT governance measures, which mandate that effective management, policies, controls, and procedures are in place to ensure that IIoT information systems support the business organizations' objectives, control access to network assets, and minimize IoT-related risk on 'central' organization (enterprise) targets.

1.2 Research Contributions

We make the following research contributions in this paper.

(1) We design a novel stealthy cyber-malware spreading framework for a parameterized family of APT-type cyber-attacks in IIoT networks, that captures the time-varying *attack-defense-impact* trio as an outcome of a time-dependent *Markov-Feller* (MF) continuous stochastic process. This stochastic process is ideally suited for an enterprise manager to broadly model, *apriori*, multiple stealthy infection spread paths from a parameterized family of APTs across a network topology, and their impact launching time periods for each path. Our proposed model extends a huge literature on *attack-defense* type models that omit providing a systematic framework to quantify the adverse impact on organizational (wireless) network assets when modeling cyber-malware spread as a time-dependent continuous stochastic process only, without modeling cyber-loss impact (see Section 2).

(2) We investigate the existence of malware spread process stability conditions for an IIoT network under which the number of infected network nodes will completely die out. We show that such a condition is not achievable in practice (primarily due to imperfect security technology), and that there will always be a mix of IIoT network nodes some of which will be susceptible to APT attacks when the MF spread process converges, and the others will be in an infected state.

We provide tight upper and lower bounds on process parameters (such as the mean of malware infection rates) as a function of individual node security strength. These bounds serve as tangible guidelines to network security managers and the organizational C-suite on ways (e.g., with respect to driving optimal resource investments) to improve node level (and consequently network level) security. Furthermore, for the MF malware spread process, we provide a closed-form analysis of the non-deterministic time-aggregate adverse impact of an APT attack on the entire IIoT network. The non-deterministic outcome represents a statistical distribution that is a result of considering multiple cyber-attack configurations within a parameterized family of APT cyber-attacks via the MF stochastic process.

This 'time-space' adverse impact value represents first-party cyber-risk impact and is equivalent to the output of the traditional FAIR model [43] extended to settings where the network connection between cyber-risk factors is explicitly accounted for, and calculated apriori over a family of APT attacks. *In other words, our research extends the traditional network oblivious FAIR model to a network aware FAIR model* and applied to an IIoT network (see Section 3).

(3) Evaluating the expectation moment of time-aggregate adverse impact distribution in an enterprise IIoT network is computationally intractable for C-suites (e.g., a CISO) and cyber-risk managers - requiring the solution of an exponential number of ordinary differential equations (ODEs). However, the mean statistic of this distribution is a bare necessity for a cyber-risk manager. We mitigate this intractability challenge by enabling cyber-risk managers to construct a simple but accurate-enough and tractable approximation of the mean value of total network-wide adverse impact distribution due to an APT cyber-attack. We use the first-order mean-field approximation (MFA) procedure to obtain such an estimate (see Section 4).

(4) We derive tight error bounds of empirical Conditional-Value-at-Risk (CVaR), i.e., a measure of APT risk, with respect to the true theoretical CVaR estimates of the time and space aggregate adverse impact distribution in an IIoT network. More specifically, the CVaR is a measure of worst-case cyber-risk and represents the tail of the time-space aggregate cyber-risk distribution. It is often the most important industry-popular risk assessment metric to risk managers after the statistical mean.

Since cyber-risk managers will only have access to empirical estimates of CVaR from loss impact samples collected over time, the former's accuracy is of paramount importance to their business. To this end, we conduct the derivation of error bounds of CVaR estimated empirically vs the ground truth using a rigorous analysis based upon the *theory of large deviations* (TLD) in probability theory. We derive tight upper and lower bound of the CVaR (our measure of worst case APT risk) estimation error using properties of the *Chernoff-Hoeffding* and the *McDiarmid* concentration inequalities from TLD, respectively.
The analysis novelty is in the tight empirical estimation of error bounds obtained in theory as a function of the finite number of empirical data samples of the cyber-loss impact distribution that a cyber-risk manager might have practical access to. The theory also states the threshold number of empirical samples from the loss impact distribution a cyber-risk manager (e.g., insurer) should demand of the IIoT network manager to satisfy its estimation error tolerance (see Section 5).

(5) An important question business managers might ask is: *do the theories that (a) ensure the stability of the APT infection spread process, and (b) promise accurate-enough estimation of the statistical mean value of the space-time adverse impact in the IIoT network, work in practice?*
We run real-world IoT testbed experiments conducted in the (Future Internet Testing) FIT-IoT Lab to validate that (a) the model parameter space under which theoretical results derived in Section 2 hold, and (b) accurate-enough MFA estimates of the true mean of the adverse impact distribution, are indeed realizable in practice for IIoT networks. Large-scale Monte Carlo simulations conducted atop the FIT-IoT Lab experimental setup indicate that first-party cyber-losses within an enterprise IIoT network (with monitoring capabilities) due to an APT cyber-breach are most likely light-tailed (see Section 6).

(6) In order to bring out the real-world practical relevance of our research, we provide an ICS case study illustrating the applicability of the *Pipedream* APT cyber-attack on cascading cyber-impact inside an organizational IIoT network.
To showcase the society-facing impact of such cyber-breaches, we then formalize and analyze the aggregate cyber-risk such intra-organization breaches might inflict on a networked society of diverse interdependent IIoT industries/enterprises in supply-chain environments. We show via theory that despite intra-organization breaches inducing a light-tailed first-party cyber-loss distribution on an organization, aggregate (first and/or third-party) losses incurred by the same in supply chain service ecosystems could be heavy-tailed. This will pose significant market scale-up challenges (as already evident from current market valuation data) to cyber-security improving commercial stand-alone cyber-insurance businesses.
We subsequently propose managerial action items to mitigate the first-party cyber-risk exposure emanating from IIoT driven businesses that further mitigates supply chain induced aggregate multi-party cyber-risk exposure (see Sections 7, 8, and 9).

The Generality of our Methodology - We emphasize that our contributions are general enough to be applied to non-IoT (wireless) communication networks that do not face the Internet. However, loss measures (e.g, revenue loss due to business disruption, total business/service downtime, i.e, unavailability) directly derived from APT malware induced and space-time dependent adverse impact on system and/or device performance (as modeled and discussed in Section 3), is mostly a characteristic of modern day society-serving IIoT networks such as industrial control systems that usually face the Internet. Throughout the paper, we use the following event terms inter-changeably: cyber-attack, cyber-breach, cyber-incident, signifying adversaries successfully bypassing IIoT network defense and causing an adverse impact with financial consequences for an enterprise.

1.3 Contribution Novelty and Their Implications to Cyber-Risk Management

Our proposed research is technically novel and/or different than existing related research in the following five significant ways - each with significant implications to cyber-risk management.

(1) Our proposed model (in Section 2) overcomes modeling shortcomings of a huge literature on *attack-defense only* type models (see Section 10). These existing models, unlike our *attack-defense-impact* type in this work, *omit* quantifying any *non-binary adverse impact measure* (e.g., cost incurred to company per unit of downtime multiplied by the total downtime) on organizational network assets/sub-systems while modeling cyber-malware spread as a time-dependent continuous stochastic process. As is often said in the enterprise management circles: *if you cannot quantify loss impact, you cannot manage it.*

(2) We are the first (to the best of knowledge) to overcome (in Section 4) the computational intractability challenge that is characteristic of ODE-based statistical mean field models trying to output the statistical mean of a performance measure (e.g, in our work - the total time-aggregate network-wide adverse impact distribution generated by an spread-based APT malware cyber-attack).

We propose a tractable approximation method (i.e., algorithm) to tightly approximate the mean of our time-aggregate adverse impact measure for both light-tailed and heavy-tailed adverse impact distributions, and empirically validate the accuracy. *The statistical mean of a loss-impact distribution is a basic operational unit of cyber-risk management.*

(3) Using concentration inequalities from the theory of large deviation in probability theory, we are the first to provide mathematical guarantees (in Section 5) on the accuracy of empirically obtained worst-case tail-risk (measured via CVaR) induced by a time-aggregate network-wide APT-generated adverse impact distribution, with respect to its non-empirical ground truth value of the relevant CVaR measure. *In practice, cyber-risk managers will only have empirically obtained samples of network and time aggregate cyber-loss post a cyber-attack event to estimate CVaR.* While some existing literature on cyber-risk characterization (but not in networked settings like ours) provides model-based estimates of the CVaR - they make no effort to characterize the accuracy between model and empirical estimates, leave alone providing mathematical guarantees of the accuracy measure.

(4) We are the first to formally analyze (in Section 7) the role of a network of IIoT networks (e.g., an inter-connected society of ICSs) on the sustainability of cyber (re-)insurance solutions commercially managing accumulative tail risk. Such a risk extent is derived from an aggregate (sourced out of multiple IIoT networks) of first/third-party IIoT cyber-risks - a characteristic of service supply chain ecosystems, when the latter are impacted due to the cascading impact of spread-based APT malware cyber-attacks on certain IIoT networks.

While some existing literature has studied the problem of sustainably managing aggregate cyber-risk in inter-dependent service network settings, they have not explicitly modeled the topological structure of underlying network of networks.

(5) Our proposed research is different and orthogonal to all existing research focused on analyzing epidemic cyber-risk spreading and recovery models in sensor-driven communication networks (see Section 10 for details). Unlike these works, our goal is not to analyze and compare which spreading and recovery model will result in better system (network) performance. Moreover, these works do not comment of system loss impact.

In contrast, our main goal is to first fix a particular APT malware spread and recovery model that is practically relevant (e.g., SIS in our case). Now having done that, we extend all related existing research by proposing the first mathematical framework with provable performance guarantees to *apriori* evaluate, compute, and tail-bound for enterprise managers - the cyber-risk statistics of space-time aggregated cyber-loss impact sampled over a family of malware spread based APT cyber-attacks. *This task will aid managerial cyber-protection budget planning to boost enterprise cyber-resilience in the future event of a cyber-attack.*

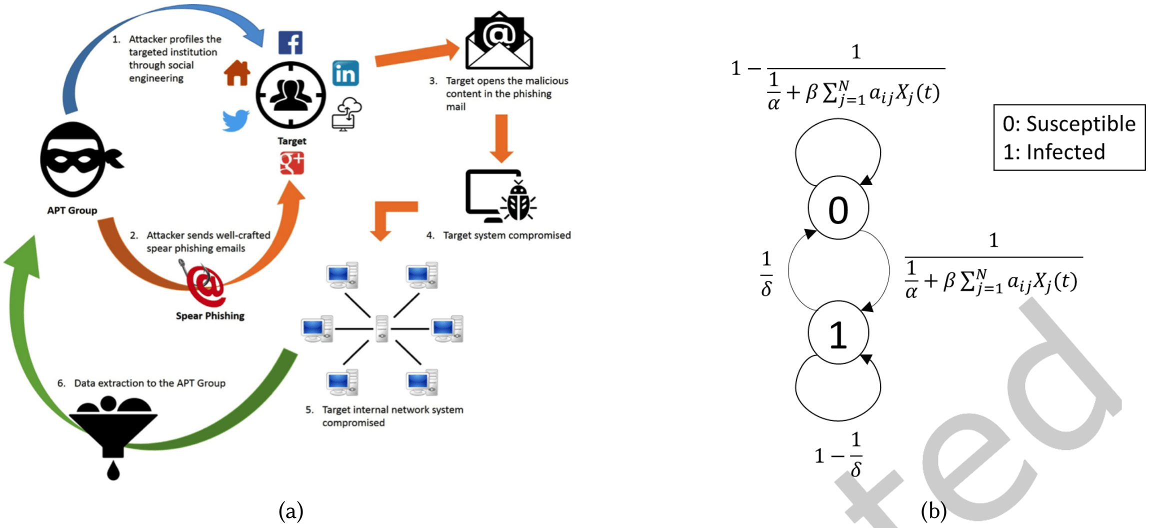

1. Attacker profiles the targeted institution through social engineering
Target
APT Group
2. Attacker sends well-crafted spear phishing emails
Spear Phishing
3. Target opens the malicious content in the phishing mail
4. Target system compromised
5. Target internal network system compromised
6. Data extraction to the APT Group
(a)
$1 - \frac{1}{\frac{1}{\alpha} + \beta \sum_{j=1}^{N} a_{ij} X_j(t)}$
0: Susceptible
1: Infected
0
$\frac{1}{\delta}$
$\frac{1}{\frac{1}{\alpha} + \beta \sum_{j=1}^{N} a_{ij} X_j(t)}$
1
$1 - \frac{1}{\delta}$
(b)

Fig. 2. An Illustration of (a) Steps Associated with a General APT Cyber-Attack, and (b) the Markov Chain (state transitions) of an Infected/Susceptible IIoT Node

2 THE APT MALWARE SPREAD PROCESS SPECIFICS

In this section, we propose models to (a) capture stealthy malware spread dynamics in an IIoT network characterized by an APT attack, and (b) formulate the node and time aggregate adverse impact of the spread in the IIoT network. Henceforth, we synonymize the term 'adverse' with 'cyber-loss' throughout the paper. *Wherever applicable, we complement modeling elements with real-world parallels borrowed from popular cyber-attacks conducted on IIoT-driven organizations.* Table I showcases important notations frequently used in this paper.

2.1 The Spread Model

Network Model - We consider an IIoT network of IoT devices labeled $1, \cdots, N$ on a simple unweighted bidirectional graph. Each IoT device (inside an IoT-driven ICS subnet and/or across subnets) is capable of getting infected (a) *indirectly*, i.e., by malware transmission (e.g., via emails, AUTORUN, open port exploitation through message broadcasting) from neighboring infected nodes (e.g., post the event when a DMZ inside an ICS is breached), and (b) *directly*, for example via it downloading malicious code from the Internet or the code being injected on them via a backdoor (e.g., the event when the DMZ is breached due to social engineering attacks or due to infected plug-able external devices). The network is represented by a symmetric adjacency matrix $A \in \{0, 1\}^{N \times N}$, with $a_{ii} = 0$ for all i, and $a_{ij} = 1$ indicates a connection between network nodes i and j, and $a_{ij} = 0$ indicates otherwise.

As examples[1] of a direct cyber-infection from popular cyber-attacks, we have (a) phishing-driven *BlackEnergy3* malware infecting CPS components (acting as nodes) of the Ukraine power grid (in 2015) via which login credentials for these components were obtained by hackers, (b) camera software vulnerabilities exploited by hackers to get entry into computers (both cameras and the computers acting as nodes) of the SCADA systems of a Turkish oil pipeline (in 2008), and (c) the *Stuxnet* worm utilizing four zero-day exploits to infiltrate the Supervisory Control and Data Acquisition (SCADA) systems controlling uranium centrifuges (acting as nodes among other CPS components that include PLC-controlled variable frequency drives (VFDs) and print spoolers)

[1] As mentioned in Section I, our examples are not restricted to critical infrastructure networks.

Table 1. Table of Important Notations for Sections 2-4

N	number of nodes in the IIoT network
deg(v)	degree of node v
X_i	susceptible/infected state of node i
α	node probability of direct infection
β	rate of a node getting infected indirectly
δ	rate of a node getting cured
γ	Pr[susceptible v getting infected by neighbor u]
A	adjacency matrix of the IIoT network graph G
$i_v(t)$	probability of node v being infected at time t
i_v^*	$\lim_{t\to\infty} i_v(t)$
$\underline{\lim}_{t\to\infty} i_v(t)$	lower bound of i_v^*
X	N-dimensional APT infection stochastic process
$\overline{\lim}_{t\to\infty} i_v(t)$	upper bound of i_v^*
L	N-dimensional adverse impact stochastic process
M	counting process of number of infected nodes
$\lambda(t)$	stochastic intensity of M
$\lambda_{i,A}$	i-th eigen value of matrix A
$f(x_1, \ldots, x_N)$	node-aggregate adverse impact function

in an Iranian nuclear plant. As examples of corresponding indirect cyber-infection (post the direct infection) from the above-mentioned cyber-attacks, we have (a) hackers opening switches that distribute power to the Ukrainian power grid and overwriting switch-controlling firmware controlling serial-to-ethernet controllers, (b) causing the Turkish oil pipeline to become over-pressurized via control commands on the IoT-controlled SCADA computers, and (c) causing the uranium centrifuges to slow up and down, crossing through mechanical resonances, till their failure via compromised PLC controller controlled VFDs. More generally, the direct-indirect nature of cyber-attacks on ICSs have been studies in [47, 56, 67, 105, 135].

Threat Model - We consider cyber-threats that are representative of the malware-induced advanced persistent threat (APT) family (e.g, *WannaCry, NotPetya*) popularly affecting many IIoT networks today. The *initial stage* of an APT, i.e., the spread of cyber-infection (malware such as the *BlackEnergy3*) through an IIoT network (e.g., by open port scanning of vulnerable IoT devices) post initial malicious code injection on a set of devices (see Figure 2a.), is (often) dynamically modelled using the seminal susceptible-infected-susceptible (SIS) methodology [95]. WLOG, we adopt the SIS model in our work. The rationale being that (a) it is intractable to plug all security deficiencies in a computer device, leave alone a system of devices [97], and (b) consequently, *IoT device i in the network is never immune*, i.e., *always eventually susceptible to infection in the cyber-world, despite measures taken via technology and/or human efforts to repair it post attack or prevent it from being attacked* [6]. The latter point is because IIoT network security is primarily about the use of IDSs and security-monitoring driven alarm systems that are merely detective measures, and not preventive measures (though all existing preventive measures are inevitably imperfect).

2.2 The Spread Dynamics

State Evolutions - The state of a node i at time t is denoted by $X_i(t)$, where $X_i(t) = 1$ indicates that i is infected at time t, and $X_i(t) = 0$ indicates that it is susceptible. Each node can be infected by its neighbors but is cured independently of all other nodes in the network[40]. Each node in the IoT network is endowed with an independent exponential clock and changes its states when the exponential clock rings. The rate of state changes by node i is given as a Markov chain (see Figure 2b.):

$$\begin{aligned} X_i : 0 \longrightarrow 1, &\quad \text{with rate } \left(\frac{1}{\alpha} + \beta \sum_{j=1}^{N} a_{ij} X_j(t)\right) \\ X_i : 1 \longrightarrow 0, &\quad \text{with rate } \delta, \end{aligned} \tag{1}$$

for $\alpha, \beta, \delta > 0$. Here (a) α is the probability that a node becomes infected directly, e.g., by downloading malicious code from the Internet [127]; (b) $\beta = \frac{1}{\gamma i_u(t)}$, where γ is the probability of a susceptible node v being infected by an infected neighbor u, i.e., $a_{uv} = 1$; $i_v(t)$ is the probability that node v is infected at time t; and (c) δ is the rate at which a node becomes susceptible from the infected state [129] (due to detection and response activities). The state evolution logic adheres to reality that IDS and security-monitoring alarm-triggered incident response are imperfect, and takes time within which cyber-adversaries take network control.

We assume, for the main purpose of analytical tractability that α, β, γ, and δ are uniform for all the nodes in the network (similar to many of our predecessors modeling malware spread in theory). *However, we relax this assumption in our simulation exercise in Section 6.* Having uniform values of $\alpha, \beta, \gamma, \delta$ across all nodes do not affect the insights obtained via theory - only the scale of parameter and quantitative expression values differ between uniform and non-uniform settings.

How Much Do the Managers Need to Estimate the 'Greeks'? - We emphasize that a network manager only needs to know the range (upper and lower bound) of values for $\alpha, \beta, \gamma, \delta$ to run large scale simulations that would provide it with *apriori* estimates of the extent of first-party cyber-risk within an IIoT network. The range reflects the set of feasible values for the 'greek' parameters. *After all, the manager planning ahead to boost enterprise cyber-resilience cannot afford to wait for the cyber-attack to happen to estimate and work with real-time values of β, γ, δ that can provide it with high-accuracy loss-impact estimates post incident.* The manager can simply conduct a simulation exercise much ahead of an incident by sampling the 'greeks' from within an interval and generate apriori loss-impact distributions to aid its cyber-protection budget planning tasks. The range of β, γ, δ can be estimated by a manager from a dataset of historical internal measurements and observations through network telemetry using market products such as *Wyebot* telemetry solutions. However, unlike β, γ, δ, it is relatively much difficult to have access to the range of α. Having a proper estimate of α would imply that the defender has perfect knowledge of the ability of the attacker - something not true in practice. Hence, the best a manager could do to work with α for generating apriori cyber-risk statistics is to assume bounds obtained from historical data. However, it is possible that the upper bound of α obtained from historical data is lower than that induced by an attacker with significantly higher potency than observed historically. To account for such attackers in simulation studies, the upper bound of α during such studies should be made higher than that observed from historically obtained data.

The Underlying Stochastic Process - We first provide an intuition, based on the work by [40], for the general audience of the N-dimensional Markov process X stitched out by the above-mentioned one-dimensional Markov chain, where N is the number of nodes in the IIoT network. More specifically, at any time instant t, $X(t)$ is the vector of random variables $X_i(t)$'s - capturing the I/S state of each node in the IIoT network. Hence, $X(t)$ is an N-dimensional random function over time, where each random instance of the single-dimensional $X_i(t)$ over t evolves according to one-dimensional Markov chain X_i. Geometrically, at each time instant t, the N-dimensional function represents a random (I/S) configuration of the entire IIoT network. Each such random instance of the N-dimensional function $X(t)$ evolves according to a special N-dimensional Markov chain (contributing to special stochastic process known as a Feller process [101]) that is decomposable using the one-dimensional Markov chains X_i.

