Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S.J.E. Taylor, eds.

AI ON SMALL AND NOISY DATA IS INEFFECTIVE FOR ICS CYBER RISK MANAGEMENT

Yaphet Lemiesa[1], Ranjan Pal[2], and Michael Siegel[2]

[1]Department of EECS, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

ABSTRACT

Modern industrial control systems (ICSs) are increasingly relying upon IoT and CPS technology to improve cost-effective service performance at scale. Consequently, the cyber vulnerability terrain is largely amplified in ICSs. Unfortunately, the historical lack of (a) sufficient, non-noisy ICS cyber incident data, and (b) intelligent operational business processes to collect and analyze available ICS cyber incident data, demands the attention of the Bayesian AI community to develop cyber risk management (CRM) tools to address these challenges. In this paper *we show with sufficient Monte Carlo simulation evidence that Bayesian AI on noisy (and small) ICS cyber incident data is ineffective for CRM.* More specifically, we show via a novel graphical sensitivity analysis methodology that even small amounts of statistical noise in cyber incident data are sufficient to reduce ICS intrusion/anomaly detection performance by a significant percentage. Hence, ICS management processes should strive to collect sufficient non-noisy cyber incident data.

1 INTRODUCTION

The modern industrial control system (ICS) is equipped with IoT and CPS technology to promote increasingly proactive and data-driven approaches to managing industrial operations. The primary (overlapping) benefits for ICS management to invest in such technology include: (a) real-time monitoring, response, and (automated) decision control of industrial and business processes, (b) improved operational visibility and asset tracking complementing remote work environments, (c) predictive sensor data driven maintenance to improve system reliability and equipment/process quality control, (d) data-driven energy optimization for reduced business operation costs, and (e) improved environmental, operational, and equipment safety to mitigate hazard externality and physical damage to people and equipments. It is not surprising then that the global IoT/CPS driven ICS market is projected (by *Statista*) to reach USD 275 billion by 2025, and to USD 450 billion by 2029 at a high CAGR of around 13.5%.

1.1 Cybersecurity Challenges in ICSs

While on one hand, the IoT and CPS technology brings in a large number of systems management benefits, it opens up a huge cyber vulnerability terrain spanned by hundreds and thousands of sensor and actuator devices. More specifically, large and complex (distributed) communication networks formed by sensor/actuator equipped ICS physical equipment have multiple critical vulnerable points in them. Research in applied computer science theory on systems networks have shown that in the worst case it is computationally infeasible to decipher all these critical points within practical time constraints (Pal et al. 2023; Pfleeger and Cunningham 2010; Pal et al. 2021; Pal et al. 2024). Add to this, budget constrained ICS managements (Dewri et al. 2007) in general are yet to invest a sufficient amount in cyber protection solutions and processes - partly because of their lack of awareness on cyber (increasing but insufficient) for a relatively new digitally new ICS industry. Moreover, there is a lack of seamless and easy-to-understand cyber risk KPIs across top-down ICS management to enable the board and upper management to approve increased expenditure on cybersecurity improvement and cyber risk management. As a consequence an ICS is exposed to a considerable exploitation risk of zero-day or other unknown-known vulnerabilities.

As examples (Stamp et al. 2009; Ten et al. 2011; Wang et al. 2013), adversaries could exploit such vulnerabilities to (a) gain root access of Human Machine Interface (HMI) devices in smart grid LANs operators to destabilize various aspects and critical parameters of the electrical system, including voltage levels, power flow, and circuit breaker status, (b) launch man-in-the-middle (MiTM) attacks to modify or falsify control data flow on communication links.

1.2 Cyber Risk Management Challenges in ICSs

The fundamental requirement for ICS management en route effective and practically deployable cyber risk management (CRM) is to understand the series of sequential steps that adversaries use to launch different types of cyber attacks on ICS infrastructure. Such knowledge is usually obtained (a) from domain experts in the ICS industry, and/or (b) a well documented central database (designed) and maintained by organizations such as MITRE through their popular MITRE ATT&CK for ICS database that is leading database of tactics, techniques, and procedures (TTPs) populated from real-world observations over multiple ICS industries around the USA and the world. Once such knowledge is obtained, the sequential nature of a cyber attack is traditionally transformed into an attack graph for analyzing how ICS network hosts and the communication links can be successfully compromised by target adversaries (Ammann et al. 2002; Sheyner et al. 2002). This analysis feeds into an anomaly detection (synonymous to detecting malicious intruder actions in the context of cybersecurity, as in this paper) exercise for an ICS management, the outcome of which supports CRM processes that subsume incident anticipation, response, and recovery.

Such anomaly detection tasks in ICSs is not a new thing, atleast for certain ICSs such as the smart/power grid. However, even for such systems, there are two challenges: (a) not all sensor-ed equipment (hosts) are continuously or periodically monitored, primarily due to lack of proper inventory management - this leading to the construction of 'incorrect' attack graphs, and (ii) the lack of sufficient historical data on cyber incidents on ICSs that leads to a lack of confidence in the anomaly detection chance outcome numbers that ICS management works with. Both (i) and (ii) are non-conducive to effect ICS CRM. As a result many significant feature variables across all the sensor-ed equipment that might together indicate (via the use of AI/ML ops) an anomaly within an ICS subsystem, is absent. This makes training AI and ML models on such data challenging to be able to perform anomaly pattern recognition substantially inaccurate (i.e., generates high false positives/negatives). It goes without saying that for ICSs that are relatively new to becoming sensor smart, a lack of inventory management combined with lack of data to run effective AI/ML ops makes anomaly detection and cyber risk management an uphill task.

1.3 Challenge-Mitigating (AI) Solutions

Traditional methodologies that can complement attack graph analysis to alleviate system uncertainty challenges such as the Inter-Domain Evidence Theoretic Approach for Inference (IDEA-I) that is based upon the seminal *Dempster-Shafer* Theory of Evidence (Li et al. 2012) is known to reduce false alarms towards the betterment of ICS CRM in certain settings (Sahu and Davis 2022). However, it suffers from the drawbacks of not being able to account for causal relationships between incident risk variables/features, finding it difficult to include domain expertise information in the decision making process, and quantitatively not suited for discrete modeling environments that often arise in practice. Without mitigating these drawbacks it is hard to see the benefits of *Demspter-Shafer* theory applications to generalizable and effective anomaly detection in ICS CRM.

There have been applications of *System-theoretic Accident Model and Processes* (STAMP) methodologies (developed at MIT by Leveson) for improving cybersecurity and cyber risk management (Leveson et al. 2003; Young and Leveson 2013). These applications are built upon the STAMP-driven *System-Theoretic Process Analysis* (STPA) and *Causal Analysis using Systems Theory* (CAST) methodologies for accidents and hazards. The STPA and CAST methodologies models an entire system and inter-component dependencies that can be studied to identify and mitigate the 'pain points' (i.e., system cyber vulnerabilities) of the

system that can be exploited by adversaries to cause system instability and cyber non-safety (Khan and Madnick 2021). In doing so, they alleviate a drawback of the *Demspter-Shafer* theory to account for causal relationships between system variables. However, the major drawback of STAMP applications is that they are not suited for modeling cyber attack sequence steps (e.g., those obtained from the MITRE ATT&CK ICS database) within individual components of a system at the network and protocol levels to identify anomalies. Alternatively, STAMP applications can be used to model an attack graph at a system network and protocol granularity (in an unnecessary complex manner), but such applications are not data-driven to smartly generate alarms (using AI/ML ops) to detect cyber anomalies for effective ICS CRM.

The field of Bayesian AI has been an industry and academic research favorite over the years to tackle CRM challenges mentioned in Section 1.2. and alleviating drawbacks mentioned in this subsection. Bayesian networks (a probabilistic inference tool in Bayesian AI) are probabilistic graphical models that helps (a) account for uncertainty in system (cyber) risk variables/parameters and incorporate domain expert knowledge and priors (Koller 2009; Pearl 1988), (b) can easily integrate cyber attack graphs at the network and protocol granularity (Pamula et al. 2006; Wang et al. 2017; Sun et al. 2018) into a Bayesian anomaly detection AI framework, (c) subsequently capture the dependencies between risk variables pertaining to a cyber attack TTP mentioned in MITRE ATT&CK like databases (Maccarone et al. 2022), and (d) can dynamically improve cyber anomaly inference performance via learning over time and with new network and protocol granularity data arriving over time (Sahu and Davis 2022).

1.4 Research Motivation and Contributions

It might seem that Bayesian networks with all its above-mentioned benefits is a standard methodology for effective CRM in ICSs - especially when Bayesian inference can be easily ported atop a cyber attack graph that is obtained from expert domain knowledge of ICS managers. However, all these above benefits are realized when there is sufficient data to train Bayesian AI models. On the contrary, the modern ICS ecosystem is mostly comprised of service industries that have relatively recently gotten on the IoT/CPS bandwagon, and most (a) do not have historical data about cyber incidents on their industries to populate cyber attack graphs for effective Bayesian inference on anomaly detection driven CRM tasks, and/or (b) are susceptible to increasing insider (and outsider) adversary manipulation (as reported by major ICS service firms such as *Nozomi Networks*, *Dragos*, and *Nanolock*) of (Bayesian AI) training data that might result in false negatives or excessive false positives (harmless alarms) on detecting malicious ICS intruders. It is evident that both (a) and (b) detrimental to the quality of CRM overall.

Research Motivation - The authors hypothesize, based on their 30+ combined years of experience in (Bayesian) statistics and working (and consulting) with industry research teams on causal cybersecurity analysis, that small and noisy data is too much of a challenge to design and deploy high performance Bayesian network driven CRM solutions rooted in anomaly detection. Hence, we are motivated to design a rigorous, systematic, and novel simulation framework that puts sufficient weight behind the hypothesis in a manner useful enough for ICS cyber risk managers to develop strategic action items improving CRM data collection and processing within a constrained budget.