This intuition can be formally represented as follows. Let $(\Omega, \mathcal{F}, P)$ be a probability space with filtration $\mathbf{F}$ = $\mathcal{F}_t | t \geq 0$, where $\mathbf{F}$ is right continuous and $\mathcal{F}_0$ contains all the P-null sets [98]. Ω is the sample space, and P

is the probability function mapping each event in F to $[0,1]$. Intuitively, $\mathcal{F}_t$ can be thought of as the family of all N-dimensional random functions $X(t)$ charted out till time t, with $\mathcal{F}_0$ consisting of all functions whose probability measure of occurrence is zero (e.g., a point function in N dimensions, where the points are random initial (I/S) configurations of the N-node IIoT network). The process X is a Markov process with state space $E = \{0,1\}^N$ with $X_0 = x \in E$. We assume that X is a Feller process with generator $G : C(E) \rightarrow \mathbf{R}$, notated by $G(f(x))_{|f \in C(E), x \in E}$ and expressed as:

$$\sum_{i=1}^{N}\left((1-x_i)\frac{1}{\alpha}+\beta(1-x_i)\sum_{j=1}^{N}a_{ij}x_j+x_i\delta\right)(f(x^i)-f(x)),$$

where state transitions $x^i_j = x_j$ for $i \neq j$ and $x^i_i = 1 - x_i$.

The family $C(E)$ consists of all cadlag functions on E (these are N-dimensional random functions admitting jumps, i.e., discontinuities, in the stochastic process $\{X_t\}$), but form a Skorohod space on which probability measures are always defined [62]. The concept of the jump is relevant in the N-dimensional setting because the (I/S) network configuration changes over time may have abrupt changes. The cadlag functions lying on the Skorohod space just enables us to be able to quantify the probability mass of a collection of random functions. The Markov process X has exponential waiting times between jumps and an exponential state space of cardinality 2^N.

Why Does the MF Process Cover the Entire Malware Spread Space? - A thing of importance to note is that for any fixed $\alpha, \beta, \gamma, \delta$ instance, a (APT) malware can take numerous attack paths inside the IIoT network. This is simply because an attack spread can start at any point within the network, and there are possibly an exponential number of spread paths in a connected network from a source to a destination(s). Each attack path will result in a binary vector $X(t) = [X_1(t), \ldots, X_N(t)]$ of I/S states at any given time instant t with the dynamics of vector $X(t)$ evolving with t. In other words, each attack path over time is a dynamically changing N-dimensional vector of I/S states. Now each attack path evolving over time is also a *single sample path* of the MF process. Note that any stochastic process by mathematical definition is a large collection of sample paths (attack paths) that captures all possible (possibly exponential in number) variations over time in the states of a dynamically evolving system modeled through the stochastic process. In our paper, the system is the combination of IIoT network with embedded I/S node states. Each attack path then represents the infection dynamics of nodes in the IIoT network over time, with multiple paths reflecting a variation in the system dynamics for any fixed $\alpha, \beta, \gamma, \delta$ instance. We selected the MF process (due to the above-mentioned attractive mathematical properties of C(E)) to capture the infection dynamics of all possible attack paths given any fixed $\alpha, \beta, \gamma, \delta$ instance. *Hence, the MF process successfully covers the entire malware spread attack path space in an N-node IIoT network for any fixed $\alpha, \beta, \gamma, \delta$ instance.*

2.3 The Adverse (Cyber-Loss) Impact Model

Adverse Impact Generating Stochastic Processes - The *end goal* of an APT is to cause system-wide damage (e.g., revenue loss via business disruption) in the IIoT network, after it has stealthily infected (a subset of) nodes in its initial stage, where infection does not imply node/device damage that is left for later. More specifically, at certain times post its infection stage, the APT attack decides to launch damage on certain infected nodes making them dysfunctional (incapable of providing QoS). These times, unknown to a a system manager, are denoted by $(T_n)_{n \in \mathbf{N}}$. The corresponding number of cumulative damage launches by these time instants are counted through by a stochastic process $M = (M(t))_{t \geq 0}$ (from [40]). The size of the negative/adverse impact on infected nodes becoming dysfunctional at any time instant t is modeled by another N-dimensional stochastic process $L = (L(t))_{t \geq 0}$, where $L(t) = (L_1(t), \ldots, L_N(t))^{\mathrm{T}}$, where each dimension represents the adverse impact on a given node in the IIoT network. The value $L_i(t)$ for node i is zero if it is either not infected at time t or the APT attack does not launch damage on i at t.

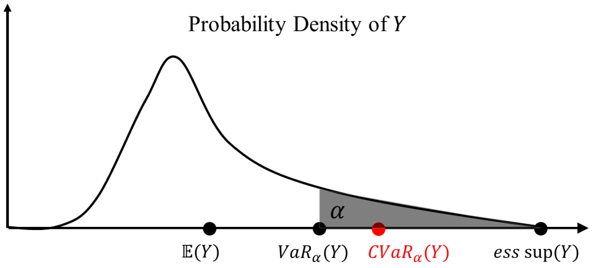


Fig. 3. Conditional Value-at-Risk (CVaR) is a measure of one-sided tail risk of the random variable $Y = \int_0^T f(t; L(t) \circ X(t))\lambda(t)dt$. For this continuous bounded random variable Y representing a cost and $\alpha \in (0, 1]$, we illustrate $\text{CVaR}_\alpha(Y)$ as the APT risk measure. The area of the shaded region is α. The expectation of Y, the Value-at-Risk of Y at level α (the smallest cost in the $\alpha \cdot 100\%$ worst cases), and the essential supremum of Y are also shown.

Note that M is a left-continuous counting stochastic process with a deterministic and non-deterministic component (according to the seminal Doob-Meyer decomposition theorem [21]), and adapted to the filtration $\mathbf{F}$, i.e., random functions spanned upto $M(t)$ is inside the family of functions $\mathbf{F}_t$. One of the components is its stochastic intensity $(\lambda(t))_{t\geq 0}$, where λ is a non-negative $\mathbf{F}$-predictable process (because it is a continuous time-adapted process that is left-continuous). The other component, $M(t \wedge T_n) - \int_0^{t\wedge T_n} \lambda(s)ds$, is the martingale, i.e., the non-deterministic component for all $n \in \mathbf{N}$. The martingale property ensures that the conditional expectation, $M'(t + \delta t) - \mathbb{E}[M'(t + \delta t)]|\mathbf{F}_t$, equals $M'(t)$ for a stochastic process $M'(t)$. In our setting, since M is a counting process, it is a sub-martingale by the Doob-Meyer decomposition theorem. In practical jargon, the sub-martingale property simply and evidently implies that the expected number of launched damages by the APT at time $t + \Delta t$ given time history is greater than the number of launched damages by time t. L is predictable (again due to it being a time-adapted process that is left-continuous) and non-negative. Both M and L are independent from the Markovian infection spread process X.

Network and Time Aggregate Adverse Impact - *We (e.g., cyber-loss managers) are concerned with the expected network aggregate adverse impact of the APT attack over a fixed time window* $[0, T]$ *with* $T > 0$, where we assume that the tangible impact units are the same for all the nodes. In practice, such units can be mapped to the loss in quality of experience (QoE) derived from dysfunctional nodes (e.g., dysfunctional IoT-embedded CNC industrial machines) [109][73][106]. Consider a measurable function $f(\cdot;\cdot) : \mathbf{R}_+ \times \mathbf{R}_+^N \to \mathbf{R}_+$, where the first argument refers to the time, and the second - to the N-dimensional (for N nodes) adverse impact (i.e., FAIR for networks) generated by an APT cyber-attack. The *expected* aggregate impact (adapted from [40]) incurred over the period $[0, T]$, given by $\mathbb{E}\Big[\int_0^T f(t; L(t) \circ X(t))dM(t)\Big]$, obey the following equivalent equalities:

$$\mathbb{E}\Big[\int_0^T f(t; L(t) \circ X(t))dM(t)\Big] = \mathbb{E}\Big[\int_0^T f(t; L(t) \circ X(t-))dM(t)\Big],$$

$$\mathbb{E}\Big[\int_0^T f(t; L(t) \circ X(t))dM(t)\Big] = \mathbb{E}\Big[\int_0^T f(t; L(t) \circ X(t-))\lambda(t)dt\Big],$$

$$\mathbb{E}\Big[\int_0^T f(t; L(t) \circ X(t))dM(t)\Big] = \mathbb{E}\Big[\int_0^T f(t; L(t) \circ X(t))\lambda(t)dt.\Big]$$

Here, the $\circ$ operator denotes the component-wise Hadamard product of vectors. The first equality is due to the fact that X and M are independent and never jump at the same time with probability 1 (in practical terms an IIoT node is not infected and launched an attack upon at the same time). The second equality follows from the

F-predictability of the integrand [21] (because M is left-continuous), and the third equality holds since the paths of X, i.e., the N-dimensional random functions, possess at most countably many jumps in $[0, T]$ and constitute a Lebesgue null set, i.e., a set with a probability measure zero, for each path (because in practice there are countably many malware spread paths and the probability of the individual occurrence of each is zero amongst an infinite continuum of possible paths). *Specific practical forms of $f(\cdot)$ will be discussed in Section 6.* Of equal importance as the mean, are the *tail properties* of the distribution $\int_0^T f(t; L(t) \circ X(t))\lambda(t)dt$ that reflects the statistical spread of the adverse impact (risk) distribution (the probability that network aggregate adverse impacts of an APT attack exceeds a certain percentile). The spread metric will characterize the notion of APT risk that we model in our work with the widely popular Conditional-Value-at-Risk (CVaR) risk measure of $\int_0^T f(t; L(t) \circ X(t))\lambda(t)dt$ (see Figure 3 for an illustration) from risk theory [70]. A detailed analysis of this CVaR metric characterizing APT risk is deferred till Section 5.

One could argue the feasibility of quantifying the adverse impact over time and space if the attack type and parameter space is (partially) unknown. *However, the aim of this paper is not to quantify this impact post an (APT) attack event. We want to quantify (for cyber-protection budget planning CISOs and CEOs) how much cyber-risk (induced by statistics of the adverse impact) an ICS can potentially be subject to if it were to be affected by a family (characterized by all feasible $\alpha, \beta, \gamma, \delta, t$) of malware-spreading APT cyber-attacks.*

To complement our contribution rationale, note that it is usual in practice for managers to find it difficult to estimate the hypothetical impact of an adverse cyber-incident [24]. This difficulty is aggravated via uncertainties in the knowledge of the cyber-risk terrain, system complexity, lack of cyber-incident data and cyber-loss impact metrics, and the inability to predict future cyber-incidents [47, 55]. Add to this is the role of cognitive biases that prevent even the most experienced of system managers to assess the impact of cyber-risk accurately enough [66, 108, 111]. In such environments, it is best that managers consider worst case cyber-risk impact (the analysis of which needs to consider a family of parameters) into account rather than work with a faulty perception of exact cyber-risk impact.

3 ON ACCURATELY ASSESSING APT ATTACK IMPACT OVER ‘SPACE’ AND TIME

In this section, we provide a closed form analysis of the (M)arkov-(F)eller malware-spread process in an APT compromised IIoT network towards enabling a cyber-risk manager to accurately assess the network-wide and time-aggregate attack impact. More specifically, we first study and derive conditions under which the spread will become stable. Once this condition is achieved, we then derive a closed form expression for the network aggregate adverse impact, for a pre-specified time period of duration T. Given that IoT device dysfunctions contribute to loss in QoS/E, *we associate the term ‘adverse impact’ with the term ‘cyber-loss impact’ throughput the rest of the paper.*

3.1 Will the APT Malware Spread Process Stabilize Over Time?

An important first question any cyber-risk manager would be interested to know is: *will the APT malware spread process stabilize over time, and in what manner?* More specifically, the question simplifies to: *will the infection spread die out, and if and when it does how many IIoT network nodes will be compromised for the risk manager to ideate a coverage budget?* The stability condition for the Markov-Feller malware-spread process is mathematically defined via the condition that $\lim_{t\to\infty} i_v(t) = i_v^*$ holds for $v = 1, \ldots, N$, regardless of the number of initially malware-infected nodes in an IIoT network. Here, $i_v(t)$ is the probability that network node v is infected at time t. Intuitively, this implies that the fraction of the network nodes in an infected state at any time t is (asymptotically) constant over time. The following result, spinning from [127], gives a sufficient condition for the malware-spread process to be asymptotically stable.

Theorem 3.1. *Let $m = \max_{v \in V=\{1,\ldots,N\}} \deg(v)$. Also let $\lambda_{1,A}, \ldots, \lambda_{N,A}$ be the eigenvalues of the adjacency matrix A of IIoT network G with $\lambda_{1,A} \geq \ldots \lambda_{N,A}$ (in modulus). In the case when $\frac{1}{\delta} < (1-\alpha)\frac{1+(1-\gamma)^m}{2}$, if $\lambda_{1,\text{ A}} < \frac{\alpha + \frac{1}{\delta}}{\gamma(1-\alpha)}$. then the APT malware-spread in the IIoT network will become asymptotically stable (i.e., will have a fixed number of nodes in (I/S) states) regardless of the initial number of infected nodes. In the case when $\frac{1}{\delta} \geq (1-\alpha)\frac{1+(1-\gamma)^m}{2}$, if $\lambda_{1,\text{ A}} < \frac{1-\frac{1}{\delta}+(1-\alpha)(1-\gamma)^m}{\gamma(1-\alpha)}$, then the malware-spread process will become asymptotically stable irrespective of the initial number of infected nodes. Moreover, let $\overline{\lim}_{t\to\infty} i_v(t)$ denote the upper bound of the limit of $i_v(t)$, and $\underline{\lim}_{t\to\infty} i_v(t)$ denote the lower bound of the limit of $i_v(t)$.*
Then, we have $\overline{\lim}_{t\to\infty} i_v(t) \leq \theta_v^+$ and $\underline{\lim}_{t\to\infty} i_v(t) \geq \theta_v^-$,

$$\theta_v^+ = \frac{1-(1-\alpha)(1-\gamma)^{\deg(v)}}{\min\left\{1+\frac{1}{\delta}-(1-\alpha)(1-\gamma)^{\deg(v)},1\right\}}$$

$$\theta_v^- = \begin{cases} \dfrac{1-(1-\alpha)(1-\gamma\nu)^{\deg(v)}|_{\{(1-\alpha)(1-\gamma\nu)^{\deg(v)} \geq \frac{1}{\delta}\}}}{1+\frac{1}{\delta}-(1-\alpha)(1-\gamma\nu)^{\deg(v)}} \\ \left(\kappa - \frac{1}{\delta}\right)\theta_v^+ + 1 - \kappa|_{\kappa - \frac{1}{\delta}} < 0 \end{cases}$$

with, $\nu = \min\{1-\frac{1}{\delta}, \alpha\}$ $\kappa = (1-\alpha)(1-\gamma\nu)^{\deg(v)}$.

Implications for Cyber-Risk Management - The theorem states that for any general IIoT network, irrespective of the number of initially APT infected nodes, *malware spread will eventually converge (under weak assumptions) to an equilibrium state where infected nodes necessarily do not die out - a practically reality in all IT (networked) systems. In other words, a certain fraction (less than 1) of nodes (in absolute number and not in identity) will always be infected (as is usual practice due to imperfect security technology*) at any point in time. To make things more clear, we can never have a situation where all network nodes will be in an ideal susceptible state. i.e., the infection dies out. As malware spreads and susceptible nodes get infected, some nodes that had been infected at a prior time step might have recovered (due to detection and response activities) to the susceptible state, while some will still remain in an infected state with some getting newly infected, and this behavior will be invariant over time. Hence, at any point in time there will be some nodes (not a constant number but one that is increasing and decreasing over time) in the network that will be in an infected state while the others will be in a susceptible state.

We also observe that when the infection rate ≪ 1, the lower bound of the limit $i_v(t)$ is tighter; on the other hand when the infection rate is near 1, the upper bound is tighter. Since the infection spread process is not necessarily always convergent to a static limit value of $i_v(t)$, it is challenging to quantify the exact thresholds at which such behavior is observed. In such scenarios, characterizing upper and lower bounds of this variable will provide cyber-risk managers with precise-enough estimates of an equilibrium value of $i_v(t)$. *As a practical takeaway, the bound variables are clearly a function of the node vulnerability, that can be mitigated through investing in strong organizational cyber-security practices by the IIoT managers.* This will further tighten the bounds.

3.2 A Cyber-Risk Manager's Closed-Form Expression for Aggregate Cyber-Loss

A closed form expression is often a necessary first step for cyber-risk managers to accurately estimate space-time aggregate cyber-loss in any part of the IIoT network. To this end, we first propose a framework, adapted from [40], to derive a closed form expression for the node-aggregate cyber-loss impact, $\int_0^T f(t; L(t) \circ X(t))\lambda(t)dt$, in a given time period for an APT compromised IIoT network. We will first propose a method for a cyber-risk manager to accurately approximate any general f in closed form as a polynomial function for analysis tractability (for ease of taking an integral), and follow that up with theoretically bounding the function approximation error.
Assumptions - Though X is a Markov process with cadlag paths, we assume that all $f \in C(E)$ are continuous functions over time, i.e., $f(t; L(t) \circ X(t)$ is continuous. However, *it is a challenging problem to compute aggregate (over time and space) loss values directly on top of general continuous loss functions.* Thankfully, the applicability

of the *Stone-Weierstrass* theorem [22] allows us to work with polynomial loss functions that are outcomes of uniformly approximating arbitrary continuous functions defined over a compact (closed and bounded) set in $\mathbf{R}^n$ - as in our case, with the Hadamard product lying in a compact set. We also assume that time-aggregated loss function f will map into a one-dimensional tangible scalar value, as is usual in practice. For a multi-variate f, we convert its range to a single dimensional output, i.e., $f : \mathbb{R}_+^N \to \mathbb{R}_+$, via the transformation:

$$f\left(x_1, \ldots, x_N\right) = g\left(\Lambda\left(x_1, \ldots, x_N\right)\right),$$

where we assume $\Lambda : \mathbb{R}_+^N \to \mathbb{R}_+$ to be a linear increasing aggregation function, and $g : \mathbb{R}_+ \to \mathbb{R}_+$ to be continuous and increasing. g is also assumed to be bounded on $[0, \|\Lambda(L)\|_\infty)$, where $\|\cdot\|_\infty$ denotes the L^∞ norm. An example of Λ is $\Lambda(x_1, \ldots, x_N) = \sum_{i=1}^N \alpha_1 x_i$, $\alpha_i \geq 0$. g, for example, could be of the form $g\left(\Lambda\left(x_1, \ldots, x_N\right)\right) = \Lambda\left(x_1, \ldots, x_N\right)$.
Polynomial Approximation of a Continuous f - As mentioned above, it can be quite cumbersome for a cyber-risk manager to accurately evaluate any general f (and consequently the aggregate space-time cyber-loss within an IIoT network) for a given IIoT network. It would be good if any given f could be approximated through a polynomial function that is much amenable to computational and analysis ease. In view of recent developments in function approximation theory following [30], a polynomial closed form approximation to a general f can indeed be constructed via the following steps. *This approximation to f is the first extension of the traditional FAIR metric [43] (that does not explicitly model the underlying infrastructure network structure) for evaluating cyber-loss through a closed form expression, applied to networked settings.*

(1) Choose (a) $d \in \mathbb{N}$ - a pre-specified choice for the degree of the polynomial, and (b) a bound $\varepsilon > 0$.
(2) Select a constant $u \in \mathbb{R}_+$ (based on prior network and cyber-loss impact knowledge) such that the node-aggregated cyber-loss impact is bounded as per the following relation:

$$\mathbb{P}(\Lambda(L) > u) \leq \epsilon.$$

(3) From the space of all degree-d polynomials, choose the best uniform approximation $p_d(x) := \sum_{\ell=0}^d a_\ell x^\ell$ ($a_0, a_1, \ldots, a_d \in \mathbb{R}$) [30] of g on the compact interval $[0, u]$. The subsequent function approximation error is denoted by $e_d(g)$, and given as

$$e_d(g) =: \max_{x \in [0,u]} |g(x) - p_d(x)| = \|g - p_d\|_{\infty,[0,u]},$$

where the L_∞ norm is used to extract out the maximum possible error.

How Much is the Approximation Error for a Cyber-Risk Manager? - We now quantify the 'cost' (error) of the approximation borne by the cyber-risk manager as an outcome of approximating f. The optimal degree-d polynomial approximation of $f(L \circ X)$ obtained from Steps 1-3 above is given by

$$\bar{f}_d(L \circ X) := \begin{cases} p_d(\Lambda(L \circ X)), & \text{if } \Lambda(L) \leq u \\ 0, & \text{if } \Lambda(L) > u \end{cases}$$

We then have the following result following developments in [30][40] in relation to the approximation error induced by the polynomial $\bar{f}_d$.