Research Contributions - We make the following (research) contributions in this paper.

- We propose a Bayesian network inference model atop a cyber attack graph to detect network and protocol level anomalies in critical infrastructure environments. Our model is without loss of generality and scale, and adapted from existing literature on Bayesian AI based anomaly detection in IT systems. The sole goal behind this contribution is to (a) lay down how the steps of any (ICS) cyber attack can be sequentially modeled into an attack graph of cause and effect nodes (i.e., risk variables) that showcase the series of (pre)conditions under which an adversary target is compromised, and (b) populate the attack graph nodes with data-driven Bayesian network parameters that eventually decide the likelihood that an adversary target is compromised (see Section 3).

- To account for practical (worst case) ICS settings with a low number of (and potentially noisy) data samples driving Bayesian network parameters on a cyber attack graph, we first propose an inference methodology to detect traffic anomalies indicating adversarial ICS intrusion success, and follow it up with providing a closed form expression of the least number of non-noisy data samples, i.e., the *sample complexity*, per Bayesian network parameter, for effective inference on the graph. We then derive a rigorous graphical Monte Carlo sensitivity analysis simulation methodology that simultaneously accounts for (a) the topological centrality of a cyber attack graph nodes, (b) the statistical nature of adversarial noise injected into conditional probability tables (CPTs), and (c) the height of the attack graph at which an adversary injects noise in the CPTs, to generate the impact of noise on the anomaly detection (and consequently CRM) performance in ICSs (see Section 4).
- We show via the run of Monte Carlo simulations on a real-world Bayesian network formed atop a cyber attack graph of a miniature ICS sub-unit that Bayesian inference is too sensitive to CPT noise. In other words, *even a small amount of adversarial noise is enough to boost the anomaly detection error margins by a significant amount.* Our results are very conservative given that we simulate using a small Bayesian inference network. A scaled network is only going to non-linearly amplify inference error margins with adversarial noise. This further implies that ICS managers need to gather a very high number of accurate CPT entries (not the status quo in ICSs) for the Bayesian network nodes in order to gain leverage from Bayesian inference for CRM - else Bayesian AI is not the suitable technology for ICS CRM. We also show via our simulation exercise that topologically central nodes of the Bayesian network (which we will call as the Bayesian attack graph) are not that critical with respect to anomaly detection performance when compared to the relatively non-central leaf (lowest depth) and low-depth nodes. Hence, ICS managers must ensure that these low-depth nodes are populated with accurate CPTs for effective Bayesian inference (see Section 6).

We briefly study related work in Section 2, and summarize the paper in Section 6.

2 RELATED WORK

We discuss related work in the specific area of Bayesian network probabilistic inference and its relation to (ICS) IT and operational technology (OT) anomaly detection. In the context of cybersecurity (and in this paper), this is often synonymous to malicious intruder detection.

The Application of Bayesian AI in (ICS) Anomaly Detection - The field of Bayesian AI has been an industry and academic research favorite over the years to tackle anomaly detection problems in IT and OT systems. Bayesian networks (a probabilistic inference tool in Bayesian AI) are probabilistic graphical models that help (a) account for uncertainty in system (cyber) risk variables/parameters and incorporate domain expert knowledge and priors (Koller 2009; Pearl 1988), (b) can easily integrate cyber attack graphs at the network and protocol granularity (Pamula et al. 2006; Wang et al. 2017; Sun et al. 2018; Yang et al. 2023) into a Bayesian anomaly detection AI framework. This technology has then been applied on SCADA-based critical infrastructure systems to detect anomalies and raise alarms for subsequent CRM (Zhang et al. 2015; Ten et al. 2007; Frigault and Wang 2008; Nzoukou et al. 2013; Sommestad et al. 2009). A survey of general Bayesian AI methodologies for cybersecurity anomaly detection tasks that are a super set of Bayesian AI tools applicable for cyber attack graphs can be found in (Perusquía et al. 2022). The *biggest drawback of the above-mentioned methods is their applicability only when the CPT parameters in a Bayesian attack graph are driven by sufficient and high accuracy attack (pre)conditions data.* General ICS environments are far from this idealistic setting, and anomaly detection performance in noisy and small data environments is an open study.

The Application of Bayesian AI for Noisy (Small) Data Anomaly Detection - The field of designing high accuracy (a need for ICS CRM) Bayesian parameter inference under noisy and small data on CPT parameters has been an open problem for many years, and every reason for it to be so. The main reason for this is the inherent and proven computational intractability of the general exact and approximate Bayesian

network probabilistic inference problem (Cooper 1990; Dagum and Luby 1993), wherein leave alone humans, even the world's most powerful computers cannot in the worst case (for certain network structures and network sizes) compute the optimal CPT parameters for *exact or approximate* inference. This difficulty increases manifold (and possibly even for average cases) when the CPT data to train these Bayesian networks are noisy and/or small in size as in the case in ICS environments. There has been some recent research on the Bayesian network probabilistic inference on noisy and small data that propose heuristics (that are based upon the parameter extension under constraints methodology) to alleviate the challenges of noisy and/or small data in Bayesian network inference (Hou et al. 2020; Ru et al. 2023; Chen and Ge 2020). However, the performance of such heuristics have not been good enough to be useful for ICS settings.

3 BAYESIAN NETWORK INFERENCE MODEL FOR INTRUDER DETECTION

In this section, we first propose the basics of a Bayesian network and its inference mathematics for the general audience. We then showcase a practical application of the Bayesian network inference model for a malicious intruder setting in a communication network.

3.1 Bayesian Network Inference Basics

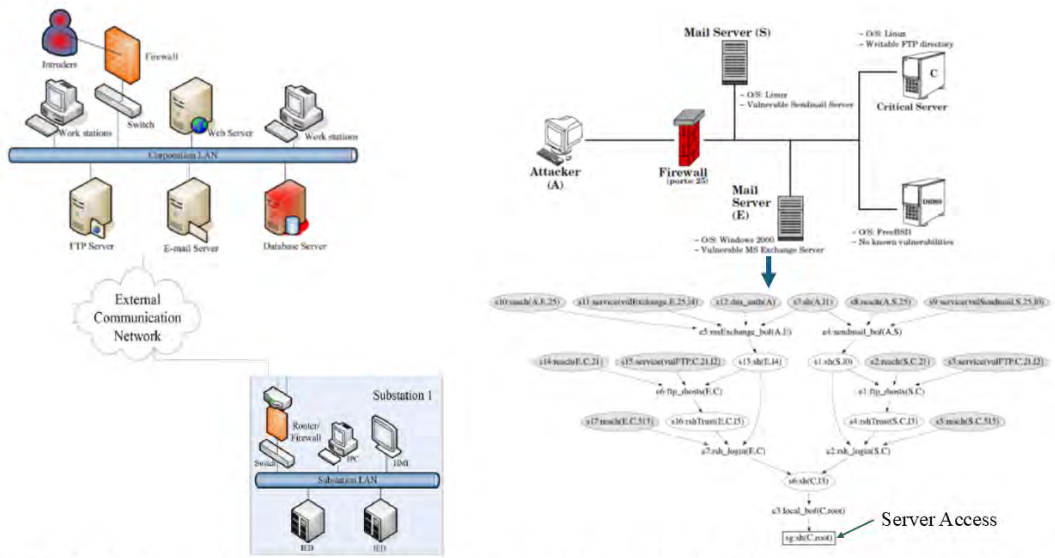
A Bayesian network (BN) is a probabilistic graphical model in AI that is represented as a tuple $BN(G,\theta)$, where G is a directed acyclic graph comprising of (a) nodes acting as random variables and (b) directed edges denoting causal relationships between the random variables (Verma and Pearl 1990). θ consists of a set of parameters that denote conditional probability distributions of node values conditioned on their parent node values. A salient property (principle) of Bayesian networks is the d-separation property that states that each node (random variable) is conditionally independent of other nodes in the network given the parameters of the parent nodes (Verma and Pearl 1990; Koller 2009). The parameters of a Bayesian network associated with each node are popularly called as conditional probability tables (CPTs).

The probabilistic inference problem on Bayesian networks is to perform a joint probability distribution estimation (inference) of N random variables which are represented as the N nodes of the Bayesian network. The structure, principle, and the mathematics of BNs make it possible to efficiently decompose the joint probability computation of N random variables into a series of smaller, connected CPT computations that are significantly and relatively much easier to compute (than a direct computation of a joint distribution) and then can be individually multiplied to get a very good approximation of the joint probability distribution. The d-separation criterion is a necessary and sufficient condition for a joint probability distribution to be compatible (but not necessarily equivalent) with a causal graph.i.e., the Bayesian network.

In mathematical terms, a BN defines a joint probability distribution over nodes (random variables) X_i (with $1 \leq i \leq N$, where N is the total number of BN nodes) and is given by:

$$\mathbf{P}(X_1, X_2 \cdots X_N) = \prod_{i=1}^{N} \mathbf{P}(X_i | P_a(X_i)),$$

where $P_a(X_i)$ for each i represents the parent node set of i in G, and $\mathbf{P}(X_i|P_a(X_i)$ is the conditional probability of each value of X_i given all possible values i's parent nodes in G can assume. $\mathbf{P}(X_i|P_a(X_i)$ is expressed as parameter $\theta_{ijk} = \mathbf{P}(X_i = k|P_a(X_i) = j)$, $\theta_{ijk} \in \theta$, $1 \leq i \leq N$, $1 \leq k \leq r_i$, $1 \leq j \leq q_i$, where r_i represent the number of discrete states of X_i and q_i represent the number of discrete states of parent node $P_a(X_i)$. The assumption of discrete states holds for most practical applications. The collection of all values of form θ_{ijk} for every node i form the CPT for node i (this CPT synonymously called i's parameter in Bayesian network theory). It is then evident that there are $r_i \times q_i$ tabular values populating the CPT for node X_i. *Given G, the eventual task in Bayesian AI is to best estimate the CPT parameters for each node that optimizes the value of the joint probability distribution* $\mathbf{P}(X_1, X_2 \cdots X_N)$ *via computing* $\prod_{i=1}^{N} \mathbf{P}(X_i|P_a(X_i))$.