Theorem 3.2. *The function approximation error for $f = f\left(x_1, \ldots, x_N\right)$ incurred by the cyber-risk manager is defined through the following inequation $\left\|f(Z) - \bar{f}_d(Z)\right\|_{L_1} \leq e_d(g) + m \cdot \varepsilon$, where $Z = L \circ X$, and m is a real number that satisfies*

$$|f(l \circ X)| = |g(\Lambda(L \circ X))| \leq |g(\Lambda(L))| \leq m,$$

for all possible realizations of L, where the L_1 prevents amplifying outlier effects during function approximation.

Implications for Cyber-Risk Management - The theorem implies that there always exists a closed form expression for approximate tangible space-time aggregate cyber-loss impact post an APT cyber-attack event on an ICS network guaranteed through polynomial approximation mathematics. The L_1 norm of the function approximation error induced by $\bar{f}_d$ is (a) tightly bounded from above by $e_d(g) + m \cdot \varepsilon$, and (b) is very low for small enough ϵ. From the viewpoint of cyber-risk management, the theorem results are useful as they simply denote that it is within the scope of the IIoT cyber-risk manager to influence the quality of its approximate for aggregate space-time cyber-loss by controlling its risk appetite (via ϵ).

4 THE CHALLENGE TO EVALUATE MOMENTS OF AGGREGATE CYBER-RISK

Thus far, we derived a math formula for managers that reflects the statistical mean of space-time aggregate cyber-risk within an IIoT network due to a malware driven APT cyber-attack. However, computing this mean is non-trivial due to some nuances in the above-mentioned closed form formula. In this section, we first state the mathematical nuances that make the mean computation task difficult. We then derive a mathematical framework, the outcome of which will help managers obtain a provably accurate enough approximation of the statistical mean value (see Theorem 4.1).

Nuances to Compute Statistical Mean - Once a cyber-risk manager has a tight accurate approximation of a general f an IIoT network instance may throw up, its next step is to evaluate the expectation moment of the *time integral of that f* (also known as expectation of the APT risk distribution), i.e., $\mathbb{E}\left[\int_0^T f(t; L(t) \circ X(t))\lambda(t)dt.\right]$. In other words, a cyber-risk manager wants an estimate of the much important expected value of aggregate APT-induced cyber-risk in the IIoT network.

However, this requires individually computing the moments $\mathbb{E}[X_{i_1}(t)], \mathbb{E}[X_{i_1}(t)X_{i_2}(t)], \ldots, \mathbb{E}[X_{i_1}(t) ... X_{i_N}(t)]$, that necessitates the solution of $2^N - 1$ ODEs induced by the Markov process X. This is clearly an intractable task for large N. To mitigate this challenge and ensure computational tractability, we approximate all these moments using the seminal *first-order N-intertwined mean-field approximation* technique, NIMFA, a widely popular interacting particle (node) method proposed by Van Mieghem et.al. [113].

The Mean-Field Approximation (MFA) - *The fundamental intuition behind the MFA is to split the expression* $\mathbb{E}[X_{i_1}(t) ... X_{i_N}(t)]$ *using probabilistic independence arguments*. To achieve this, the MFA seeds from the following time-variant infinitesimal dynamics, $\Delta_t(X_i(t)) \mid \mathbf{F}_t$ of $X_i(t)$ between getting infected and getting cured during time Δt. Specifically, $\Delta_t(X_i(t)) \mid \mathbf{F}_t$ equates to

$$\left((1 - X_i(t))\,\beta \sum_{j=1}^{N} a_{ij} X_j(t) - \delta X_i(t) \right) \Delta t + o(\Delta t),$$

where intuitively, dividing by Δt and taking the expectation on both $\Delta_t(X_i(t)) \mid \mathbf{F}_t$, and its expanded expression; and letting $\Delta t \longrightarrow 0$, we obtain the following exact expression for the derivative of the probability $\mathbb{E}[X_i(t)] = P(X_i(t) = 1)$:

$$\frac{d\mathbb{E}[X_i(t)]}{dt} = -\delta \mathbb{E}[X_i(t)] + \beta \sum_{j=1}^{N} a_{ij} \mathbb{E}\left[X_j(t)\right] - \beta \sum_{j=1}^{N} a_{ij} \mathbb{E}\left[X_i(t) X_j(t)\right], \tag{2}$$

for $i = 1, \ldots, N$. This follows from Kolmogorov's forward equations for Markov process X_i of IIoT network node i.

The approximation relies on finding a function $F : [0, 1] \longrightarrow [0, 1]$ that splits mixed terms of the form $\mathbb{E}\left[X_i(t)X_j(t)\right]$ as

$$\mathbb{E}\left[X_i(t)X_j(t)\right] \simeq F(\mathbb{E}[X_i(t)]) \cdot F(\mathbb{E}[X_j(t)]).$$

This split, motivated by the principle of independence of expectations, results in the approximation:

$$\frac{d\mathbb{E}[X_i(t)]}{dt} \approx -\delta \mathbb{E}[X_i(t)] + \beta \sum_{j=1}^{N} a_{ij} \mathbb{E}\left[X_j(t)\right] - \beta \sum_{j=1}^{N} a_{ij} F\left(\mathbb{E}[X_i(t)]\right) \cdot F\left(\mathbb{E}\left[X_j(t)\right]\right). \tag{3}$$

We denote by $z_i^{(1)}(t)$ the corresponding first-order approximation of $\mathbb{E}\left[X_i(t)\right]$ and get the following system of ODEs:

$$\frac{dz_i^{(1)}(t)}{dt} = -\delta z_i^{(1)}(t) + \beta \sum_{j=1}^{N} a_{ij} z_i^{(1)}(t) - \beta \sum_{j=1}^{N} a_{ij} F\left(z_i^{(1)}(t)\right) \cdot F\left(z_j^{(1)}(t)\right), \tag{4}$$

for $i = 1, \ldots, N$, (similar to [57]) solving which helps us easily compute the first-order mean-field approximation of the moment, $\mathbb{E}\left[\int_0^T f(t; L(t) \circ X(t))\lambda(t)dt.\right]$, corresponding to the mean-field function F. It has been shown in [40] that the solution to this set of ODEs always exist. Now suppose $F(x) = x$ (a popular choice from [113]), then NIMFA results in

$$\mathbb{E}\left[X_i(t)X_j(t)\right] \simeq F(\mathbb{E}[X_i(t)]) \cdot F(\mathbb{E}[X_j(t)]) = \Pi_i \mathbb{E}[X_i(t)],$$

that eases out the computation of $\mathbb{E}[X_{i_1}(t) \ldots X_{i_N}(t)]$.

Practical Importance of the Existence and Uniqueness of MFA - A topic of immense practical importance for cyber-risk management is investigating the existential and uniqueness properties of the MFA post solving (4). The existential argument is necessary for a cyber-risk manager's confidence to be able to derive an accurate-enough approximation of the mean time and network aggregate cyber-loss for any IIoT network. The uniqueness argument is necessary to provide confidence to a cyber-risk manager that the existing above-mentioned mean estimate is indeed the best possible estimate. We first have from (3), the following relation upper-bounding $\frac{d\mathbb{E}[X_i(t)]}{dt}$, given $\mathbb{E}\left[X_i(t)X_j(t)\right] = \mathbb{E}\left[X_i(t)\right]\mathbb{E}\left[X_j(t)\right] + \mathrm{Cov}\left(X_i(t), X_j(t)\right)$ with $\mathrm{Cov}\left(X_i(t), X_j(t)\right) \geq 0$:

$$\frac{d\mathbb{E}\left[X_i(t)\right]}{dt} \leq -\delta \mathbb{E}\left[X_i(t)\right] + \beta \sum_{j=1}^{N} a_{ij} \mathbb{E}\left[X_j(t)\right] - \beta \sum_{j=1}^{N} a_{ij} \mathbb{E}\left[X_i(t)\right] \mathbb{E}\left[X_j(t)\right], \tag{5}$$

for $i = 1, \ldots, N$. Setting $V := (v_1, v_2, \ldots, v_N)^\top$, where $v_i(t) := z_i^{(1)}(t)$, the system of new ODEs can be written in matrix notation (for ease of subsequent analysis) as follows:

$$\frac{d}{dt}V = (\beta A - \delta \mathbb{I})V - \beta\, \mathrm{diag}(V) AV, \tag{6}$$

where $\mathrm{diag}(V)$ denotes the diagonal matrix with entries $v_1, v_2, \ldots, v_N$ and $\mathbb{I} \in \mathbb{R}^{N \times N}$ denotes the identity matrix. We now have the following result, derived and adapted from [40], related to the *existence* and *uniqueness* of the MFA, and its accuracy with respect to the theoretical ground truth.

Theorem 4.1. *A cyber-risk manager always has access to a mean field approximation (MFA) of the expected value of time and network aggregate cyber-risk in an IIoT network. This MFA solves (6) for any choice of parameters δ and β with arbitrary non-negative initial conditions. Moreover, the MFA is unique in $C([0, \infty) : \mathbb{R}^N)$. In terms of the accuracy of the MFA estimate, let (a) $y_i(t) := \mathbb{E}\left[X_i(t)\right] - v_i(t)$ and (b) $\hat{\mu}$ the largest eigenvalue of the adjacency matrix A. Then, for any $t \geq 0$, we have (in line with [114],that focuses on the time-varying covariance matrix as a rough measure for representing MFA accuracy)*

$$\|y(t)\|^2 \leq e^{(-2\delta + 4\beta\hat{\mu} + \beta)t} \beta \int_0^t \|R(s)\|^2 ds,$$

where $y(t) = (y_1(t), .., y_N(t))$ and $R(t) = (R_1(t), .., R_N(t))$ with $R(t)$ satisfying

$$\frac{d}{dt}\mathbb{E}\left[X_i(t)\right] = -\delta \mathbb{E}\left[X_i(t)\right] + \beta\left(1 - \mathbb{E}\left[X_i(t)\right]\right) \sum_{k=1}^{N} a_{ki} \mathbb{E}\left[X_k(t)\right] - \beta R_i(t). \tag{7}$$

Table 2. Important Notations for Section 5

n	#empirical samples of cyber-loss impact
X	r.v. characterizing aggregate cyber-loss
U	maximum value of X
α	tail percentile/quantile
δ	a probability marker
ε	precision/accuracy amount
$\text{CVaR}_\alpha(X)$	CVaR for r.v. X at an α-quantile
$\widehat{\text{CVaR}}_\alpha(X)$	empirical CVaR for r.v. X at an α-quantile
$\text{APTRisk}_\alpha(X)$	$\text{CVaR}_\alpha(X)$
$\widehat{\text{APTRisk}}_\alpha(X)$	$\widehat{\text{CVaR}}_\alpha(X)$

$R(t)$ captures the time-dependent accuracy (measured as the difference between the exact and approximate dynamics of $\mathbb{E}\,[X_i(t)]$)of the instantaneous MFA estimate of mean APT cyber-risk, where the error term $R_i(t) := \sum_{k=1}^{N} a_{ki}\,\text{Cov}\,(X_i(t), X_k(t))$.

Implications for Cyber-Risk Management - The theorem guarantees the existence of a unique first-order MFA solving (6), and makes computing approximate statistics of node and time aggregate cyber-loss impact moments analytically tractable for IIoT and cyber-risk managers. Note that if $X_i(t)$, $X_k(t)$ are independent, the MFA of the instantaneous (and time-aggregate) cyber-loss impact moments are identical to the true moments. The theorem also states that with small $\frac{\beta}{\delta}$ ratios - indicating a high curing rate compared to the infection rate, the MFA accuracy is high, and the approximation error decays exponentially. *In practice, this evidently implies that cyber-risk managers (e.g., cyber-insurance agencies) will only gain value from their estimated cyber-loss impact approximations when a strong cyber-security posture exists in the IoT network.* Hence, from a policy perspective, the C-suite of IIoT-driven organizations should be made to significantly invest in strong technical and behavioral cyber-security practices.

5 ANALYZING TAIL APT RISK (CVAR) IN AN IIOT NETWORK

Thus far, we derived analytical and tractable closed-form solutions to the first moment, i.e., expectation, of the node-aggregate cyber-loss distribution in an IIoT network. While this metric is useful to a cyber-risk manager, in practice, it is interested, both, in expected cyber-loss estimates, as well as *in the knowledge of the tail, i.e., loss spread, of a cyber-loss distribution.* An industry standard cyber-risk metric to measure this tail is the Conditional-Value-at-Risk (CVaR) metric. *In our paper, we synonymously term this metric as the APT risk, and provide accuracy guarantees of empirically estimating the APT risk, when compared to the theoretical ground-truth. This is because any cyber-risk manager can only have access to discrete samples of cyber-loss to infer tail cyber-risk.*

To do so, we propose a rigorous framework based on the theory of large deviations (TLD) that (a) first uses *concentration inequalities* [20] to analyse the deviations of empirical estimates of the APT risk from the empirical mean of the cyber-loss impact, and (b) subsequently provides upper and lower bounds of the deviations of the empirical estimates of the APT risk from theoretical ground truth (see Theorems 5.3, 5.6, and 5.8). Concentration inequalities in the TLD deal with deviations of functions of independent random variables from their expectation. *A table of important notations for this section is showcased in Table II.*

5.1 Background for Analyzing Empirical CVaR

Here, we provide a background on *empirically* analysing the sample-driven CVaR metric of any risk distribution simply because the empirical version of the CVaR metric is what a cyber-risk manager will have access to. The

CVaR at level $\alpha \in (0, 1]$ of a random variable X is

$$\mathrm{CVaR}_\alpha(X) \triangleq \inf_v \left\{ v + \frac{1}{\alpha}\mathbb{E}\left[(X-v)^+\right] \right\}.$$

It is well known from [100] that, when X has a continuous distribution, that
$\mathrm{CVaR}_\alpha(X) = \mathbb{E}\left[X \mid X \geqslant \mathrm{VaR}_\alpha(X)\right]$, where $\mathrm{VaR}_\alpha(X) \triangleq \sup_x\{x \mid \mathbb{P}(X \geqslant x) \geqslant \alpha\}$ is the α-quantile (or VaR) of X. While this relation does not necessarily hold if X has a discontinuous distribution, *we can nonetheless roughly interpret* $\mathrm{CVaR}_\alpha(X)$ as the expected loss over the $\alpha\%$ worst cases. We now make the following assumption necessary for deriving tight bounds of finite sample CVaR on aggregate network cyber-loss impact.

Assumption 1. The random variable X satisfies $\mathrm{supp}(X) \subseteq [0, U]$, and its samples $X_1, \ldots, X_n$ are independent.

Suppose one were to map the outcome of a stochastic process modeling the aggregate cyber-loss over time and space in an IIoT network, the loss values obtained on multiple sample paths of the stochastic process would be independent. This would closely map empirical estimates of such aggregate loss estimates from the real network at different points in time - thereby justifying the 'independence' aspect of the assumption. An upper bound criterion (as stated in the assumption) is common to the application of many concentration inequalities, and without loss of generality we use supp $(X) \subseteq [0, U]$, to reflect the fact that the minimum loss impact value is zero.

For the case of CVaR_α, we denote the simple empirical CVaR estimator by $\widehat{\mathrm{CVaR}}_\alpha$, and express it as:

$$\widehat{\mathrm{CVaR}}_\alpha\left(X_1, \ldots, X_n\right) \triangleq \inf_v \left\{ v + \frac{1}{n\alpha}\sum_{i=1}^{n}\left(X_i - v\right)^+ \right\}, \tag{8}$$

where $(y)^+ = \max(y, 0)$. This empirical estimator is an intuitive and typically popular one [100] based on the method of moments. Such estimators are *efficiently computed* for large n and most convex loss functions - the CVaR function, CVaR_α, being one having a piece-wise linear loss function [13]. *In the analysis to follow, we denote* $X = X_T$ *to the random variable instantiating the node-aggregate cyber-loss upto a pre-specified time period* T *in an IIoT network.*

5.2 The Lower Bound of (Empirical APT Risk - True APT Risk)

A powerful result from McDiarmid [69][23] allows us to quantify the probability of the empirical estimates of the CVaR of the aggregate space-time cyber-loss in an IIoT network exceeding the true value of CVaR *atleast* by an error margin, as a function of the margin and quantity of empirical samples.

THEOREM 5.1. *(McDiarmid [69]). Consider a function* $f : S^n \to \mathbb{R}$ *which satisfies*

$$\sup_{x_1,..,x_n,x_i' \in S}\left|f\left(x_1, \ldots, x_n\right) - f\left(x_1, .., x_i', .., x_n\right)\right| \leqslant c_i$$

for all $i = 1, \ldots, n$. *Let* $X_1, \ldots, X_n$ *be independent random variables taking values in S. Then*

$$\mathbb{P}\left(\left|f\left(X_1, \ldots, X_n\right) - \mathbb{E}\left[f\left(X_1, \ldots, X_n\right)\right]\right| \geqslant \varepsilon\right) \leqslant 2\mathrm{e}^{\frac{-2\varepsilon^2}{\sum_{i=1}^n c_i^2}}$$

McDiarmid's inequality says we need on the order of $n_H \triangleq (\sum_{i=1}^n c_i^2/\varepsilon^2)\log(1/\delta)$ samples to estimate the sample mean within a precision of ε with probability at least $1 - \delta$. Using this above theorem, we have the following result that will be used (alongside McDiarmid's result) to derive the main result in this section.

LEMMA 5.2. *The estimator,* $\widehat{\mathrm{CVaR}}_\alpha$, *satisfies the following:*

$$\mathbb{E}\left[\widehat{\mathrm{CVaR}}_\alpha\left(X_1, \ldots, X_n\right)\right] \leqslant \mathrm{CVaR}_\alpha\left(X_1, \ldots, X_n\right).$$

The result states that that estimator $\widehat{\mathrm{CVaR}}_\alpha$ lower bounds CVaR_α in expectation, and is used in conjunction with McDiarmid's result to derive our following main result.

THEOREM 5.3. *Consider a cyber-risk manager having access to* $X = (X_1 \dots X_n)$ *- a vector of* n *samples of the node and time aggregate cyber-loss impacts in an IIoT network. Then*

$$\mathbb{P}\left(\widehat{\mathrm{CVaR}}_\alpha\,(X) \geqslant \mathrm{CVaR}_\alpha\,(X) + \varepsilon\right) \leqslant \mathrm{e}^{-2\frac{\varepsilon^2\alpha^2}{U^2}\cdot n}, \tag{9}$$

or, with $\mathrm{APTRisk}_\alpha(X) = \mathrm{CVaR}_\alpha(X)$,

$$\mathbb{P}\left(\widehat{\mathrm{APTRisk}}_\alpha\,(X) \geqslant \mathrm{APTRisk}_\alpha\,(X) + \varepsilon\right) \leqslant \mathrm{e}^{-2\frac{\varepsilon^2\alpha^2}{U^2}\cdot n},$$

where $\widehat{\mathrm{APTRisk}}_\alpha(X) = \widehat{\mathrm{CVaR}}_\alpha(X)$.

Implications for Cyber-Risk Management - The result provides a closed form expression of the lower bound of the deviation in empirical APT risk (measured as empirical CVaR) with the theoretical true value of the APT risk, as a function of sample count. *In practice, a higher sample count reduces the deviation*, and hence a cyber-risk manager (e.g., enterprise cyber-risk officer, cyber-insurer) should ideally get access to sufficient number of samples prior to policy under-writing to reduce cyber-risk estimation uncertainty.

5.3 The Upper Bound of (Empirical APT Risk - True APT Risk)

A powerful result from Hoeffding [45][23] allows us to quantify the probability of the empirical estimates of the CVaR of the aggregate space-time cyber-loss in an IIoT network exceeding the true value of CVaR *atleast* by an error margin, as a function of the margin and quantity of empirical samples.

THEOREM 5.4. *(Hoeffding [45]). Let* $X_1, \dots, X_n$ *be i.i.d. random variables with supp*$(X) \subseteq [0, U]$. *Then, for any* $\varepsilon \geqslant 0$, *we have* $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| \geqslant \varepsilon\right) \leqslant 2\mathrm{e}^{-2(\frac{\varepsilon}{U})^2 n}$

Hoeffding's inequality states that we need on the order of $n_H \triangleq (U/\varepsilon)^2 \log(1/\delta)$ samples to estimate the sample mean within a precision of ε with probability at least $1 - \delta$. Given that CVaR of a distribution requires a conditional expectation evaluation of its α-tail, we would expect for CVaR_α to need on the order of n_H/α samples, and approximately α% samples falling in the α -tail. Using the Hoeffding's theorem, we have the following result that will be used to derive the main result in this section.

LEMMA 5.5. *The inequality* $\widehat{\mathrm{CVaR}}_\alpha\,(X_1, \dots, X_n) \geqslant \frac{1}{n\alpha}\sum_{i=1}^{\lfloor n\alpha \rfloor} X_{(i)}$ *holds, where* $X_{(i)}$ *are the decreasing order statistics of* X_i, *i.e.,* $X_{(1)} \geqslant X_{(2)} \geqslant \dots \geqslant X_{(n)}$.

The result states that that estimator $\widehat{\mathrm{CVaR}}_\alpha$ upper bounds CVaR_α in expectation, and is used in conjunction with Hoeffding's result to derive our following main result.

THEOREM 5.6. *Consider a cyber-risk manager having access to* $X = (X_1 \dots X_n)$ *be a vector of* n *samples of the node and time aggregate cyber-loss impacts in an IIoT network. Then for any* $\varepsilon \leqslant 0$,

$$\mathbb{P}\left(\widehat{\mathrm{CVaR}}_\alpha\,(X) \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon\right) \leqslant 3\mathrm{e}^{(-\frac{1}{5})\alpha(\frac{\varepsilon}{U})^2\cdot n} \tag{10}$$

or, with $\mathrm{APTRisk}_\alpha(X) = \mathrm{CVaR}_\alpha(X)$,

$$\mathbb{P}\left(\widehat{\mathrm{APTRisk}}_\alpha\,(X) \leqslant \mathrm{APTRisk}_\alpha\,(X) - \varepsilon\right) \leqslant 3\mathrm{e}^{(-\frac{1}{5})\alpha(\frac{\varepsilon}{U})^2\cdot n},$$

where $\widehat{\mathrm{APTRisk}}_\alpha(X) = \widehat{\mathrm{CVaR}}_\alpha(X)$.

Implications to Cyber-Risk Management - The result provides a closed form expression of the upper bound of the deviation in empirical APT risk (measured as empirical CVaR) with the theoretical true value of the APT risk, as a function of sample count. *In practice, a higher sample count reduces the deviation*, and hence a cyber-risk manager (e.g., cyber-insurer) should ideally get access to sufficient number of samples prior to policy under-writing to reduce cyber-risk estimation uncertainty.

5.4 Two-Sided Bound of (Empirical APT Risk - True APT Risk)

It would always be best for a cyber-risk manager to have access to simultaneous two-sided (in comparison to the abover-mentioned one-sided bounds) CVaR bounds from a vector of n samples of aggregate cyber-loss estimates. One could obtain a single powerful and tighter *two-sided* bound of (Empirical APT Risk - True APT Risk), using the Wasserstein (Kantorovich-Rubinstein) distance metric [115]. More specifically, we provide such a two-sided concentration bound for the empirical APT risk estimate by the following steps based on [15]: (a) drawing a mathematical relationship between the estimation error, $\left|\widehat{\mathrm{CVaR}}_\alpha - CVaR_\alpha(X)\right|$, and the Wasserstein distance between the true and empirical cyber-loss impact distribution functions, and (b) subsequently bounding the Wasserstein distance between these two distributions. However, before going through these steps, we need the following assumption.

Assumption 2. There exist $\beta > 1$ and $\gamma > 0$ such that $\mathbb{E}\left(\exp\left(\gamma|X|^\beta\right)\right) < \top < \infty$.

The assumption states that a r.v. X, in our case reflective of a node-aggregate cyber-loss random variable for IIoT networks over a given time period, satisfies an exponential moment bound. We now define the Wasserstein (Kantorovich-Rubinstein) distance [115].

Definition 5.7. The Wasserstein (Kantorovich-Rubinstein) distance between two cumulative distribution functions (CDFs) F_1 and F_2 on $\mathbb{R}$ is defined by

$$W_{p=1}(F_1, F_2) \triangleq \left[\inf_{F \in \Gamma(F_1,F_2)} \int_{\mathbb{R}^2} |x-y| dF(x,y)\right], \tag{11}$$

where $\Gamma(F_1, F_2)$ is the set of all joint distributions on $\mathbb{R}^2$ having F_1 and F_2 as marginals. *The function $f : \mathbb{R} \rightarrow \mathbb{R}$ is L-Lipschitz if it is L-Hölder of order* 1, and consequently, a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is L-Hölder of order p if $|f(x) - f(y)| \leq L|x-y|^p$ for all $x, y \in \mathbb{R}$.

THEOREM 5.8. *Consider a cyber-risk manager having access to $X = (X_1 \dots X_n)$ be a vector of n samples of the node and time aggregate cyber-loss impacts in an IIoT network. Suppose X be the node-aggregate cyber-loss r.v. with CDF F and mean μ satisfying Assumption 2, for some $\beta > 1$. Then, for any $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\left|\widehat{\mathrm{APTRisk}}_\alpha(X) - \mathrm{APTRisk}_\alpha(X)\right| > \varepsilon\right) \leq G(n, c_1, c_2, \alpha, \varepsilon, \beta), \tag{12}$$

where $G(n, c_1, c_2, \alpha, \varepsilon, \beta)$ is given by the expression

$$c_1 \left[\exp\left[-c_2 n(1-\alpha)^2\varepsilon^2\right] \mathbb{I}\{(1-\alpha)\varepsilon \leq 1\} + \exp\left[-c_3 n(1-\alpha)^\beta \varepsilon^\beta\right] \mathbb{I}\{(1-\alpha)\varepsilon > 1\}\right].$$

Here, the constants c_1, c_2 and c_3 obey

$$\mathbb{P}\left(W_1(F_n, F)\right) > \varepsilon) \leq B(n, \varepsilon, \beta),$$

where $B(n, \varepsilon, \beta)$ is denoted by the following expression:

$$c_1 \left(\exp\left(-c_2 n\varepsilon^2\right) \mathbb{I}\{\varepsilon \leq 1\} + \exp\left(-c_3 n\varepsilon^\beta\right) \mathbb{I}\{\varepsilon > 1\}\right),$$

for some c_1, c_2, and c_3 that depend on the parameters β, γ and $\top$ in Assumption 2.

Implications for Cyber-Risk Management - The result provides a closed form expression of the two-sided bound of the deviation in empirical APT risk (measured as empirical CVaR) with the theoretical true value of the APT risk, as a function of sample count. Similar to the implications of theorems 5.3 and 5.6, *a higher sample count in practice reduces the deviation*, and hence a cyber-risk manager (e.g., enterprise cyber-risk officer, cyber-insurer) should ideally get access to sufficient number of samples prior to policy under-writing. It is also the case that the two-bound scenario needs more amount of samples to guarantee error performance bounds on the two sides when compared to the number of samples required for one-sided bounds.

5.5 Providing Bounds When Cyber-Loss Distribution is Heavy-Tailed

Thus far we have derived concentration bounds for the network-aggregate cyber-loss distribution, specifically suited for the latter being statistically light-tailed. However, such results do not hold when these distributions are heavy-tailed, i.e., higher loss moments might not exist, even if the mean might exist. In practice, such a possibility can arise if (novel and) extremely sophisticated APT cyber-attacks can cause ICS business disruption for months, leading to an outlier-high first-party cyber-loss amount.

In such scenarios, estimates of empirical mean of network-aggregate cyber-loss gathered by managers will not converge to the true mean. In such scenarios, we can adopt the seminal methodology developed in [68] to replace sample mean with robust proxies, and obtain high-confidence bounds for the excess risk of aggregate cyber-loss estimators. We have the following result, borrowed and adapted from [68] that provides a robust proxy of an empirical sample mean estimate of a possible heavy-tailed aggregate cyber-loss distribution. The proxy estimate converges exponentially fast towards zero error with large sample size.

Theorem 5.9 (Adapted from Mathieu and Minsker, 2021). *Assume that we have n empirical samples of the network and time aggregate cyber-loss distribution and that $\sigma(\ell,\mathcal{F}) < \infty$. Then, for appropriately set k and Δ,*

$$\mathcal{E}(\widehat{f_n}) \le \bar{\delta} + C(\mathcal{F},P)\left(\frac{s}{n^{2/3}} + \left(\frac{\mathcal{O}}{n}\right)^{2/3}\right)$$

with probability at least $1-e^{-s}$ for all $s \lesssim k$. Moreover, if $\sup_{f\in\mathcal{F}}\mathbb{E}^{1/4}(\ell(f(X)) - \mathbb{E}\ell(f(X)))^4 < \infty$, then

$$\mathcal{E}(\widehat{f_n}) \le \bar{\delta} + C(\mathcal{F},P)\left(\frac{s}{n^{3/4}} + \left(\frac{\mathcal{O}}{n}\right)^{3/4}\right),$$

again with probability at least $1-e^{-s}$ for all $s \lesssim k$ simultaneously. Now assume that $\sup_{f\in\mathcal{F}}\mathbb{E}^{1/4}(\ell(f(X)) - \mathbb{E}\ell(f(X)))^4 < \infty$. There exists an estimator $\widehat{f_n}''$ such that

$$\mathcal{E}\left(\widehat{f_n}''\right) \le \bar{\delta} + C(\mathcal{F},P,\rho)\left(\frac{\mathcal{O}}{n} + \frac{s}{n}\right)$$

with probability at least $1-e^{-s}$ for all $1 \le s \le s_{\max}$ where $s_{\max} \to \infty$ as $n \to \infty$. Here, $\bar{\delta}$ is the quantity that often coincides with the optimal rate for the excess risk [4, 65].

Understanding the Theorem - Our goal is to robustly estimate in an empirical fashion - the first moment (i.e., statistical mean) of the network aggregate cyber-loss (risk) distribution when the latter is heavy-tailed. The main contribution through this theorem is the proof of high-confidence bounds for an accurate measure of the excess risk, $\mathcal{E}(f) := \mathbb{E}\ell(f(X)) - \mathbb{E}\ell(f_*(X))$ of the empirical estimators (when compared to the theoretical ground truth) $\widehat{f_n}$ and $\widehat{f_n}^U$ (as a function of sample size n), where f denotes the stochastic network-aggregate cyber-loss function, f_* is its best empirical estimate, $\widehat{f_n}^U$ is the U-statistic variant of $\widehat{f_n}$, and l is a loss function.

First, we observe from the theorem that convergence rates of order $n^{-1/2}$ are achieved with exponentially high probability if $\sigma(\ell,\mathcal{F}) = \sup_{f\in\mathcal{F}}\sigma^2(\ell,f) < \infty$ and $\mathbb{E}\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{n}}\sum_{j=1}^{N}(\ell(f(X_j)) -$ $\mathbb{E}\ell(f(X))) < \infty$. This reflects achieving high accuracy on empirical estimates of the statistical with a small number of samples. The latter is true if the class $\{\ell(f),\ f\in\mathcal{F}\}$ is P-Donsker [36], in other words, if the empirical process $f \mapsto \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(f(X_j)) - \mathbb{E}\ell(f(X)))$ converges weakly to a Gaussian limit.

Next, the theorem demonstrates that under additional assumption requiring that any $f\in\mathcal{F}$ with small excess risk must be close to f_* that minimizes the expected loss, $\widehat{f_N}$ and $\widehat{f_n}^U$ attain fast rates; the theorem states the bounds only for $\widehat{f_n}$ while the results for $\widehat{f_n}^U$ are similar, up to the change in absolute constants. Moreover, there is the design of a two-step estimator based on $\widehat{f}_n$ that is capable of achieving faster rates whenever $\bar{\delta} \ll n^{-3/4}$. Estimator $\widehat{f}_n''$ is based on a two-step procedure, where $\widehat{f}_n$ serves as an initial approximation that is refined on the second step via the risk minimization restricted to a “small neighborhood” of $\widehat{f}_n$.

6 EXPERIMENTAL EVALUATION

Thus far, we have derived theory relevant to a cyber-risk manager interested in an accurate estimate of the network and time aggregate adverse impact, i.e., cyber-loss, in an IIoT network due to an APT cyber-attack. We now run real-world test-bed experiments in the FIT IoT-Lab [2] to study salient aspects of our proposed theory that cannot be inferred via the proposed theory. More specifically, we *investigate the following questions under the influence of popular malware families*:

(1) Do parameters that ensure the stability of the malware-spread process in theory, exist in practice?
(2) Do (network/time) aggregate cyber-loss distributions significantly differ for IIoT networks of various sizes?
(3) How accurate are theoretical MFA estimates of mean aggregate cyber-loss, compared to empirical estimates?
(4) How do empirical CVaR estimates for aggregate cyber-loss distributions vary over time in an IIoT network?

We organize this section in two parts: in the first part, we describe our experimental setup; in the second part, we analyze our experimental results with respect to questions (i)-(iv).

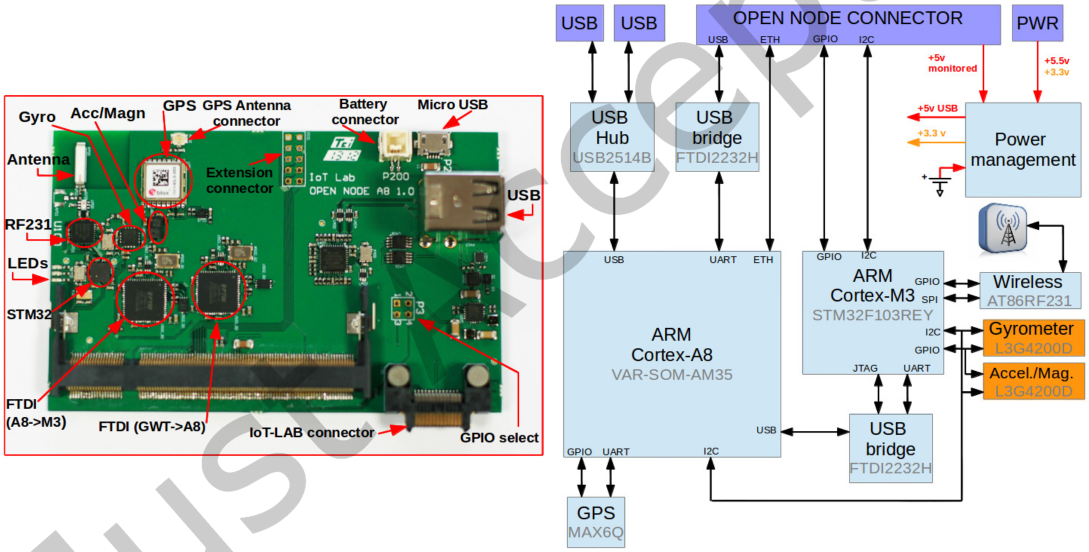


Fig. 4a: IoT-LAB A8-M3

Fig. 4b: IoT-LAB A8-M3 Architecture

6.1 Experimental Setup

Testbed - We ran experiments in *Future Internet Testing* (FIT) IoT-Lab, that provides access to a very large scale IoT testbed. We experimented on the Grenoble site, that provided us access to 228 IoT-LAB *A8-M3 boards* (see Figures 4a. and 4b.), and allowed us to create the standard *mesh* (e.g., as in industry campuses, smart homes, low-range wireless networks), *star* (e.g., as in factories, oilfields, LPWANs), and *cluster* (e.g., as in a smart grid) wireless communication topologies (see Figure 5.) using these devices. The IoT-LAB A8-M3 boards are equipped with *ARM Cortex-A8* microprocessors having 256 MB of RAM, and radio chips enabling the former to communicate with other IEEE 802.15.4 compliant (low power) objects within wireless radio range. They are installed with *Yocto OS*, that can be used to create tailored Linux images for embedded and IoT devices. We created a WPAN using 10, 50, 100, and 200 IoT-LAB A8-M3 boards (and for each topology type), and let them communicate with each other using the *6LoWPAN* (IPv6 over low-power WPANs) technology. Apart from the fact that we did not have access to a real testbed of more than 228 nodes, current ICSs are known to have a count of the number of communication devices (field devices) that are orders of magnitude lesser than traditional mobile and wireless IT systems (that could have tens of 1000s of communication devices in a dynamic network). This is because ICS networks are relatively more static in size and software/firmware configurations, when compared to these IT system networks.

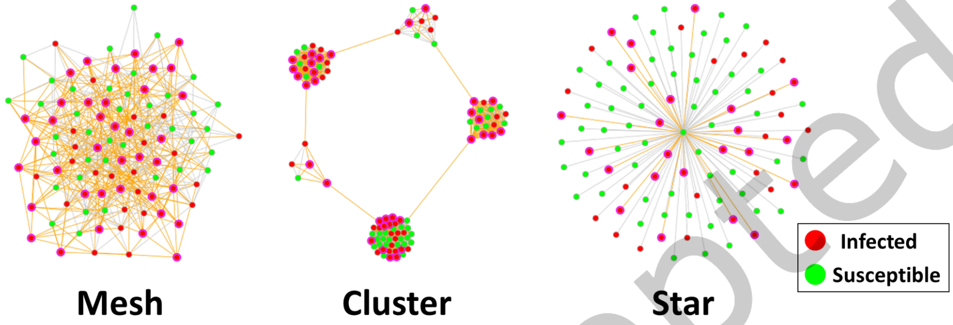


Fig. 5: Illustration of IoT Network Topology Types

Staging Malware Attacks - Summarizing from [78][77][7], malware-based IoT attacks (the focus of our paper) operate in three stages: *injection*, *infection*, and *attack*. The injection stage involves gaining "control" of the IoT device (primarily via root access) through popular mechanisms that include credential hijacking, password brute-forcing, dictionary attacks, or utilizing known device/system/firmware vulnerabilities [7][124]. In line with [78], we simulate different variants of IoT malware and built a comprehensive dataset using 1000 malware samples from the most popular malware families: *Zorro*, *Gayfgt*, *Mirai*, *Hajime*, *IoTReaper*, *Bashlite*, *nttpd*, and *linux.wifatch*. The malware samples are collected from *IoTPOT* [82], *VirusTotal*, and *OpenMalware* sources, and replicated (injected) into arbitrarily pre-selected nodes in lab-created WPAN IoT networks. As an example, the Mirai botnet used its capability of password brute-forcing behind the *modus operandi* to gain directly gain control (as part of the *injection stage*) of an IoT device. This injection stage was followed by the *infection stage* where the attacker set up communication with the bot master (C&C server). We reserved certain number of nodes on the FIT IoT-Lab necessary to act as the C&C servers that implemented the infection stage throughout the communication network - similar to the classic Mirai botnet operation. More specifically, via the use of these servers, Mirai downloaded the required toolkit having information about the C&C servers, and copied into the compromised nodes - as part of the infection phase in the IoT network and taking full control of compromised nodes. Finally, once a malware has full control of certain nodes in the network, it launches the *attack* stage through which it first stops security services on telnet compromised network nodes and executes cyber-attacks such as denial of service (DoS) attacks, ransomware attacks, and their likes [116]. *In this paper, we showcase service unavailability-led "QoE degradation" attacks.* The attack stage comprises of launching time-controlled synchronous device malfunction attacks, on each infected network node, that follow, (a) a Poisson attack arrival process (mean (per-unit of time) attack rate of $\lambda = 3$) [w.l.o.g.], and (b) a renewal attack arrival process (with inter-arrival times following a Weibull distribution with parameters $\lambda = 5$, and $k = 1.5$)[w.l.o.g.] [103]. Device malfunction is executed via changes in the the password of the device using the *passwd* command in Linux, that

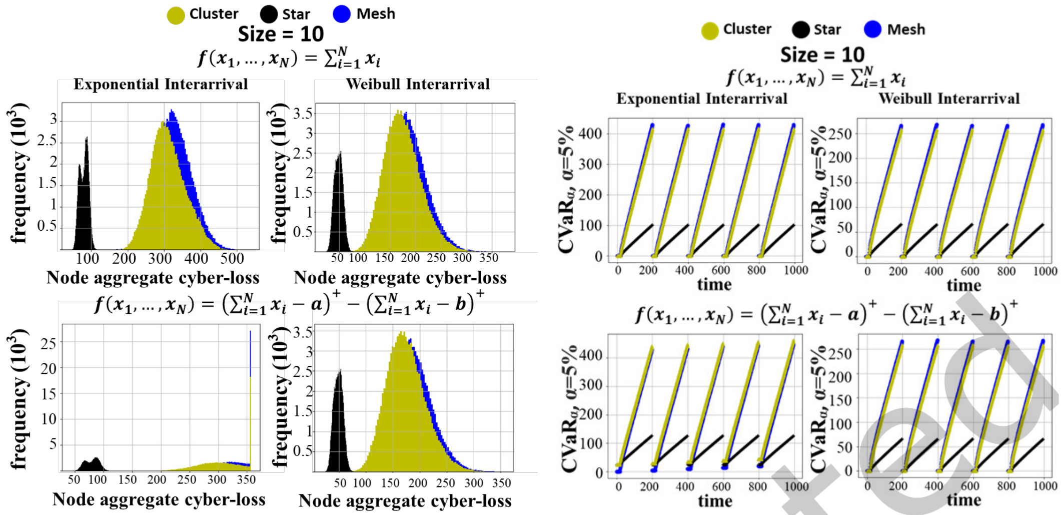


Fig. 6: Illustrating for IoT network topologies (cluster, star, mesh) of size 10, (a) **node-aggregate cyber-loss distributions** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, and (b) **CVaR time series** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, with loss-coverage every 200 time units. The star topology is the most efficient in terms of IIoT first party cyber-risk management.

locks up the device and does not allow legitimate users (ourselves) to access it. *We conduct a total of 100,000 fresh Monte Carlo runs of the three-staged process on these popular malwares.*

Per-Node Cyber-Loss Impact - We assume QoE/S loss impact (e.g., business downtime) in the IIoT network due to a node becoming dysfunctional (e.g., an exploding turbine post a successful cyber-attack) maps to a loss-normalized scale of $(0,1]$, and is proportional to the Bonacich centrality [19] of the node. The rationale behind this assumption lies in the fact that a network central IIoT device (e.g., SCADA controller) controls multiple devices connected to this node, and a reliability hit on the latter increases the likelihood of the failure of multiple devices together (amplifying business downtime). Assuming that the reliability/resilience of an IIoT network is a function of the sum of the reliability/reliability of its individual nodes, a central node in the IIoT going down adversely impacts network-wide QoE/S loss impact more when compared to the same when a non-central node becomes dysfunctional.