An ICS Architecture Block with IT/OT Divisions Cyber Attack Graph on Malicious Critical Server Access

Figure 1: An ICS block with IT/OT networks (left). The cyber attack graph to access critical server (right).

The parameter estimation task that best estimates $\mathbf{P}(X_1, X_2 \cdots X_N)$ is computationally hard (Cooper 1990; Dagum and Luby 1993), even after a Bayesian network structural reduction that mitigates the computational complexity. In practice, when sufficient training data is available alongside domain expertise on how the random variables in a BN are causally related, the *Maximum Log-Likelihood Estimation* (MLE) approach is the standard inference approach that first solves the following likelihood (L) optimization problem:

$$L(\theta : \mathbf{D}(G)) = \max_{\theta} \prod_{n=1}^{N} L_n(\theta_{X_n|P_a(X_n)} : \mathbf{D}(G)),$$

and then used the seminal Expected Maximization (EM) algorithm (Koller 2009) to iteratively solve the following log-likelihood (LL) optimization problem for the optimal BN node parameters:

$$J(\theta : \mathbf{D}(G)) = \max_{\theta} \sum_{n=1}^{N} LL_n(\theta_{X_n|P_a(X_n)} : \mathbf{D}(G)),$$

Here, $\mathbf{D}(G)$ is the set of independent data tuples on the nodes (random variables) of G. However, such an approach is not feasible when the training data on the random variables on a BN is scarce. In that scenario, MLE gives way to the *Maximum A Posteriori* approach (Koller 2009) that focuses more on domain expert inputs in the form of statistical priors of the random variables and optimizes BN node parameters outputting a robust approximation to $\mathbf{P}(X_1, X_2 \cdots X_N)$. In popular practice, flat priors and uniform Bayesian Dirichlet priors are modeled for domain expert inputs.

3.2 Application of Bayesian Network Inference in ICS ICT Intruder Detection

We now showcase how the Bayesian network inference model applies well to the problem of malicious intruder detection in ICS ICT settings.

Consider an ICT setting with a corporation local area network (LAN) that controls a substation LAN of a smart power grid (see Figure 1). The left side of Figure 1 shows an ICS block (a subset of the entire ICS) with an adversary/intruder exposed IT-driven ICS corporation LAN controlling a substation OT network. The right hand side of Figure 1 showcases a cyber attack scenario where a malicious adversary intends to target (i.e., intrude into) a critical server in the IT network that controls operations in a substation OT

network. The various possible ways (steps) that the adversary can get control of the critical server are shown via the attack graph, where the ovals are different system states of the cyber attack process and the directed arrows denote the transition from one state to another. The adversary in this case first exploits the *Remote Sendmail buffer-overflow* vulnerability on *Mail Server* (or *Microsoft Exchange* mail server's vulnerability) in order to gain shell level access inside the protected network. It then uses *Mail Server* as an intermediate host to launch the *Anonymous FTP.rhosts remote login* attack, followed by the *local buffer-overflow* gain root level access on the critical server.

The cyber attack graph being an acyclic directed graph, the states can be populated (using domain expert and traffic data information) using conditional probability tables (CPTs) that denote the probability of one state being obtained from another. The resulting graph with populated CPTs is a Bayesian attack graph (BAG). *Without the availability of such a graph it is not possible for ICS managers to know the likelihood of a malicious intrusion into the critical server.* The non-Bayesian cyber attack graph in Figure 1 only states (an important prerequisite) the conditions under which malicious intrusion is possible - *it does not quantify the likelihood of such an event happening.* Quantification is a must for ICS CRM.

4 A SENSITIVITY ANALYSIS SIMULATION FRAMEWORK FOR NOISY BN INFERENCE

We have established thus far that malicious adversary intrusion event likelihood quantification is necessary for effective ICS CRM. On the other hand, we are familiar that CPT population for ICS cyber attack environments is an arduous task due to lack of (a) quality (and sufficient) historical cyber incident data and (b) sufficient domain expertise to confidently populate the CPTs with belief priors. Hence, neither MLE nor MAP inference methodologies are ideally suited for Bayesian network inference. The authors hypothesize, based on their 30+ combined years of experience in (Bayesian) statistics and working (and consulting) with industry research teams on causal cybersecurity analysis, that small and noisy data is too much of a challenge to design and deploy high performance Bayesian network driven CRM solutions rooted in anomaly detection. To put more weight in favor of the hypothesis, the authors in this section first state (as a normative exercise) the minimum number of non-noisy samples needed per parameter to effective inference on a Bayesian network. Assuming that the normative suggestion is not practically deployable in general, the authors then propose a Monte Carlo simulation framework to perform a sensitivity analysis of noisy BN inference to study inference accuracy sensitivity to parameter (CPT) noise.