Empirically Estimating Spread/Loss Parameters - We estimate the values of rates $\frac{1}{\alpha}$, δ, and $\frac{1}{\gamma}$ for each of the 100,000 Monte Carlo runs of our three-staged malware launch experiment. We observe that for all nodes, α values lie in the range $[0.3, 0.4]$, $\frac{1}{\delta}$ values lie in the range $[0.25, 0.35]$, and the γ's lie in the range $[0.12, 0.129]$. In the interest of space, and w.l.o.g., we plot results for the 'median' setting where $\alpha = 0.35$, $\frac{1}{\delta} = 0.3$, and $\gamma = 0.125$. We arrive at the γ values in a two-stage process: (i) in the first stage, we borrow system set-up ideas from the ML-driven anomaly detection framework in [77][78] to design a similar ML framework in the FIT IoT Lab that effectively detects infected IoT boards from system log data, and (ii) in the second stage, we reboot/cure the infected IoT board. The reciprocal of the time taken for (i)+(ii) gives us values of γ. The empirical node-aggregate cyber-loss per instant of time in the IoT network is captured by two function types: (i) $f(x_1,\dots,x_N) = \sum_{i=1}^{N} x_i$, and (ii) $f(x_1,\dots,x_N) := \left(\sum_{i=1}^{N} x_i - a\right)^+ - \left(\sum_{i=1}^{N} x_i - b\right)^+$), where a is the minimum loss amount a coverage agency decides to provide coverage for, and $b - a > 0$ is the upper coverage limit set by the agency. Note that the situation

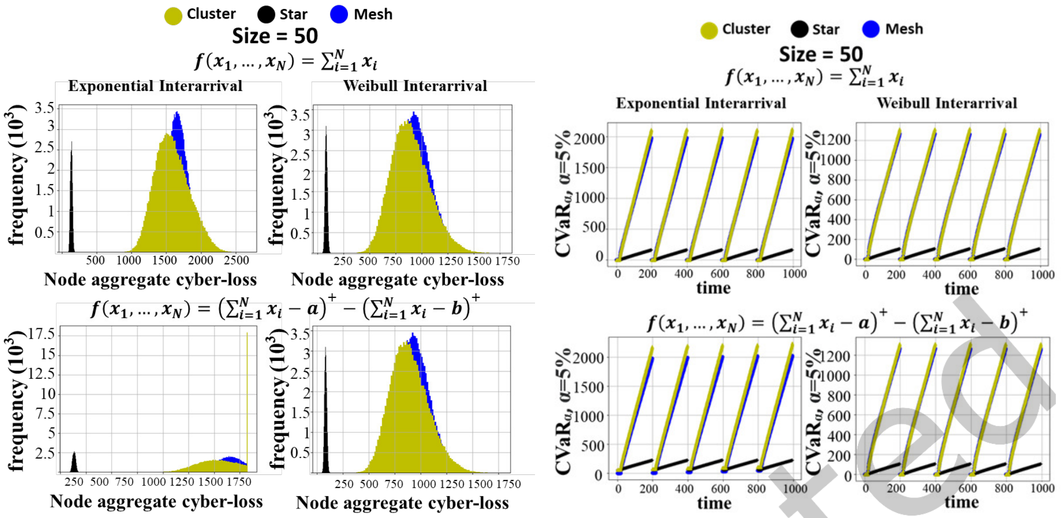


Fig. 7: Illustrating for IoT network topologies (cluster, star, mesh) of size 10, (a) **node-aggregate cyber-loss distributions** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, and (b) **CVaR time series** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, with loss-coverage every 200 time units. The star topology is the most efficient in terms of IIoT first party cyber-risk management.

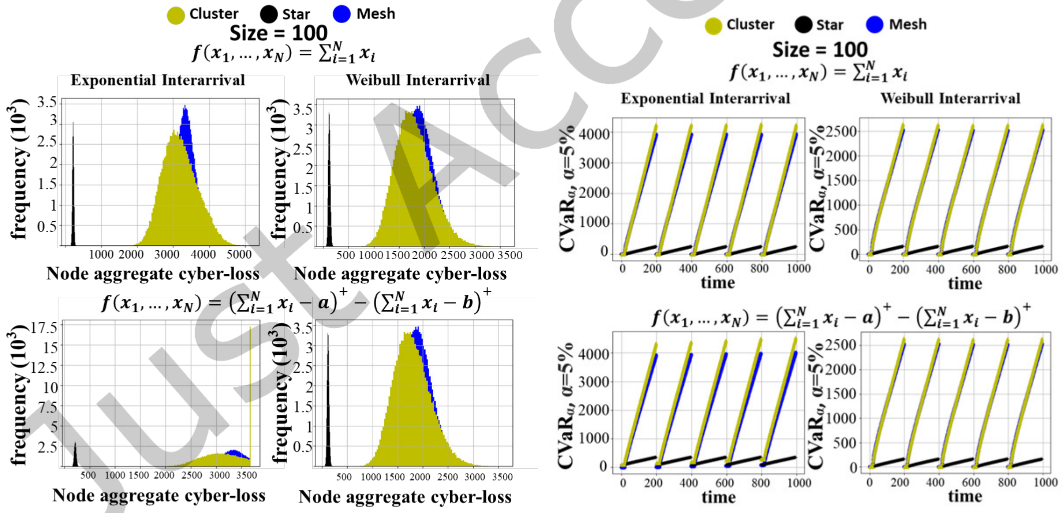


Fig. 8: Illustrating for IoT network topologies (cluster, star, mesh) of size 10, (a) **node-aggregate cyber-loss distributions** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, and (b) **CVaR time series** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, with loss-coverage every 200 time units. The star topology is the most efficient in terms of IIoT first party cyber-risk management.

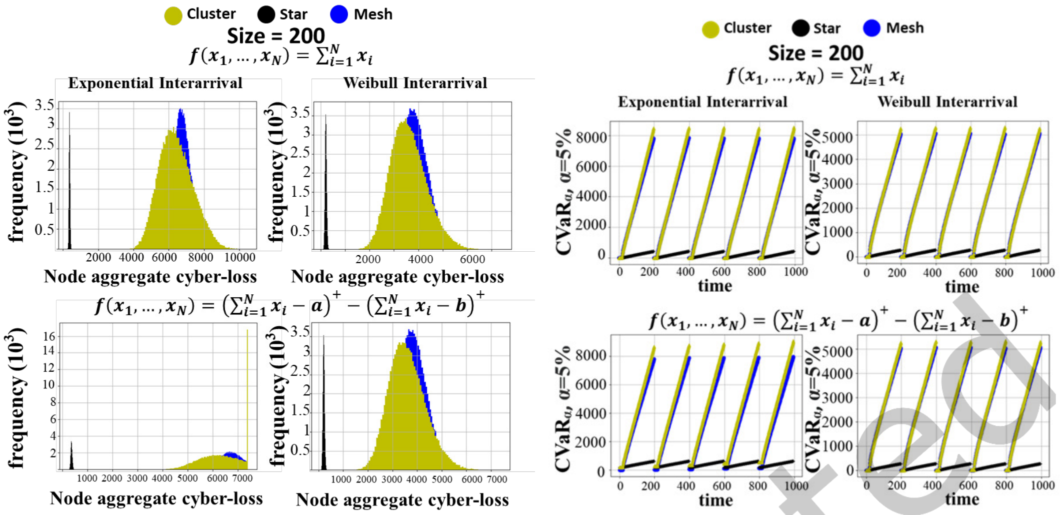


Fig. 9: Illustrating for IoT network topologies (cluster, star, mesh) of size 200, (a) **node-aggregate cyber-loss distributions** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, and (b) **CVaR time series** for *Exponential* and *Weibull* cyber-attack inter-arrival distributions, with loss-coverage every 200 time units. The star topology is the most efficient in terms of IIoT first party cyber-risk management.

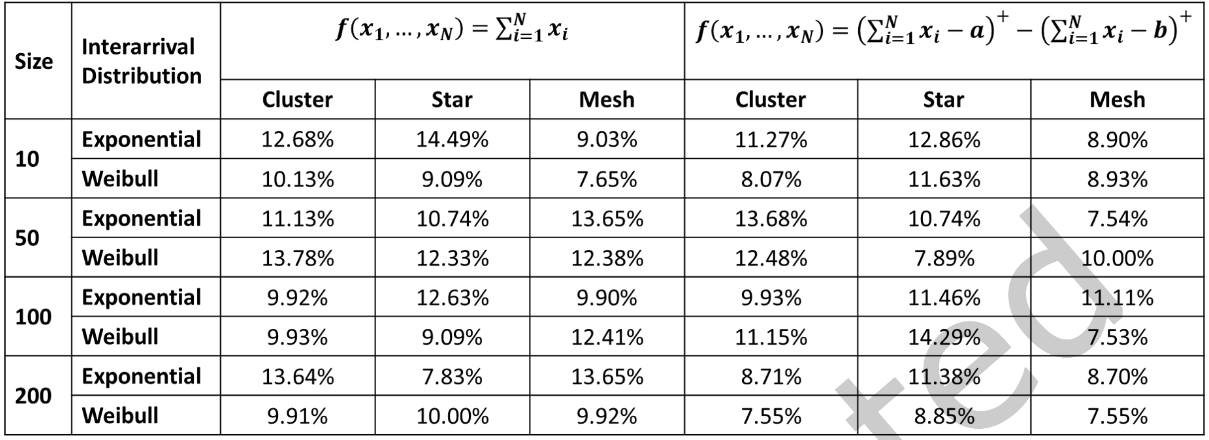
when $a = 0$ (no lower coverage limit) and $b = \sum_{i=1}^{N} x_i$ indicates a full coverage clause by a coverage agency. Both (i) and (ii) are standard functions in the loss coverage industry; satisfy the mathematical assumptions behind f as described in Section 2.2; and are continuous, linear, and increasing in $x_1, \dots, x_N$. We plot our results for the case when $a = 0$, and b is 70% of f (w.l.o.g.), for each network configuration.

6.2 Results and their Practical and Policy Connotations

We observe from Figures 6-9 that for nearly all samples from each parameter's (e.g., $\frac{1}{\alpha}$, δ, $\frac{1}{\gamma}$) empirically observed interval span, Theorem 3.1 holds. In the jargon of practice, the spread parameters that we observe from a real-world malware spread at an empirical spread equilibrium satisfy the parameter conditions derived from theory in Theorem 3.1 that ensure spread stability. *Hence, we validate our theory showcased in Theorem 3.1 w.r.t., our experimental setup.*

In relation to the shape of the intra-network space-time aggregate first party cyber-loss distribution over a time period of 200 units (Figures 6a-9a) for the various network topologies, we observe that the tail size (light-tailed in shape) does not increase with network sizes, *thanks to the 'balanced' infection-cure dynamics* of malware spread as achieved in practice. *This is a good result when keeping in mind the success of third-party cyber-risk coverage markets (e.g., cyber-insurance) that would prefer light-tailed cyber-risks to provide coverage for first-party losses, and promote targeted policy measures that incentivize organizations to establish and maintain a balanced infection-cure dynamics.* It is evident and intuitive to note that the star network topology is least susceptible to large aggregate cyber-losses. *However, we put forward the following word of caution regarding the tail of our reported and experimentally-derived aggregate cyber-risk distributions.* It is not always the case that intra-network first party aggregate cyber-losses will be light-tailed in shape. Certain form of cyber-attacks are very sophisticated (e.g., targeted ransomware, sophisticated versions of the Ukraine cyber-attack) and can render an ICS (or some of its important sub-divisions) crippled for days (upto a month) without it functioning, i.e., the time to recovery.

TABLE III: Empirical Estimation Error (%) of Mean Aggregate Cyber-Loss w.r.t. its MFA Estimates Obtained from Theory in Section 4



Size	Interarrival Distribution	$f(x_1,\dots,x_N)=\sum_{i=1}^{N} x_i$			$f(x_1,\dots,x_N)=\left(\sum_{i=1}^{N} x_i - a\right)^+ - \left(\sum_{i=1}^{N} x_i - b\right)^+$		
		Cluster	Star	Mesh	Cluster	Star	Mesh
10	Exponential	12.68%	14.49%	9.03%	11.27%	12.86%	8.90%
	Weibull	10.13%	9.09%	7.65%	8.07%	11.63%	8.93%
50	Exponential	11.13%	10.74%	13.65%	13.68%	10.74%	7.54%
	Weibull	13.78%	12.33%	12.38%	12.48%	7.89%	10.00%
100	Exponential	9.92%	12.63%	9.90%	9.93%	11.46%	11.11%
	Weibull	9.93%	9.09%	12.41%	11.15%	14.29%	7.53%
200	Exponential	13.64%	7.83%	13.65%	8.71%	11.38%	8.70%
	Weibull	9.91%	10.00%	9.92%	7.55%	8.85%	7.55%

This is because ICS technology is customized, proprietary, and often run regressive software and firmware. A large compromised network section equipped with such technology will usually need a high recovery time to get back to normal production regime. In our experimental FIT laboratory setup, we could not consider a very low rate of recovery due to scheduling constraints. In practice days of service dysfunctions might lead to very high costs in loss of production (e.g., energy, water) incurred by an ICS, and subsequently will contribute to heavy-tailed aggrgegate first party cyber-risk.

In relation to the accuracy of the empirical MFA estimate of mean aggregate first party cyber-loss within an IIoT network, with respect to the true theory estimate, we observe (see Table III) the accuracy to be distributed around the [85.5% - 92.5%] for 100,000 Monte Carlo trials (for each topology type). The accuracy gap (from 100%) is primarily due to MFA not being able to capture network asymmetries inherent in practical networked systems. Through this *we claim the first-order estimates from the NIMFA methodology to be fairly accurate*, conditioned (as per Theorem 4.1) on the $\frac{\beta}{\delta}$ ratios (an indication of the cyber-security strength of the network) in our practical setup. The main message we put forward is: *there exists real-world malware spread parameters for IoT networks for which MFA provides fairly accurate estimates of cyber-loss moments.* This has a significant implication for cyber-risk managers who would want to apply targeted intervention policies on IIoT network managers to maintain appropriate $\frac{\beta}{\delta}$ ratios for their networks, not only to lower their chances of covering large losses, but to ensure that they have access to MFA estimates in theory that are close to empirical estimates they may not always have access to. The latter point is relevant in environments where the absence of cyber-information disclosure regulatory policies might hamper effective cyber-risk auditing.

We observe from CVaR time series dynamics (see Figures 6b-9b) that it increases at a *linear rate* till 200 time units (loss mitigation happens every 200 time units), for each topology type and IIoT network size. This pattern has two positive implications for a cyber-coverage agency (e.g., an insurer): (i) the marginal loss growth rate is constant over time - *hence a strong incentive for the agency to deploy coverage contract policies that perennially maintain low values of such constants* for an organization, and (ii) loss estimates are *time-predictable* in advance. It needs to be emphasized that even though aggregate loss sizes and their probability of occurrence might be predictable, the timing of such losses might not, depending on the strategic play between the attacker and the defender [112][58][85] - not the focus of this work.

Overall, we observe from Figures 6-9 that the aggregate first party cyber-loss distributions in an intra-organizational IIoT network follow a *light-tailed distribution*. This is primarily due to the local nature of the IIoT network combined with the fact that a certain fraction of the number of network nodes are periodically monitored and recovered, if compromised by APT malware. This latter condition, satisfying a certain recovery threshold (not modeled in this work) is necessary to prevent cyber-loss distributions to become heavy-tailed.

7 AN ICS CASE STUDY

In this section, we first lay out a case study illustrating the applicability of APT breaches in real world IIoT networks through a recent example of the *Pipedream* malware (built to target machine automation devices) developed to execute on IIoT-driven industrial control systems (ICSs). *Our case study helps us connect the proposed theory in this paper with the real world.* The generality of Pipedream is that it can interact with specific industrial equipment embedded in different types of machinery leveraged across multiple industries.

The Basic Elements of the Pipedream Malware - The functioning basis of industrial automation networks are IIoT-driven equipment enabling network operators to translate digital information into manual actions in an automated fashion. Such equipment, due to their hardware and firmware diversity, typically speak different communication languages across different portions of the network using standard communication protocols. Pipedream includes the TAGRUN, CODECALL, and OMSHELL modules allowing a cyber-attacker to send instructions to ICS components using industrial network protocols such as *OPC UA*, *Modbus*, and *Codesys*. Though these modules are general enough for ICS components, Pipedream developers developed them specifically for controllers from Schneider Electric and Omron - in other words, the TAGRUN, CODECALL, and OMSHELL modules were targeted at (a) OPC servers, (b) Schneider Electric Modicon M251, Modicon M258, and Modicon M221 Nano PLCs, and (c) Omron NX1P2, NJ501 PLCs, and R88D-1SN10F-ECT servo drive. *These elements characterize tools to conduct direct and indirect, i.e., spread-based, cyber-attacks on IIoT-networked ICS components.*

The APT Functioning Methodology - *An APT cyber-breach executes through the orchestrated functioning of TAGRUN, CODECALL, and OMSHELL modules.* TAGRUN performs a reconnaissance role through its ability to scan for and enumerate OPC UA servers. OPC UA acts as a central communications protocol to collect and store data from ICS assets in industrial environments promoting the functioning of production systems and control processes, and forwards them to hackers for their access and subsequent modification. In addition, TAGRUN helps brute forcing IIoT device credentials, and outputting log files. CODECALL communicates with IIoT-driven ICS devices using the Modbus protocol, that enables CODECALL to interact with devices from different manufacturers. However, the tool contains a specific module to interact with, scan, and attack (using brute-force, DDoS) Schneider Electric's Modicon M251 (TM251MESE) PLC using Codesys, which is used by the company's proprietary *EcoStruxure Machine Expert protocol*. OMSHELL is designed to obtain shell access to Omron PLCs, that include Omron NX1P2, NJ501, and R88D-1SN10F-ECT servo drive. The tool primarily operates using the HTTP and FINS over UDP protocols for (a) scanning and device identification, (b) wiping a device's program memory and resetting the device, (c) connecting to a device backdoor to enable arbitrary command execution, (d) capturing network traffic, and (e) killing arbitrary processes running on an IIoT device. TAGRUN, CODECALL, and OMSHELL are illustrated in Figure 10. Syncing with each other, TAGRUN, CODECALL, and OMSHELL can disrupt controllers to cascadingly shutdown device operations, reprogram controllers to sabotage industrial processes, and disable safety controllers to cause physical damage (see Figure 11).

8 MULTI-PARTY SUPPLY CHAIN IMPACT OF FIRST-PARTY CYBER-RISK

APT breaches have cyber-loss implications beyond the first-party impact on an intra-organizational setting - especially when a said organization is part of a service supply chain ecosystem. It is common knowledge (courtesy *Dragos* OT-CERT reports) that big IT driven businesses have multiple small/medium supplier organizations in

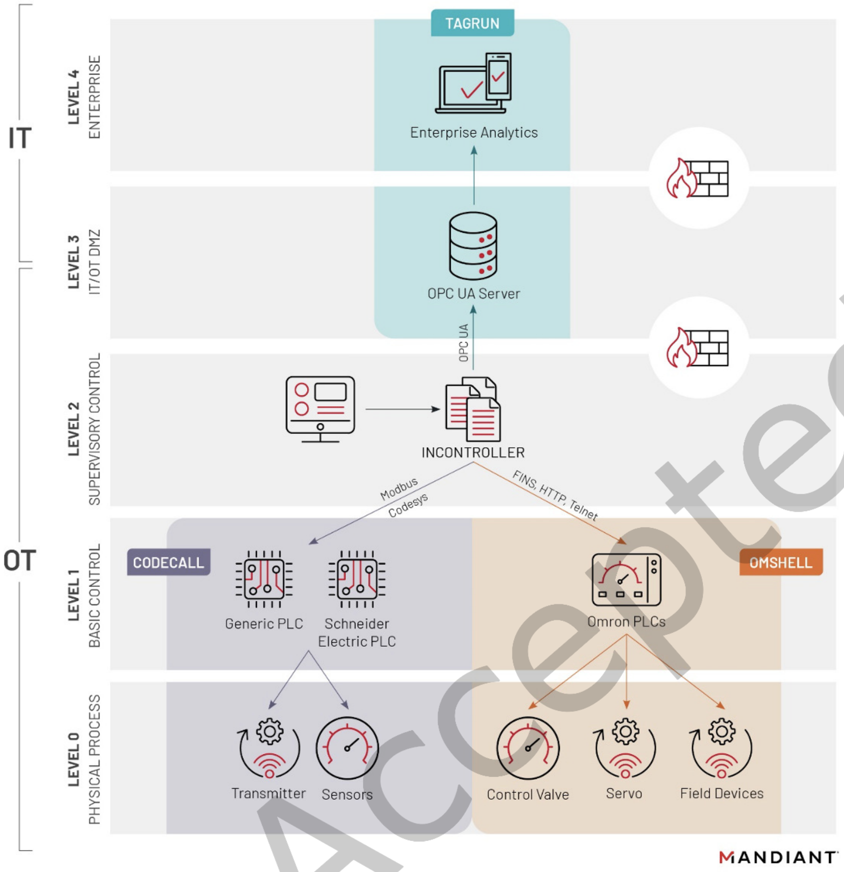
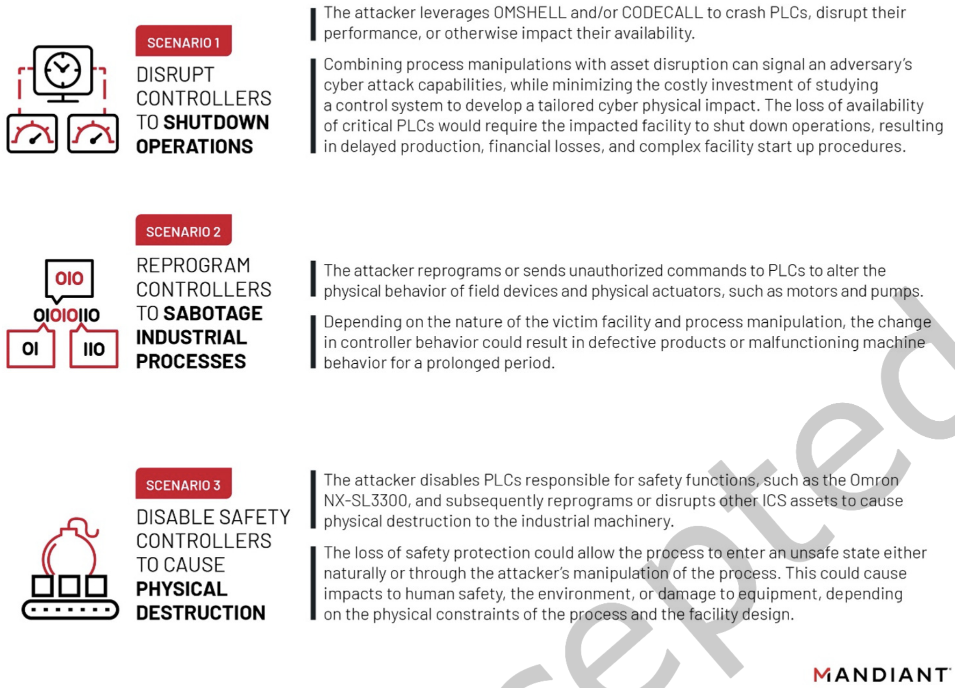


Fig. 10: The TAGRUN, CODECALL, and OMSHELL Modules (*Source:* Mandiant)

the supply chain, and a cyber-hit on any one of them could result in multi-millions US dollars in revenue loss for the source organization. Given also the fact that most small and medium businesses (SMBs) do not focus on cyber-security, it is quite likely that a (nation state) cyber-adversary could findit relatively easy to target big organizations indirectly via disrupting the services of small suppliers in the supply chain (as was done in the case of *SolarWinds* and *Kaseya* cyber-attacks).

A question of significant cyber-risk management (CRM) interest post establishing (via model-based simulations) the light-tailed nature of intra-organizational first-party loss impact due an APT breach is: *what is the nature of multi-party aggregate cyber-loss impact incurred in a supply chain environment of inter-dependent IIoT-networked ICSs?* We provide insights on this question via a general rigorous model-based analysis. In this paper, we only consider the supply chain impact due to an APT cyber-attack on one organization. In reality, the unique thing about an IIoT setting is that multiple organizations/enterprises can be affected simultaneously via a cyber-attack causing concurrent supply chain impacts. As an example this could happen when firmware, i.e., software embedded in IoT device's hardware or circuit board, that is often made by the same vendor for many IoT devices across an entire supply chain and quite vulnerable to cyber-attack due to unsecured code, is exploited by corporate adversaries or nation states. Our results in this section will show that the aggregate cyber-risk impact in networked supply


SCENARIO 1
DISRUPT CONTROLLERS TO SHUTDOWN OPERATIONS
SCENARIO 2
010
01010110
01
110
REPROGRAM CONTROLLERS TO SABOTAGE INDUSTRIAL PROCESSES
SCENARIO 3
DISABLE SAFETY CONTROLLERS TO CAUSE PHYSICAL DESTRUCTION
MANDIANT

The attacker leverages OMSHELL and/or CODECALL to crash PLCs, disrupt their performance, or otherwise impact their availability.

Combining process manipulations with asset disruption can signal an adversary's cyber attack capabilities, while minimizing the costly investment of studying a control system to develop a tailored cyber physical impact. The loss of availability of critical PLCs would require the impacted facility to shut down operations, resulting in delayed production, financial losses, and complex facility start up procedures.

The attacker reprograms or sends unauthorized commands to PLCs to alter the physical behavior of field devices and physical actuators, such as motors and pumps.

Depending on the nature of the victim facility and process manipulation, the change in controller behavior could result in defective products or malfunctioning machine behavior for a prolonged period.

The attacker disables PLCs responsible for safety functions, such as the Omron NX-SL3300, and subsequently reprograms or disrupts other ICS assets to cause physical destruction to the industrial machinery.

The loss of safety protection could allow the process to enter an unsafe state either naturally or through the attacker's manipulation of the process. This could cause impacts to human safety, the environment, or damage to equipment, depending on the physical constraints of the process and the facility design.

Fig. 11: Three APT Attack Scenarios of the Pipedream Malware (*Source:* Mandiant)

chain settings due to an APT cyber-attack on a single enterprise is sufficient to assess the aggregate cyber-risk impact due to simultaneous APT cyber-attacks on multiple enterprises.

An Interdependent Supply Chain Network-Wide QoS Model - Inspired by the contribution in [64], let the tangible performance output (covers the span of general organizational performance metrics) of any individual IIoT-driven organization i in an inter-dependent service network reflect the quality of service (QoS) for i. Let this performance output be modeled by the *Cobb-Douglas* function $q_i = z_i^{\alpha} l_i^{\alpha} \Pi_{j=1}^{n} q_{ij}^{(1-\alpha)c_{ij}}$, where l_i is the amount of human resources employed by i; q_{ij} is the amount of QoS i receives from the functioning of j (is directly a function of i's intra-organizational cyber-security posture) on which the former is reliant upon; $c_{ij} \geq 0$ denotes the degree of reliance of i on j with $\sum_j c_{ij} = 1$; z_i is the adverse impact on organization i's QoS of an APT cyber-breach event on organization i, where z_i is assumed to be independently distributed across i through $F_i \equiv \log(z_i)$; and α is the output elasticity of resource invested to generate performance (QoS). The rationale for using the Cobb-Douglas function to derive an organization's output metric stems from labor studies that use it as a standard model. An example of QoS q_i for i is the service-driven revenue for organization i. We define the weighted centrality vector for each organization in the service network of IIoT-driven organizations through $v \equiv \frac{\alpha}{n}[I-(1-\alpha)C^T]^{-1}\vec{1}$, where $[I-(1-\alpha)C^T]^{-1}$ is the Leontief inverse. We also assume that $\mathbb{E}[f_i] = 0$ (normalization), and $\mathrm{Var}[f_i]$ is finite (we will revisit the implication of this assumption later) for each organizations i, where f_i is the pdf of the adverse impact distribution on i caused by cyber-breach incident(s). It is evident that the variation in the system-wide QoS is given by $y = v^T f$, where $f = [f_1, .., f_n]$. In this paper, we are interested to investigate $\sqrt{\mathrm{Var}[y]}$.

Supply Chain Cyber-Risk Management Scalability - We have the following result from our proposed model analyzing $\sqrt{\mathrm{Var}[y]}$ (an aggregate cyber-risk manager's indicator of management scalability in an economic sense).

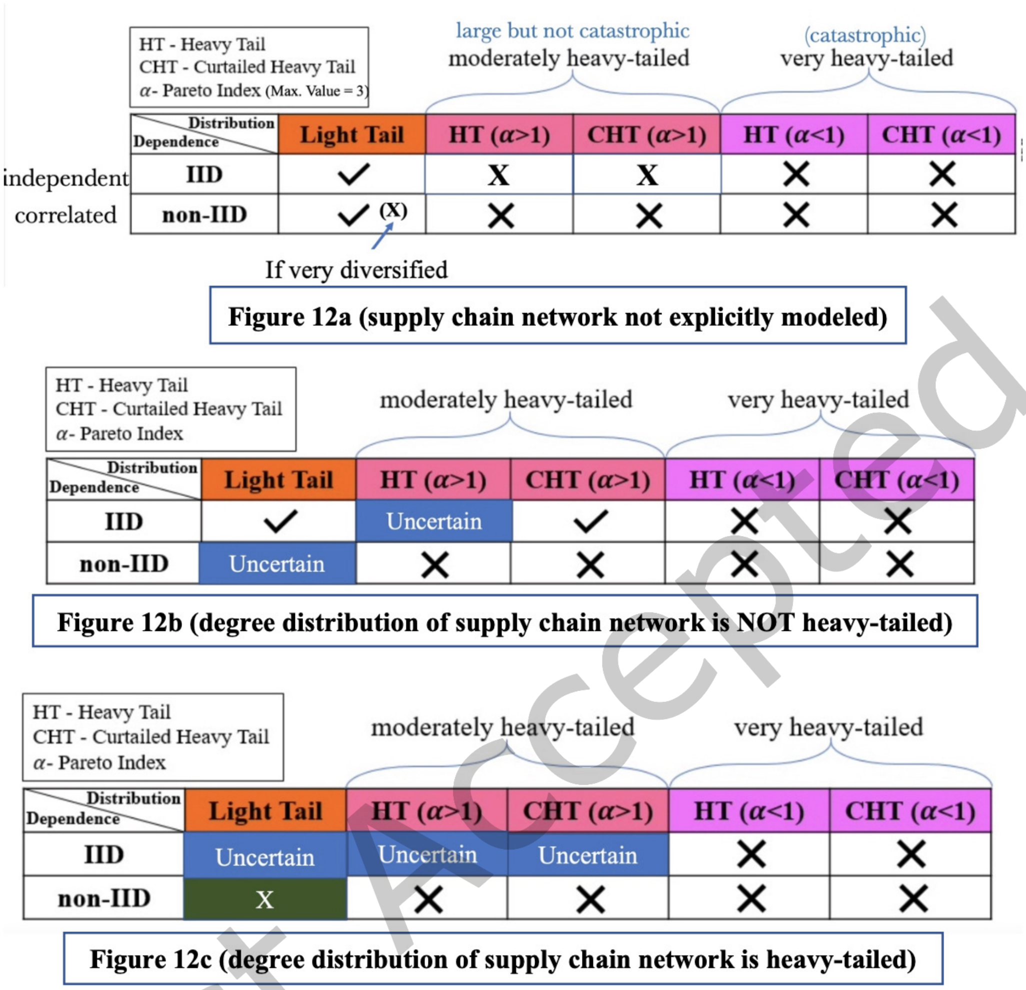
Theorem 8.1. *Let C be an inter-dependent organizational service network driven by IIoT-driven ICS enterprises. Under the event of cyber-breach event(s) whose first-party adverse impact on individual organizations is independently governed by f_i for each i, the standard deviation of the adverse impact distribution on the network-wide QoS, denoted by $\sqrt{Var[y]}$ follows the relation: $\sqrt{Var[y]} = \Omega\left(\frac{1+\frac{1}{\bar{d}}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})^2}}{\sqrt{n}}\right)$, where $\bar{d}$ is average degree of all organizational nodes, and $d_i = \sum_j c_{ji}$ is the degree of organizational node i. Moreover, if there exists a constant $\beta \in (1,2)$, a function $H(\cdot)$ such that $\lim_{t\to\infty} H(t)t^{\delta} = \infty$ and $\lim_{t\to\infty} H(t)t^{-\delta} = 0$ for all $\delta > 0$; a constant c; all $k < d_{\max} = \Theta(n^{\frac{1}{\beta}}$; and $P(k) = \frac{1}{n}ck^{-\beta}H(k)$, then $\sqrt{Var[y]} = \Omega\left(n^{\frac{-(\beta-1)}{\beta-\delta}}\right)$ for arbitrary $\delta > 0$.*

Theoretical Implications to Cyber-Risk Management - First and foremost, the theorem suggests that in the most ideal case when f_i's are independent, the standard deviation in the network-wide QoS is significantly affected post an APT cyber-breach event (only on one organization) if the distribution of inter-liability (measured through the individual node degrees of graph induced by matrix C) between organizations have a heavy tail. *In other words, in the absence of the said heavy-tailed property, aggregate cyber-loss coverage contracts post claims made as an aftermath of APT cyber-breach incidents are information asymmetry resilient.*

Alternatively, if the organizational client span of a cyber-risk manager (e.g., a cyber (re-)insurer) forms a supply chain network with a degree distribution that is not heavy-tailed, the aggregate CRM business will be network scalable (in an economic sense) under independent f_i's - irrespective of whether the cyber-risk manager has perfect information on intra-organizational cyber-loss distributions or otherwise. Here, the heavy-tailed property is ensured through the conditions: the existence of a function $H(\cdot)$ such that $\lim_{t\to\infty} H(t)t^{\delta} = \infty$ and $\lim_{t\to\infty} H(t)t^{-\delta} = 0$ for all $\delta > 0$; a constant c; all $k < d_{\max} = \Theta(n^{\frac{1}{\beta}}$; and $P(k) = \frac{1}{n}ck^{-\beta}H(k)$. Note here that $\beta \in (1,2)$ conservatively only captures the family of power-law distributions, that are specific milder instances of heavy-tailed distributions.

Practical Implications to Cyber-Risk Management - The independence assumption behind the f_i's is a conservative one when assesses the real world. In reality the f_i's are statistically dependent. i.e., correlated, and might result in high values of $\sqrt{\mathrm{Var}[y]}$ even for non-heavy tailed inter-liability cyber-risk distributions (ones that are 'normal' and not catastrophic). In the context of IIoTs, the f_i's will be quite dependent under the effect of common vulnerability driven APTs launched on IIoT-driven ICSs. In addition, and realistically enough, an APT cyber-attack launched on one organization due to firmware vulnerabilities mentioned above can very likely be concurrently launched on other IIoT driven organizations (on the same supply chain or on different ones) with common firmware. This would cause multiple f_i's on multiple supply chains to arise at the same time and all being mutually dependent to some degree.

These network-centric results extend and complement the aggregate cyber-risk coverage feasibility results for general non-networked settings, as discussed in [88][89][87]. We summarize our complete formally-driven insights on the economic scalability of aggregate CRM for IIoT (networked) supply chain environments in Figure 12. *We infer that a sufficient condition for economically scalable aggregate CRM to be feasible is for to-be-aggregated cyber-loss impact distributions sourcing from multiple IIoT-networked organizations to be (a) few in number, (b) independent, and (c) moderately heavy-tailed.* This is probably too ideal a situation cyber-insurers could hope for. This leads us to the next section where we discuss action items for managers of IIoT driven enterprises to mitigate first-party cyber-risk exposure so as to result in a situation where supply chain ecosystems expose cyber (re-)insurers to aggregate light-tailed cyber-risks.



HT - Heavy Tail
CHT - Curtailed Heavy Tail
α- Pareto Index (Max. Value = 3)

large but not catastrophic
moderately heavy-tailed
(catastrophic)
very heavy-tailed

	Dependence \ Distribution	Light Tail	HT (α>1)	CHT (α>1)	HT (α<1)	CHT (α<1)
independent	IID	✓	X	X	✕	✕
correlated	non-IID	✓ (X)	✕	✕	✕	✕

If very diversified

Figure 12a (supply chain network not explicitly modeled)

HT - Heavy Tail
CHT - Curtailed Heavy Tail
α- Pareto Index

moderately heavy-tailed
very heavy-tailed

Dependence \ Distribution	Light Tail	HT (α>1)	CHT (α>1)	HT (α<1)	CHT (α<1)
IID	✓	Uncertain	✓	✕	✕
non-IID	Uncertain	✕	✕	✕	✕

Figure 12b (degree distribution of supply chain network is NOT heavy-tailed)

HT - Heavy Tail
CHT - Curtailed Heavy Tail
α- Pareto Index

Dependence \ Distribution	Light Tail	HT (α>1)	CHT (α>1)	HT (α<1)	CHT (α<1)
IID	Uncertain	Uncertain	Uncertain	✕	✕
non-IID	X	✕	✕	✕	✕

Figure 12c (degree distribution of supply chain network is heavy-tailed)

Fig. 12: Illustrating the case-based economic feasibility of scaling aggregate cyber-risk management businesses (as a function of the number of risk sources of various statistical types and dependencies) for a supply chain network of IIoT-driven ICSs when (a) the network is not explicitly modeled, (b) the degree distribution of the network is not heavy-tailed, and (c) the degree distribution of the network is heavy-tailed.

9 ACTION ITEMS FOR IIOT DRIVEN ENTERPRISE MANAGEMENT

In this section, we propose action items for the managerial suite of IIoT-networked ICS organizations to minimize the adverse cyber-loss impact of an APT cyber-breach on organizational and supply chain business performance metrics. Many of our action items also extends to IT driven enterprises.

Action Items for Technology Management - Technology management within an ICS involves 'managing' the workings of (a) the communication protocols of the IIoT network, (b) the IIoT devices and associated architectures, and (c) adversary modules to minimize chances of cyber-breach events and their impact. More specifically, ICS managers should broadly adopt the following cyber-risk management strategies: (i) perimeter hardening, (ii) network hardening, (iii) workstation hardening, and (iv) device protection and hardening.

Perimeter hardening includes limiting outsider access to networks comprising ICS IIoT devices, placing these IIoT devices behind firewalls, extra-secure ICS IIoT devices facing the Internet, and continually monitoring for events that might indicate attempted unauthorized access. This includes inspecting traffic between systems within a data center or cloud service, and traffic seeking access to them.

Network hardening includes implementing secure access controls, disabling unused communication ports and protocols, segmenting networks to reduce an adversary's ability to move laterally across the network and compromise assets, using secure methods for remote access (especially in the work-from-home (WFH) age), and setting up measures (e.g., IDSs, IPSs, antivirus, usage logs) to detect network compromises. Both perimeter and network hardening should involve the implementation of a robust zero trust solution.

ICS workstation hardening includes implementing strong authentication controls (e.g., changing default passwords, using multi-factor authentication, implementing the account lockout feature on multiple wrong password guesses. enhance application and browser controls for improved protection), setting blocklists and allowlists to deny access to suspicious and/or malicious entities and to keep ICSs safe from unwanted software, respectively, and encourage safe and secure workstation use habits (e.g., scanning external hard drives and USB devices, storing sensitive workstation data on servers/cloud, locking workstation screens when idle).

IIoT *device protection and hardening* includes installing physical control to prevent unauthorized access, and tracking operation modes (e.g., keeping PLCs in RUN mode, otherwise overseeing whether alarms informing ICS operators are working). In addition, both workstation and device protection and hardening processes should (a) ideally use passwordless solutions (meeting the Faster Identity Online, i.e., FIDO, Alliance standard) that binds sign-on credentials in the Trusted Platform Module (TPM), (b) perform timely cleanup of stale executables on individual devices, (c) enable firmware scanning tools, memory integrity, and Secure Boot to shield from advanced firmware attacks, (d) enable memory access protection to prevent malicious hardware/software plug-ins, and (e) use dashboards related to security information and event management (SIEM) and security orchestration and response (SOAR) to monitor for anomalous or unauthorized behaviors.

Action Items for General Management - Identifying the government to be the 'highest general manager' overseeing enterprise cyber-risk management, we recommend the following action items for the general governments to be imposed upon enterprises and vendors around the globe: (i) mandatory cyber-reporting of incidents within the first 24-72 hours, (ii) implementing cyber-security contingency and response plan to ransomware and other major threats to to IT and OT systems, (iii) requiring software vendors to leverage a secure software development lifecycle and providing a software bill of materials (SBOM), (iv) mandating device manufacturers to coordinate vulnerability disclosure processes for released products.