4.1 Sample Complexity and Inference Methodology for Noisy Settings

We lay down the sample complexity and inference methodology to effectively infer from a BN.

The sample complexity is the minimum number of non-noisy samples needed per Bayesian network parameter (CPT) to accurately learn the parameter via the maximum likelihood estimation approach. Given a Bayesian network G of N nodes, with each node having a maximum parent size of K, the sample complexity (SC) of each parameter is upper bounded by the following relation (Dasgupta 1997):

$$SC(G,N,K) \leq \frac{288N^2 2^K}{\varepsilon^2} \ln^2\left(1+\frac{3N}{\varepsilon}\right) \ln \frac{1+\frac{3N}{\varepsilon}}{\varepsilon\delta},$$

where $(1-\delta)\times 100$ is the confidence percentage of accurate sample complexity, and $\varepsilon = \alpha N$ is the error rate of accurate parameter learning for a small constant $\alpha \geq 0$. The sample complexity $SC(G,N,K)$ increases in N and K and decreases in ε and δ. The sample complexity is an important metric for cyber insurance agencies managing enterprise cyber risk. It allows them to gauge the effort of ICS cyber risk managers into collecting cyber vulnerability data in favor of better CRM - board and upper management driven business processes driving such efforts, feeding into policy pricing and coverage.

Having showcased a closed form expression for the sample complexity, we propose a methodology to infer a target random variable (node) in a Bayesian network when the CPT entries, i.e., parameters are noisy. In this paper, we use the MLE approach to robustly estimate parameters, instead of the MAP

approach of inference. The reason being that in ICS environments, the managers, i.e., the domain experts, are not confident enough of the prior distributions on the BN random variables (cyber risk variables) due to lack of historical data on IT/OT related cyber incidents specific to ICSs. The situation would be different if a new IT start-up company were to design a BN framework for CRM, and the MAP inference is a better approach in small/noisy data settings due to much historic information available on IT security breaches. Hence, for ICS environments, we propose a robust variant of MLE to estimate BN parameters, where we are intent on finding the ground truth version of θ - the vector of parameters of all the BN nodes. The 'weak' prior on ICS-specific cyber domain knowledge from managers are represented partly as noise, and partly as non-noisy inputs to BN CPT entries on the cyber risk variables.

The MLE problem for noisy data $\mathbf{D}(G)$ with a total of M data samples (a mix of noisy and non-noisy random variable values) becomes

$$J(\theta^{noisy} : \mathbf{D}(G), \mathbf{W}) = \max_{\theta^{noisy}} \sum_{n=1}^{N} LL_n(\theta^{noisy}_{X_n|P_a(X_n)} : \mathbf{D}(G), \mathbf{W}),$$

where $\mathbf{W} = \{w_1, \cdots, w_M\}$ is the vector of functionals of M independent and noisy samples of N-dimensional data tuples representing the N nodes (cyber risk variables) of the Bayesian cyber attack graph G. Here, each functional w_m is denoted as $-\exp(-\eta_{\mathbf{d}} \frac{d_m}{\psi_{\mathbf{d}}})$, where $d_m = \sqrt{(x_m - x_{mean})^T S^{-1} (x_m - x_{mean})}$ is the *Mahalanobis* distance of an independent N-dimensional tuple m to the mean of all the tuples; S is the covariance matrix of the tuple matrix; $\psi(\mathbf{d})$ is the standard deviation of $\mathbf{d} = \{d_1, \cdots, d_M\}$; and $\eta_{\mathbf{d}}$ is a positive tuning factor. We have the following result in relation to the effectiveness of Bayesian network inference in the presence of noisy dataset $\mathbf{D}$.

Theorem 1 *Given a Bayesian cyber attack graph (BAG) G of N nodes, and a dataset* $\boldsymbol{D}$ *with noisy data samples of the N-dimensional tuple of cyber risk variables forming G, the following results hold in relation to estimating the joint distribution* $\mathbf{P}(X_1, X_2 \cdots X_N)$ *through the lens of estimating the CPT for an adversary target risk variable* X_N *of the BAG conditioned upon variables* $X_1, \cdots, X_{N-1}$*: (i) the ground truth* $\mathbf{P}^*(\theta^{noisy})$ *of* $\mathbf{P}(X_1, X_2 \cdots X_N)$ *converges to* $\mathbf{P}(X_N|P_a(X_N))$ *w.p. 1 when the number of non-noisy samples in* $\boldsymbol{D}$ *approaches* $SC(G,N,K)$*, (ii) out of all factorizations of* $\mathbf{P}(X_1, X_2 \cdots X_N)$ *forming a set Q, the KL-divergence measure between* $\mathbf{P}(X_N|P_a(X_N))$ *(as an element of Q) and ground truth value of* $\mathbf{P}(X_1, X_2 \cdots X_N)$ *is the minimum.*