With respect to action items on a higher managerial, i.e., board and upper management, level, as its most important broad action items to minimize the adverse cyber-loss impact of (APT) cyber-breach incidents on ICSs and their contributed supply chains, a CPS-driven ICS organization's leadership should *(1) make security a just cause by building a strong and compliant cyber-security culture, (2) invest high in residual cyber-risk mitigation methods by forming a cyber-risk division that always mandates buying insurance, and (3) encourage systemic cyber-risk resilience.*

In order to make security a just cause within a CPS-driven ICS organization, its leadership must hardwire cyber-risk thinking in OT/IT strategy by integrating security features in the design of CPS/Cloud components and processes, instead of just bolting security as an ad-on property. As an example, the security design of CPS elements should adhere to NIST guidelines. The enterprise leadership should fortify employee security knowledge through (a) a required multi-phased training reinforcement approach to support desired behavior outcomes, (b) preparing employees to handle or be resilient to business disrupting cyber-attacks via effectively designed business continuity and disaster recovery exercises, and (c) more importantly, educating business owners about cyber-risk and its impact on business process and performance so that they can take appropriate accountability for such risks. It should also promote a one-stop centralized web resource for best practices and internal cyber-incident reporting. Moreover, the enterprise leadership should have a mechanism in place to measure the efficacy (i.e., quality) and maximize the efficiency (i.e., quantity) of cyber-security skilling programs and their outcomes. At the same time, it is well-known that security is often inversely proportional to service appeal on customers (and hence is often viewed as an obstacle or afterthought in business processes) – hence the organization must strike a good profit and security culture tradeoffs by identifying only the most important IIoT driven CPS elements, i.e., the ICS chokepoints, to integrate security within design.

The C-suite and board should also budget high on investing in tech-security controls such as (a) hiring few but high quality and experienced security engineers to design and manage (basing upon STAMP [53], COA matrix, and CARVER frameworks) software-defined cyber-resilience processes covering the entire span of an ICS, (b) ensure IoT elements within an ICS do not run on default passwords and are periodically updated and security patched, (c) enable the proper use of antivirus, firewalls, and password managers, and (d) identifying critical business assets. Finally, the C-suite and board should promote security leadership to embrace security as a fun 'to-do' controls exercise (via incentives and gamification exercises) for employees at all levels of a CPS-driven organization. Specific to CPS settings, these controls would include employees using strong passwords and antiviruses for IoT devices; using secure communication settings while operating organization CPS elements over the Internet and/or the cloud while working remotely (e.g., as in the COVID age); maintaining software patch and configuration management spanning CPS components and processes; ensuring system backups with periodic data restores; and conducting periodic internal security audits and penetration scanning.

The C-suite and the board of a CPS-driven ICS organization should necessarily invest in the formation of a cyber-risk division to accurately access and mitigate the organization's exposure to cyber-risk and subsequently attract favorable coverage policies from a cyber-insurer. This initiative would involve (a) hiring a specialist chief risk officer (CRO) who should oversee periodic audit and security benchmarking activities on the CPS components inside an ICS that closely embrace NIST, CARVER, and STAMP-like cyber-safety frameworks, (b) hiring few specialized CPS cyber-risk quantification experts capable of formally accounting for component and process interdependencies and risk correlations in (large) CPS network settings to derive long-term cyber-damages, i.e., risk estimates post CPS attacks, (c) post conducting analysis on (b), adjudicating whether an organizational re-insurer would find it feasible to diversify aggregate first and third-party cyber-risks incurred by the organization post a cyber-attack event, and subsequently (d) working with the C-Suite, HR, and organizational psychologists to promote security best practices among employees handling CPS components and processes to reduce aggregate cyber-risk incurred, via methods of gamification and incentives.

Cyber-vulnerability information sharing by individual organizations is a must for cyber-insurers to appropriately price supply chain induced systemic risk in a society formed due to the networked interaction of multiple ICS networks serving diverse application sectors. To this end the C-suite should promote their organization sharing best cyber-risk governance practices among business partners; encourage various management divisions to participate in CPS cyber-threat information sharing platforms; and cooperate with insurers to release to the relevant public – accurate CPS cyber-vulnerability information in easy-to-understand, structured, and quantifiable formats using the MITRE ATT&CK framework as an example. The USA in particular has already started to take (CPS) cyber-vulnerability information sharing seriously in sectors with critical infrastructure with the recent introduction of a law signed by US President Joe Biden on March 15, 2022. According to this law, organizations must report cyber-incidents to the Cybersecurity and Infrastructure Security Agency (CISA) of the US Department of Homeland Security (DHS) within 72 hours, with an obligation to report ransomware payments within 24 hours.

10 RELATED WORK

The practical importance of cyber-risk quantification activities for IIoT networks is a recent phenomenon, especially post the wake of major targeted cyber-attacks in the last few years [126][125]. To the best of our knowledge, we are not aware of any research effort till date that undertakes a rigorous formal characterization of APT cyber-risk measures (e.g., APT risk measured via a CVaR metric) for (I)IoT networks through providing provable risk performance guarantees. To this end, ours is the first effort filling this gap.

Though four recent research efforts [90][93][87][88][92], that are focused on analytically analyzing cyber-risk in IoT societies might look quite related, there is a significant difference in their goal and ours in this paper. Our main focus in this paper is accurately estimating the APT risk apriori due to the time and node aggregate malware-induced cyber-loss distribution in an IIoT network. On the other hand, the authors in [90][93][87][88][92][91] study statistical, algorithmic, and economic feasibility conditions under which aggregate cyber-risk management services like re-insurance can be profitable. Moreover, the main focus of our work deals with intra-organizational IIoT networks, whereas [90][93][87][88][92] mainly deal with supply-chain service networks between different intra-organizational IIoT networks.

Given the inter-disciplinary building blocks to our methodology, we briefly survey related work covering the *four* broad areas much related to our research: (i) cyber-epidemic spread processes, (ii) analytical characterization of cyber-risk, and (iii) cyber-risk characterization in critical IIoT-driven systems, and (iv) the FAIR model to characterize enterprise cyber-risk.

Cyber-Epidemic Spread Processes - The use of the network-based SIS model (also known as contact processes), as a continuous time Markov chain, to model an epidemic spread is a standard in computer epidemiology studies, and has been used and analyzed in multiple research efforts [62][63][14][37][76][9][113][114][95][122] - be it related to cyber-infections, or otherwise. There has been other considerable literature (not necessarily on SIS dynamics) on the dynamics of spread of computer malware in (a) homogeneous networks, i.e., complete graphs, [52][28], (ii) specific heterogeneous (e.g., power-law) networks [12][79][11][41][121][120][25][81][80][74][75], and (iii) arbitrary heterogeneous networks [118][44][27]. In recent years, there has been a number of efforts modeling the spread of stealthy cyber-malware (e.g., APTs) in arbitrary communication networks using approaches seeded by the SIS methodology [131][132][130][129][127][128][123].

However, none of these above-mentioned works on the dynamics of spread processes account for the time and node aggregate cyber-loss impact post a cyber-breach incident. In this paper, we integrate the epidemic modeling of stealthy cyber-attacks with an rigorous formal characterization of the time and node aggrgegate cyber-loss impact post an APT cyber-attack event.

Analytical Characterization of Cyber-Risk - With respect to analytically characterizing moments of the cyber-loss impact due to a Markovian spread process, a mean-field approximation approach, as in our work, has been proposed in [113]. The benefit of using such as an approach is its power to capture the complete structure of the IIoT network, rather than only average degrees - as formalized in degree-based mean field approaches [96][17]. Moreover, such a complete network-based approach enables us to analyze the influence of the network topology on the spread of the infection and also on the total cyber-loss impact accrued in an IIoT network. *In addition, unlike us, the works in [96][17][114] do not characterize the time-dependent accuracy of mean-field approximation estimates.*

It is widely known from general risk management studies that though *statistical variance* may be a suitable cyber-risk measure when the corresponding risk distribution is close to be symmetric, it is not an appropriate measure in the general case where the distribution may be asymmetric (as in our work), as it equally penalizes losses and gains. The reader is referred to extensive literature on the shortcomings of the expected value risk measure [3][38][50][100]. *To overcome this limitation, we adopt* shortfall-based *or* quantile-based *risk measures that have rapidly gained wide popularity during the first decade of the 21st century.* The most used of such measures is the conditional value-at-risk (CVaR), first developed by the authors in [100], and its powerful mathematical properties investigated in [1][8]. While the asymptotic convergence properties of various estimators for CVaR have been investigated in [1], *less is known about finite-sample convergence properties for estimators.* The authors in [110] have shown some finite empirical convergence results for CVaR; although these results apply to the case of optimization of CVaR, *the bounds rely on statistical learning and, as a result, suffer from the conservatism of their theory.*

In contrast, our proposed(one-sided and two-sided) bounds based on empirical finite samples do not rely on statistical learning inputs. On providing two-sided concentration bounds for empirical finite-sample CVaR estimation, since the latter is a weighted average of the underlying distribution quantiles, one could employ concentration results for quantiles such as in [54]. *While such an approach can provide bounds with better constants, the resulting bounds also involve distribution-dependent quantities.* In contrast, our approach provides a unified method of proof that is distribution-agnostic.

Characterization of Cyber-Risk in Critical IIoT-Driven Systems - There have been quite a few works specifically related to characterizing failure risks in (IIoT-driven) critical infrastructure hosting organizations. The authors in [10] propose the cellular automata driven Abelian sandpile model that can be used to estimate the loss impact of cascading failures in critical infrastructure systems. Using the sandpile model, the authors in [26] show that (cyber-)loss impact post a cyber-breach event in an industrial power grid will exhibit a statistical power-law distribution. *However, the major drawbacks of the analysis in [26] is that (a) it does not generalize to a setting where multiple nodes within an IIoT network are directly (cyber-)infected to start with, and (b) it does not explicitly model recovery strength of network nodes post detection of their (cyber-)failure.* It is in fact the recovery dynamics that increases the likelihood of an IIoT-driven network system to incur a light-tailed first-party cyber-loss distribution post an APT cyber-breach event (that we show in our work), when compared to a heavy-tailed power-law distribution aforementioned.

The authors in [71] use the OPA theory proposed in [29][31][99], specific to electrical distribution systems, to derive first-party (cyber-)loss related performance statistics post a cascading failure event. Though the authors in [71] do not conclude on particular first party cyber-loss statistics, *two major drawbacks of their work are that (a) it does not generalize to a setting where multiple nodes within an IIoT network are directly (cyber-)infected to start with, and (b) it does not explicitly model recovery strength of network nodes post detection of their (cyber-)failure.* In [33], the authors propose the CASCADE framework that accounts for multiple (IoT) components inside an industrial system being directly (cyber-)infected initially through a uniform distribution, and derive statistics pertaining to the number of non-functioning components post a cascading impact of direct and indirect cyber-infection dynamics related to a (cyber-)breach event. In [32], the authors extend the CASCADE framework to model the initial distribution of directly cyber-infected components to be a Poisson distribution, and subsequently derive statistics pertaining to the number of non-functioning components post a cascading impact in an industrial system. *However, the major drawbacks of the contributions in [32][33] is (a) the lack of a dynamic time-dependent analysis that generally characterizes cascading processes in practice, and (b) the lack of statistics pertaining to the cyber-loss impact that commercial cyber-risk managers (e.g., cyber-insurance companies) are interested in.*

The authors in [34] model time dependency of cascading cyber-infections in an industrial (IIoT) network, and derive the probability distribution of the total number of failed components. *However, like [32][33], [34] does not derive statistics pertaining to the cyber-loss impact that commercial cyber-risk managers might be interested in.* In [39], the authors (in sync with our vision of quantifying ICS cyber-risk) present a system that calculates *Cyber Value-at-Risk* of an organisation. CVaR is a probabilistic density function for losses (not to be confused with the traditional CVaR metric) from cyber-incidents, for any given threats of interest and risk control practice. It takes into account varying effectiveness of security controls, the consequences for risk propagation through infrastructures, and the cyber-harms that result.

The major drawback of the contribution in [39] is the lack of generalization through a rigorous formal analysis to quantify cyber-risk for a broad family of intra-networked ICSs of different shapes, i.e., network structure, and sizes. More specifically, the authors in [39] do not model the recovery-induced spread process of cyber-malwares as a function of the size and shape of an ICS network (leave alone generalizing across a broad space of ICS networks) to derive network-aggregate cyber-loss impact, and its tail properties. The fallout of this drawback is a lack of principled insight on the statistics of quantified cyber-risk for general-enough ICSs. In this paper we explicitly model (a) a broadly generalized intra-organizational IIoT network as a function of its size and shape, (b) multiple initial directly infected nodes, and (c) the time-dependent cascading behavior of (cyber-)infection to derive cyber-loss impact statistics.

The FAIR Model to Characterize Enterprise Cyber-Risk - The *Factor Analysis of Information Risk* (FAIR) model is a cyber risk assessment tool that connects the cyber domain to the field of operational risk management and allows to calculate an expected losses following value at risk calculation approach [48]. FAIR provides a framework, an ontology and taxonomy supported with statistical means to do cyber risk quantification [49, 117] The core of this model is focused on the frequency of adversarial threat and the magnitude of impacted assets while considering the presence of vulnerabilities, the strength of the adversary, and the defenses in place [48, 119]. FAIR as cyber risk assessment tool is both accepted in practice [46] and science [35, 59, 60, 94, 104] It has been applied in many including Government-to-Citizen (G2C) e-services [35] cloud computing [104], smart grid [59, 60] and malware on mobile systems [94].

The traditional FAIR model is not without its potential limitations. *The model may introduce inaccuracies when using different distributions for input parameters [5, 42, 117] and does not handle heavy-tailed distributions. Moreover, FAIR does not fully account for either the dynamic complex nature of cyber risk [134], or the networked effects between cyber-risk targets.*

Recent work solves some of these limitations by using a complementary Bayesian network approach [117] or applying the FAIR Controls Analytics Model (FAIR-CAM) [51]. A complementary Bayesian network approach allows for using data sets with different distributions for input and increases goals and modelling purpose [117]. However, this approach does not handle heavy-tailed cyber-risk distributions and does not account for the networked effects between cyber-risk targets. FAIR-CAM allows to include a more detailed level of security controls relevant only to specific types of loss event scenarios [51]. These controls relate to quality in decision-making, managing the variance in operational performance, and affecting the magnitude of a loss [51].

In contrast to all the above-mentioned works, our novel version of the FAIR model handles heavy-tailed distributions, accounts for either dynamic complex nature of cyber risk, and captures the networked effects between cyber-risk targets.

11 SUMMARY

According to management guru Peter Drucker, there is no effective (cyber-risk) management without an accurate-enough (cyber-)risk assessment. The success (or lack thereof) of the entire cyber-risk management industry is pivoted upon the latter fact. In this paper, we proposed the first theoretical framework for ICS enterprise managers to accurately estimate *apriori* (to the occurrence of cyber-attack events) and tightly bound APT risk in general IIoT driven ICS networks for a parametric family of stealthy spread-based APT cyber-attacks. We first modeled the time-varying *attack-defense-impact* trio pertaining to our threat model as a Markov-Feller continuous stochastic process. We then rigorously evaluated and tightly bounded the mean of spread parameters at an equilibrium of the spread process. Subsequently, we provided (a) a closed form expression for the node and time aggregate cyber-loss impact in an IIoT network due to a spread-based APT cyber-attack - a random variable of interest to cyber-risk managers (e.g., insurers) and a novel extension of the FAIR model for networked settings, and (b) a computationally tractable first-order mean-field approximation method to scalably and accurately estimate the mean value of the aggregate cyber-loss impact distribution (we call it the APT risk distribution) in the network when the number of nodes in an IIoT network is large.

The mean metric is often not enough for a cyber-risk manager who also wants to study the tail of a cyber-risk distribution to gauge catastrophic cyber-risk impact effects. The popular metric in the risk management industry to study the tail is CVaR, which we synonymized in our work with the tail estimate of the APT cyber-risk distribution. To this end, we proposed a rigorous analysis motivated by concentration bounds from large deviation theory in probability, to derive tight non-asymptotic bounds of the difference between the empirical estimation of the APT risk measure (evaluated as the sample CVaR of the node and time aggregate cyber-loss impact distribution) and the true (ground truth) value of the APT risk. We complemented our theory with trace-driven Monte Carlo simulations based upon IoT test-bed experiments run on the FIT IoT-Lab.

To illustrate the society-facing impact of APT cyber-breaches on IIoT-driven industries, we showed via theory that despite intra-organization breaches inducing a light-tailed first-party cyber-loss distribution on a single IIoT-networked organization, aggregate multi-party losses incurred by the same in supply chain service settings of interdependent IIoT-networked organizations could be heavy-tailed - posing significant challenges to commercial cyber-risk management businesses such as stand-alone cyber-insurance. We subsequently proposed managerial action items to mitigate the non-aggregate cyber-risk exposure emanating from any given IIoT-driven organization to minimize supply chain induced aggregate cyber-risk incurred by an IIoT society.

ACKNOWLEDGMENTS

The listing of authors roughly follow a decreasing order of contributions. R. Pal is the senior and corresponding author. R. Pal and R. Sequeira are the lead authors. The work is supported by internal funds available from MIT CAMS. Parts of text in Section 7 related to the ICS case study is borrowed from Internet sources in a fair use fashion. A few symbols and model constructs in this work have been borrowed from [40] for the reasons of consistency and self-sufficiency.

REFERENCES

[1] Carlo Acerbi. 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance* 26, 7 (2002), 1505–1518.

[2] Cedric Adjih, Emmanuel Baccelli, Eric Fleury, Gaetan Harter, Nathalie Mitton, Thomas Noel, Roger Pissard-Gibollet, Frederic Saint-Marcel, Guillaume Schreiner, Julien Vandaele, and others. 2015. FIT IoT-LAB: A large scale open experimental IoT testbed. In *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*. IEEE, 459–464.

[3] Maurice Allais. 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society* (1953), 503–546.

[4] Pierre Alquier, Vincent Cottet, and Guillaume Lecué. 2019. Estimation bounds and sharp oracle inequalities of regularized procedures with Lipschitz loss functions. *The Annals of Statistics* 47, 4 (2019), 2117–2144.

[5] Chris Anderson and Mia Poletto Andersson. 2013. *Long tail.* Bonnier fakta.

[6] Ross Anderson and Tyler Moore. 2009. Information security: where computer science, economics and psychology meet. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367, 1898 (2009), 2717–2727.

[7] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J Alex Halderman, Luca Invernizzi, Michalis Kallitsis, and others. 2017. Understanding the mirai botnet. In *26th {USENIX} security symposium ({USENIX} Security 17)*. 1093–1110.

[8] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. Coherent measures of risk. *Mathematical finance* 9, 3 (1999), 203–228.

[9] Norman TJ Bailey and others. 1975. *The mathematical theory of infectious diseases and its applications.* Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.

[10] Per Bak, Chao Tang, and Kurt Wiesenfeld. 1988. Self-organized criticality. *Physical review A* 38, 1 (1988), 364.

[11] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.

[12] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. 2008. *Dynamical processes on complex networks.* Cambridge university press.

[13] Aharon Ben-Tal and Marc Teboulle. 2007. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance* 17, 3 (2007), 449–476.

[14] Carol Bezuidenhout and Geoffrey Grimmett. 1990. The critical contact process dies out. *The Annals of Probability* (1990), 1462–1482.

[15] Sanjay P Bhat and LA Prashanth. 2019. Concentration of risk measures: A Wasserstein distance approach. In *Advances in Neural Information Processing Systems*. 11762–11771.

[16] Christian Biener, Martin Eling, and Jan Hendrik Wirfs. 2015. Insurability of cyber risk: An empirical analysis. *The Geneva Papers on Risk and Insurance-Issues and Practice* 40, 1 (2015), 131–158.

[17] Marián Boguná and Romualdo Pastor-Satorras. 2002. Epidemic spreading in correlated complex networks. *Physical Review E* 66, 4 (2002), 047104.

[18] Jean-Chrysostome Bolot and Marc Lelarge. 2008. A new perspective on internet security using insurance. In *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. IEEE, 1948–1956.

[19] P. B. Bonacich. 1987. Power and Centrality: A Family Of Measures. *Amer. J. Sociology* 92 (1987).

[20] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. 2013. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

[21] Pierre Brémaud. 1981. *Point processes and queues: martingale dynamics*. Vol. 50. Springer.

[22] Bruno Brosowski and Frank Deutsch. 1981. An elementary proof of the Stone-Weierstrass theorem. *Proc. Amer. Math. Soc.* (1981), 89–92.

[23] David B Brown. 2007. Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters* 35, 6 (2007), 722–730.

[24] Shawn A Butler. 2002. Security attribute evaluation method: a cost-benefit approach. In *Proceedings of the 24th international conference on Software engineering*. 232–240.

[25] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. 2000. Network robustness and fragility: Percolation on random graphs. *Physical review letters* 85, 25 (2000), 5468.

[26] Benjamin A Carreras, Vickie E Lynch, Ian Dobson, and David E Newman. 2002. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos: An interdisciplinary journal of nonlinear science* 12, 4 (2002), 985–994.

[27] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. 2008. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)* 10, 4 (2008), 1–26.

[28] Zesheng Chen and Chuanyi Ji. 2005. A self-learning worm using importance scanning. In *Proceedings of the 2005 ACM workshop on Rapid malcode*. 22–29.

[29] Alexander E David, Blazhe Gjorgiev, and Giovanni Sansavini. 2020. Quantitative comparison of cascading failure models for risk-based decision making in power systems. *Reliability Engineering & System Safety* 198 (2020), 106877.

[30] Kenneth R Davidson and Allan P Donsig. 2009. *Real analysis and applications: theory in practice*. Springer Science & Business Media.

[31] Ian Dobson, Benjamin A Carreras, Vickie E Lynch, and David E Newman. 2001. An initial model for complex dynamics in electric power system blackouts. In *hicss*.

[32] Ian Dobson, Benjamin A Carreras, and David E Newman. 2004. A branching process approximation to cascading load-dependent system failure. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. IEEE, 10–pp.

[33] Ian Dobson, Benjamin A Carreras, and David E Newman. 2005. A loading-dependent model of probabilistic cascading failure. *Probability in the Engineering and Informational Sciences* 19, 1 (2005), 15–32.

[34] Hui Dong and Lirong Cui. 2015. System reliability under cascading failure models. *IEEE Transactions on Reliability* 65, 2 (2015), 929–940.

[35] Richard Dreyling, Eric Jackson, and Ingrid Pappel. 2021. Cyber security risk analysis for a virtual assistant G2C digital service using FAIR model. In *2021 Eighth International Conference on eDemocracy & eGovernment (ICEDEG)*. IEEE, 33–40.

[36] Richard M Dudley. 2014. *Uniform central limit theorems*. Vol. 142. Cambridge university press.

[37] Richard Durrett and Xiu-Fang Liu. 1988. The contact process on a finite set. *The Annals of Probability* (1988), 1158–1173.

[38] Daniel Ellsberg. 1961. Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics* (1961), 643–669.

[39] Arnau Erola, Ioannis Agrafiotis, Jason RC Nurse, Louise Axon, Michael Goldsmith, and Sadie Creese. 2022. A system to calculate cyber-value-at-risk. *Computers & Security* 113 (2022), 102545.

[40] Matthias A Fahrenwaldt, Stefan Weber, and Kerstin Weske. 2018. Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin: The Journal of the IAA* 48, 3 (2018), 1175–1218.

[41] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. 1999. On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review*, Vol. 29. ACM, 251–262.

[42] Sergey Foss, Dmitry Korshunov, Stan Zachary, and others. 2011. *An introduction to heavy-tailed and subexponential distributions*. Vol. 6. Springer.

[43] Jack Freund and Jack Jones. 2014. *Measuring and managing information risk: a FAIR approach*. Butterworth-Heinemann.

[44] Ayalvadi Ganesh, Laurent Massoulié, and Don Towsley. 2005. The effect of network topology on the spread of epidemics. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, Vol. 2. IEEE, 1455–1466.

[45] Wassily Hoeffding. 1994. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*. Springer, 409–426.

[46] Isaca. 2009. *The risk IT framework.* ISACA.

[47] Mohammad S Jalali, Michael Siegel, and Stuart Madnick. 2019. Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. *The Journal of Strategic Information Systems* 28, 1 (2019), 66–82.

[48] Jack Jones. 2006. An introduction to factor analysis of information risk (fair). *Norwich Journal of Information Assurance* 2, 1 (2006), 67.

[49] Nathan Jones and Brian Tivnan. 2018. *Cyber Risk Metrics Survey, Assessment and Implementation Plan.* Technical Report. MITRE CORP BEDFORD MA.

[50] Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I.* World Scientific, 99–127.

[51] Kamalanathan Kandasamy, Sethuraman Srinivas, Krishnashree Achuthan, and Venkat P Rangan. 2020. IoT cyber risk: A holistic analysis of cyber risk assessment frameworks, risk vectors, and risk ranking process. *EURASIP Journal on Information Security* 2020, 1 (2020), 1–18.

[52] Jeffrey O Kephart and Steve R White. 1992. Directed-graph epidemiological models of computer viruses. In *Computation: the micro and the macro view.* World Scientific, 71–102.

[53] Shaharyar Khan and Stuart E Madnick. 2021. Cybersafety: A System-theoretic Approach to Identify Cyber-vulnerabilities & Mitigation Requirements in Industrial Control Systems. *IEEE Transactions on Dependable and Secure Computing* (2021).

[54] Ravi Kumar Kolla, LA Prashanth, Sanjay P Bhat, and Krishna Jagannathan. 2019. Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters* 47, 1 (2019), 16–20.

[55] Dragan Komljenovic, Mohamed Gaha, Georges Abdul-Nour, Christian Langheit, and Michel Bourgeois. 2016. Risks of extreme and rare events in Asset Management. *Safety science* 88 (2016), 129–145.

[56] Halima Ibrahim Kure and Shareeful Islam. 2019. Assets focus risk management framework for critical infrastructure cybersecurity risk management. *IET Cyber-Physical Systems: Theory & Applications* 4, 4 (2019), 332–340.

[57] Henry Lam and Haofeng Zhang. 2022. Prediction Intervals for Simulation Metamodeling. (2022).

[58] Aron Laszka, Gabor Horvath, Mark Felegyhazi, and Levente Buttyán. 2014. FlipThem: Modeling targeted attacks with FlipIt for multiple resources. In *International Conference on Decision and Game Theory for Security.* Springer, 175–194.

[59] Anhtuan Le, Yue Chen, Kok Keong Chai, Alexandr Vasenev, and Lorena Montoya. 2017. Assessing loss event frequencies of smart grid cyber threats: Encoding flexibility into fair using bayesian network approach. In *Smart Grid Inspired Future Technologies: First International Conference, SmartGIFT 2016, Liverpool, UK, May 19-20, 2016, Revised Selected Papers.* Springer, 43–51.

[60] Anhtuan Le, Yue Chen, Kok Keong Chai, Alexandr Vasenev, and Lorena Montoya. 2019. Incorporating FAIR into bayesian network for numerical assessment of loss event frequencies of smart grid cyber threats. *Mobile Networks and Applications* 24 (2019), 1713–1721.

[61] Marc Lelarge and Jean Bolot. 2009. Economic incentives to increase security in the internet: The case for insurance. In *IEEE INFOCOM 2009.* IEEE, 1494–1502.

[62] Thomas Milton Liggett. 2012. *Interacting particle systems.* Vol. 276. Springer Science & Business Media.

[63] Thomas M Liggett. 2013. *Stochastic interacting systems: contact, voter and exclusion processes.* Vol. 324. springer science & Business Media.

[64] John B Long Jr and Charles I Plosser. 1983. Real business cycles. *Journal of political Economy* 91, 1 (1983), 39–69.

[65] Gabor Lugosi and Shahar Mendelson. 2019. Risk minimization by median-of-means tournaments. *Journal of the European Mathematical Society* 22, 3 (2019), 925–965.

[66] Stuart Madnick, Mohammad S Jalali, Michael Siegel, Yang Lee, Diane Strong, Richard Wang, Wee Horng Ang, Vicki Deng, and Dinsha Mistree. 2016. Measuring stakeholders' perceptions of cybersecurity for renewable energy systems. In *International workshop on data analytics for renewable energy integration.* Springer, 67–77.

[67] Ignacio J Martinez-Moyano, Rogelio Oliva, Donald Morrison, and David Sallach. 2015. Modeling adversarial dynamics. In *2015 Winter Simulation Conference (WSC).* IEEE, 2412–2423.

[68] Timothée Mathieu and Stanislav Minsker. 2021. Excess risk bounds in robust empirical risk minimization. *Information and Inference: A Journal of the IMA* 10, 4 (2021), 1423–1490.

[69] Colin McDiarmid. 1989. On the method of bounded differences. *Surveys in combinatorics* 141, 1 (1989), 148–188.

[70] Alexander J McNeil, Rüdiger Frey, Paul Embrechts, and others. 2015. Quantitative risk management: Concepts. *Economics Books* (2015).

[71] Sheng-wei Mei, Xiao-feng Weng, An-cheng Xue, and others. 2006. Blackout model based on OPF and its self-organized criticality. In *2006 Chinese Control Conference.* IEEE, 1673–1678.

[72] JR Minkel. 2008. The 2003 Northeast Blackout–Five Years Later. *Scientific American* 13 (2008).

[73] Dimitar Minovski, Christer Åhlund, Karan Mitra, and Roman Zhohov. 2020. Defining Quality of Experience for the Internet of Things. *IT Professional* 22, 5 (2020), 62–70.

[74] Michael Molloy and Bruce Reed. 1995. A critical point for random graphs with a given degree sequence. *Random structures & algorithms* 6, 2-3 (1995), 161–180.

[75] Michael Molloy and Bruce Reed. 1998. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, probability and computing* 7, 3 (1998), 295–305.

[76] Thomas Mountford, Jean-Christophe Mourrat, Daniel Valesin, and Qiang Yao. 2016. Exponential extinction time of the contact process on finite graphs. *Stochastic Processes and their Applications* 126, 7 (2016), 1974–2013.

[77] Anand Mudgerikar, Puneet Sharma, and Elisa Bertino. 2019. E-spion: A system-level intrusion detection system for iot devices. In *proceedings of the 2019 ACM Asia conference on computer and communications security*. 493–500.

[78] Anand Mudgerikar, Puneet Sharma, and Elisa Bertino. 2020. Edge-Based Intrusion Detection for IoT devices. *ACM Transactions on Management Information Systems (TMIS)* 11, 4 (2020), 1–21.

[79] Mark Newman. 2018. *Networks*. Oxford university press.

[80] Mark EJ Newman. 2007. Component sizes in networks with arbitrary degree distributions. *Physical review e* 76, 4 (2007), 045101.

[81] Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. 2001. Random graphs with arbitrary degree distributions and their applications. *Physical review E* 64, 2 (2001), 026118.

[82] Yin Minn Pa Pa, Shogo Suzuki, Katsunari Yoshioka, Tsutomu Matsumoto, Takahiro Kasama, and Christian Rossow. 2015. IoTPOT: Analysing the rise of IoT compromises. In *9th {USENIX} Workshop on Offensive Technologies ({WOOT} 15)*.

[83] Ranjan Pal and Leana Golubchik. 2010. Analyzing self-defense investments in internet security under cyber-insurance coverage. In *2010 IEEE 30th International Conference on Distributed Computing Systems*. IEEE, 339–347.

[84] Ranjan Pal, Leana Golubchik, Konstantinos Psounis, and Pan Hui. 2014. Will cyber-insurance improve network security? A market analysis. In *INFOCOM, 2014 Proceedings IEEE*. IEEE, 235–243.

[85] Ranjan Pal, Leana Golubchik, Konstantinos Psounis, and Pan Hui. 2016. Security Pricing as an Enabler of Cyber-Insurance: A First Look at Differentiated Pricing Markets. *arXiv preprint arXiv:1607.02598* (2016).

[86] Ranjan Pal, Leana Golubchik, Konstantinos Psounis, and Pan Hui. 2018. Improving Cyber-Security via Profitable Insurance Markets. *ACM SIGMETRICS Performance Evaluation Review* 45, 4 (2018), 7–15.

[87] Ranjan Pal, Ziyuan Huang, Sergey Lototsky, Xinlong Yin, Mingyan Liu, Jon Crowcroft, Nishanth Sastry, Swades De, and Bodhibrata Nag. 2021. Will Catastrophic Cyber-Risk Aggregation Thrive in the IoT Age? A Cautionary Economics Tale for (Re-) Insurers and Likes. *ACM Transactions on Management Information Systems (TMIS)* 12, 2 (2021), 1–36.

[88] Ranjan Pal, Ziyuan Huang, Xinlong Yin, Mingyan Liu, Sergey Lototsky, and Jon Crowcroft. 2020. Sustainable catastrophic cyber-risk management in IoT societies. In *2020 Winter Simulation Conference (WSC)*. IEEE, 3105–3116.

[89] Ranjan Pal, Ziyuan Huang, Xinlong Yin, Sergey Lototsky, Swades De, Bodhibrata Nag, Mingyan Liu, Jon Crowcroft, and Nishanth Sastry. 2021. Will Catastrophic Cyber-Risk Management Thrive in the IoT Age?: A Cautionary Economics Tale for (Re) Insurers and Likes. *To Appear in ACM Transactions on Management Information Systems* (2021).

[90] Ranjan Pal, Ziyuan Huang, Xinlong Yin, Sergey Lototsky, Swades De, Sasu Tarkoma, Mingyan Liu, Jon Crowcroft, and Nishanth Sastry. 2020. Aggregate Cyber-Risk Management in the IoT Age: Cautionary Statistics for (Re) Insurers and Likes. *IEEE Internet of Things Journal* (2020).

[91] Ranjan Pal, Peihan Liu, Taoan Lu, and Edward Y Hua. 2022. How Hard is Cyber-Risk Management in IT/OT Systems? A Theory to Classify and Conquer Hardness of Insuring ICSs. *ACM Transactions on Cyber-Physical Systems* 6, 4 (2022).

[92] Ranjan Pal, Taoan Lu, Peihan Liu, and Xinlong Yin. 2021. Cyber (re-) insurance policy writing is NP-hard in IoT societies. In *2021 Winter Simulation Conference (WSC)*. IEEE, 1–12.

[93] Ranjan Pal, Konstantinos Psounis, Jon Crowcroft, Frank Kelly, Pan Hui, Sasu Tarkoma, Abhishek Kumar, John Kelly, Aritra Chatterjee, Leana Golubchik, and others. 2020. When Are Cyber Blackouts in Modern Service Networks Likely? A Network Oblivious Theory on Cyber (Re) Insurance Feasibility. *ACM Transactions on Management Information Systems (TMIS)* 11, 2 (2020), 1–38.

[94] Mookyu Park, Junwoo Seo, Jaehyeok Han, Haengrok Oh, and Kyungho Lee. 2018. Situational Awareness Framework for Threat Intelligence Measurement of Android Malware. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.* 9, 3 (2018), 25–38.

[95] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. 2015. Epidemic processes in complex networks. *Reviews of modern physics* 87, 3 (2015), 925.

[96] Romualdo Pastor-Satorras and Alessandro Vespignani. 2001. Epidemic dynamics and endemic states in complex networks. *Physical Review E* 63, 6 (2001), 066117.

[97] Shari Pfleeger and Robert Cunningham. 2010. Why measuring security is hard. *IEEE Security & Privacy* 8, 4 (2010), 46–54.

[98] Philip Potter. 2004. Stochastic Integration and Differential Equation. *Stochastic Modeling and Applied Probability* 21 (2004).

[99] Junjian Qi, Wenyun Ju, and Kai Sun. 2016. Estimating the propagation of interdependent cascading outages with multi-type branching processes. *IEEE Transactions on Power Systems* 32, 2 (2016), 1212–1223.

[100] R Tyrrell Rockafellar, Stanislav Uryasev, and others. 2000. Optimization of conditional value-at-risk. *Journal of risk* 2 (2000), 21–42.

[101] L Chris G Rogers and David Williams. 1994. Diffusions, Markov Processes and Martingales, Volume 1: Foundations. *John Wiley & Sons, Ltd., Chichester* 7 (1994).

[102] Sasha Romanosky, Lillian Ablon, Andreas Kuehn, and Therese Jones. 2019. Content analysis of cyber insurance policies: how do carriers price cyber risk? *Journal of Cybersecurity* 5, 1 (2019), tyz002.

[103] Sheldon M Ross. 2014. *Introduction to probability models*. Academic press.

[104] Alireza Shameli Sendi and Mohamed Cheriet. 2014. Cloud computing: A risk assessment model. In *2014 IEEE International Conference*

on Cloud Engineering. IEEE, 147–152.

[105] Daniel A Sepúlveda Estay. 2021. A system dynamics, epidemiological approach for high-level cyber-resilience to zero-day vulnerabilities. *Journal of Simulation* (2021), 1–16.

[106] René Serral-Gracià, Eduardo Cerqueira, Marilia Curado, Marcelo Yannuzzi, Edmundo Monteiro, and Xavier Masip-Bruin. 2010. An overview of quality of experience measurement challenges for video applications in IP networks. In *International Conference on Wired/Wireless Internet Communications*. Springer, 252–263.

[107] Nikhil Shetty, Galina Schwartz, Mark Felegyhazi, and Jean Walrand. 2010. Competitive cyber-insurance and internet security. In *Economics of information security and privacy*. Springer, 229–247.

[108] Detmar W Straub and Richard J Welke. 1998. Coping with systems risk: Security planning models for management decision making. *MIS quarterly* (1998), 441–469.

[109] Muhammad Suryanegara, Dimas Agung Prasetyo, Fery Andriyanto, and Nur Hayati. 2019. A 5-Step framework for measuring the Quality of Experience (QoE) of Internet of Things (IoT) services. *IEEE Access* 7 (2019), 175779–175792.

[110] A Takeda and T Kanamori. 2005. *A robust optimization approach based on conditional value-at-risk measure and its applications to statistical learning problems*. Technical Report. Working paper.

[111] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science* 185, 4157 (1974), 1124–1131.

[112] Marten Van Dijk, Ari Juels, Alina Oprea, and Ronald L Rivest. 2013. FlipIt: The game of ?stealthy takeover? *Journal of Cryptology* 26, 4 (2013), 655–713.

[113] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. 2008. Virus spread in networks. *IEEE/ACM Transactions On Networking* 17, 1 (2008), 1–14.

[114] P Van Mieghem and Ruud van de Bovenkamp. 2015. Accuracy criterion for the mean-field approximation in susceptible-infected-susceptible epidemics on networks. *Physical Review E* 91, 3 (2015), 032812.

[115] Cédric Villani. 2008. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media.

[116] Jack Wallen. 2017. Five nightmarish attacks that show the risks of IoT security. *URL: http://www. zdnet. com/article/5-nightmarish-attacks-that-show-the-risks-of-iot-security* (2017).

[117] Jiali Wang, Martin Neil, and Norman Fenton. 2020. A Bayesian network approach for cybersecurity risk assessment implementing and extending the FAIR model. *Computers & Security* 89 (2020), 101659.

[118] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. 2003. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings*. IEEE, 25–34.

[119] Gaute Wangen, Christoffer Hallstensen, and Einar Snekkenes. 2018. A framework for estimating information security risk assessment method completeness: Core Unified Risk Framework, CURF. *International Journal of Information Security* 17 (2018), 681–699.

[120] Herbert S Wilf. 2005. *generatingfunctionology*. CRC press.

[121] Walter Willinger, David Alderson, and John C Doyle. 2009. Mathematics and the internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society* 56, 5 (2009), 586–599.

[122] Daniel W Woods and Rainer Böhme. 2021. Systematization of Knowledge: Quantifying Cyber Risk. In *IEEE Symposium on Security & Privacy*.

[123] Yingbo Wu, Pengdeng Li, Lu-Xing Yang, Xiaofan Yang, and Yuan Yan Tang. 2017. A theoretical method for assessing disruptive computer viruses. *Physica A: Statistical Mechanics and its Applications* 482 (2017), 325–336.

[124] Jacob Wurm, Khoa Hoang, Orlando Arias, Ahmad-Reza Sadeghi, and Yier Jin. 2016. Security analysis on consumer and industrial IoT devices. In *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)*. IEEE, 519–524.

[125] Christos Xenofontos, Ioannis Zografopoulos, Charalambos Konstantinou, Alireza Jolfaei, Muhammad Khurram Khan, and Kim-Kwang Raymond Choo. 2021. Consumer, commercial and industrial iot (in) security: attack taxonomy and case studies. *IEEE Internet of Things Journal* (2021).

[126] Liudong Xing. 2020. Cascading failures in internet of things: review and perspectives on reliability and resilience. *IEEE Internet of Things Journal* 8, 1 (2020), 44–64.

[127] Shouhuai Xu, Wenlian Lu, and Li Xu. 2012. Push-and pull-based epidemic spreading in networks: Thresholds and deeper insights. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 7, 3 (2012), 1–26.

[128] Shouhuai Xu, Wenlian Lu, Li Xu, and Zhenxin Zhan. 2014. Adaptive epidemic dynamics in networks: Thresholds and control. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 8, 4 (2014), 1–19.

[129] Shouhuai Xu, Wenlian Lu, and Zhenxin Zhan. 2011. A stochastic model of multivirus dynamics. *IEEE Transactions on Dependable and Secure Computing* 9, 1 (2011), 30–45.

[130] Luxing Yang, Moez Draief, and Xiaofan Yang. 2017a. Heterogeneous virus propagation in networks: a theoretical study. *Mathematical Methods in the Applied Sciences* 40, 5 (2017), 1396–1413.

[131] Lu-Xing Yang, Moez Draief, and Xiaofan Yang. 2015. The impact of the network topology on the viral prevalence: a node-based approach. *PloS one* 10, 7 (2015), e0134507.

[132] Lu-Xing Yang, Pengdeng Li, Xiaofan Yang, and Yuan Yan Tang. 2017b. Distributed interaction between computer virus and patch: A modeling study. *arXiv preprint arXiv:1705.04818* (2017).

[133] Zichao Yang and John CS Lui. 2014. Security adoption and influence of cyber-insurance markets in heterogeneous networks. *Performance Evaluation* 74 (2014), 1–17.

[134] Sander Zeijlemaker and Michael Siegel. 2023. Capturing the Dynamic Nature of Cyber Risk: Evidence from an Explorative Case Study. (2023).

[135] S Zeijlemaker, JD Uriega, and G Pasaoglu Kilanc. 2018. Malware dynamics: how to develop a successful anti-malware defense reference architecture policy. In *Proceedings of the 36th International Conference of the System Dynamics Society, Reykjavik, Iceland*. [Sl: sn].