Proof Sketch - To prove the first part of the theorem, we resort to the concept of the Law of Large Numbers (LLN). Let $\mathbf{P}^*(X_1, X_2 \cdots X_N)$ factorizes to an element S in set Q then $\mathbf{P}^* = \mathbf{P}^S | argmax_{\theta^{noisy}} J(\theta^{noisy} : \mathbf{D}(G), \mathbf{W})$. Here, $\mathbf{P}^S = \mathbf{P}(X_N|P_a(X_N))$. Let $\hat{\mathbf{P}}$ be the empirical value of $\mathbf{P}(X_1, X_2 \cdots X_N)$. Using LLN, we show that $\mathbf{P}^S | argmax_{\theta^{noisy}} J(\theta^{noisy} : \mathbf{D}(G), \mathbf{W}) = \hat{\mathbf{P}}(X_1, X_2 \cdots X_N)$ when M approaches $SC(G,N,K)$. To prove the second part, note that the KL-divergence measure $KL(P^*, R) = \sum_{\mathbf{x}} P^*(\mathbf{x}) \log \frac{P^*(\mathbf{x})}{R(\mathbf{x})}$ (where $P^*(\mathbf{x}) = \mathbf{P}^*(X_1, X_2 \cdots X_N)$; $R(\mathbf{x})_{\in Q} = \mathbf{P}(X_N|P_a(X_N))$ is maximized when $\sum_{\mathbf{x}} P^*(\mathbf{x}) \log R(\mathbf{x})$), and is proved using mathematical induction.

4.2 Sensitivity Analysis Simulation Framework to Decipher Impact of Noisy Datasets

The above result states, as part, the high accuracy of Bayesian network inference when the non-noisy data sample count of BAG parameters nears $SC(G,N,K)$ for a BAG of N nodes, with each node variable having a maximum of K parents. The authors hypothesize, based on their 30+ combined years of experience in (Bayesian) statistics and working (and consulting) with industry research teams on causal cybersecurity analysis, that small and noisy data is too much of a challenge to design and deploy high performance Bayesian network driven CRM solutions rooted in anomaly detection. In other words, BAG inference is too sensitive to dataset noise - *the degree of such sensitiveness, as a function of structure-driven noise addition, is not captured in the closed form expression* $SC(G,N,K)$*, and is not studied in existing literature.* It is the structural insights that are important to defend against strategic BN structure-motivated adversaries. Such insights when combined with $SC(G,N,K)$ analysis (see Figure 3) will result in the most effective and practically deployable CRM action items for noisy/small data ICS environments.

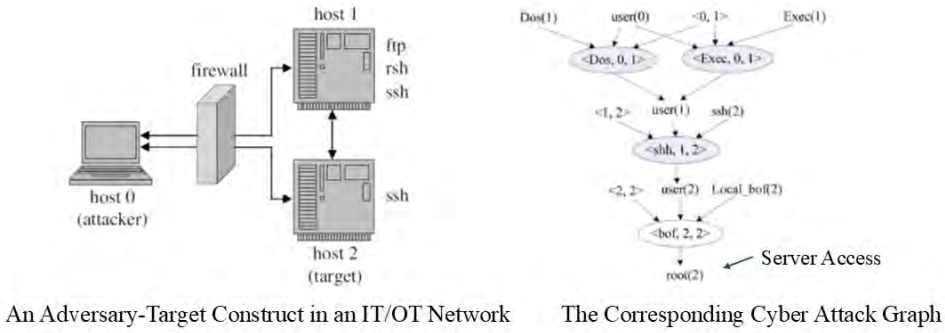


Figure 2: An adversarial setting in an IT/OT network (left). The corresponding cyber attack graph (right).

4.2.1 Framework Design Motivation and Real-World Setup

Design Motivation - In this section, to put weight on our hypothesis and more importantly think from an adversarial viewpoint which nodes of a BAG are most critical and sensitive to noise injection in their CPTs with respect to Bayesian network inference, we design a Monte Carlo simulation framework for sensitivity analysis. The closed form expression for $SC(G,N,K)$ indeed provides the variations of SC with acceptable error rates, but does not provide a microscopic view into how the BAG structure affects inference error - *an important adversarial cyber risk management task.*

Real-World Setup - We perform Bayesian network sensitivity analysis on a cyber attack graph that has been operated on in practice for research simulations at The MITRE Corporation. The *adversary-target* construct in a miniature real-world network prototype is shown in Figure 2 (left), and the corresponding cyber attack graph is shown in Figure 2 (right). The cyber attack on the left of Figure 1 represents two zero day attacks on the *ssh* protocol services on *host 1* and *host 2*, and this is followed by a buffer overflow attack on target *host 2*. It is interesting and important to note that even for this miniature network, a simple *adversary-target* construct and cyber attack sequence leads to an attack graph with more than 20 nodes (the non-circled pre-conditions for a cyber attack that attached to edges on the graph are also treated as nodes). Hence, a cyber attack graph size grows significantly fast in the size of a communication network and cyber attack complexity - *an important point for consideration in Section 5*.

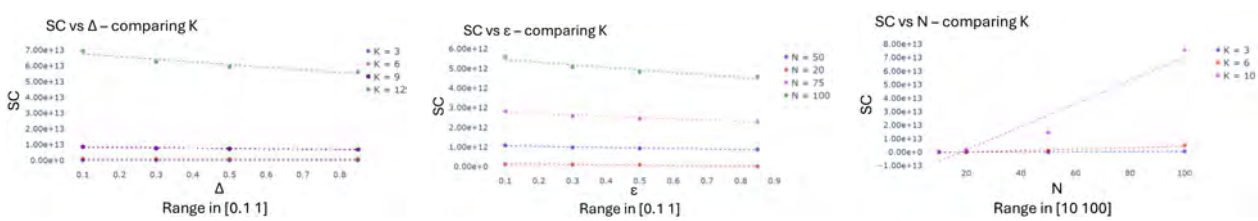


Figure 3: Sample Complexity (SC) vs Δ, ε, and N, varying N and K for comparisons

4.2.2 Monte Carlo Sensitivity Analysis Simulation Setup

In this section, we lay down the Monte Carlo simulation setup to conduct sensitivity analysis on a Bayesian attack graph derived from the cyber attack graph in Figure 2. We strategize adversarial moves in two dimensions: (i) the nodes of a BAG that adversaries are interested to compromise, and (ii) the amount of statistical adversarial noise injected into CPT of targeted nodes. As an 'exhaustive' BAG node selection criteria, we assume (without loss of generality) that adversaries target (a) top three nodes with highest in-degree centrality, (b) top three nodes with highest out-degree centrality, (c) top three nodes with highest betweenness centrality, (d) three nodes at different heights of the BAG but with the same in-degree centrality

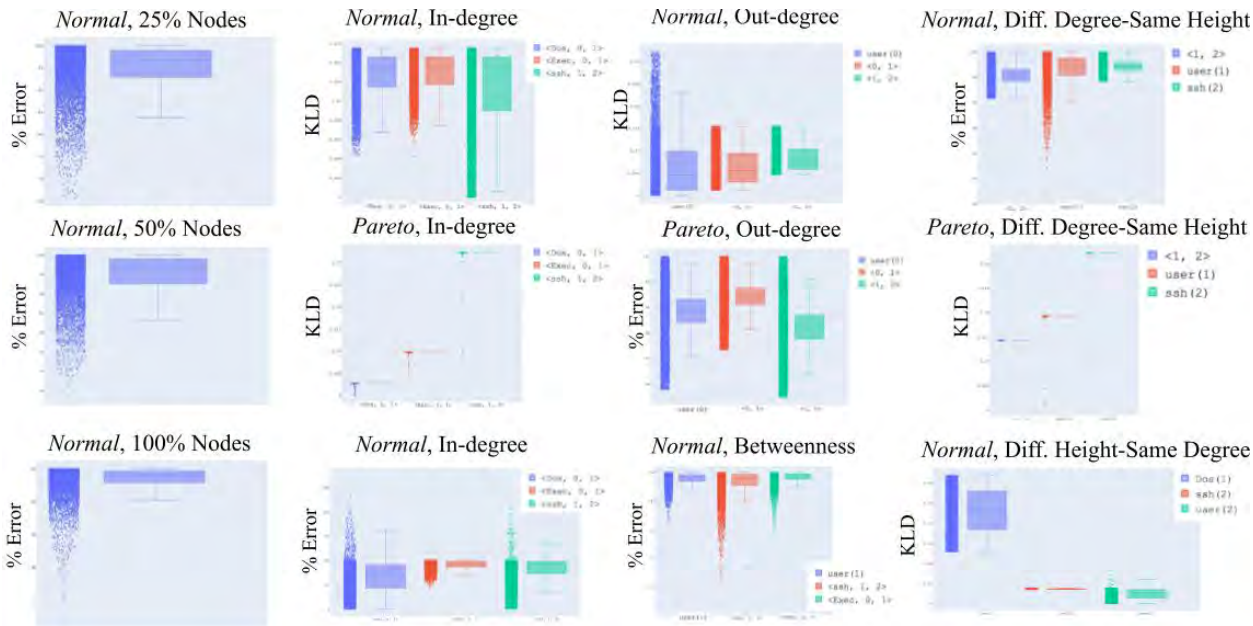


Figure 4: Simulation Graphs Modeling Error with Types of Node Selection, Noise, and Comparison Statistic. (The *y*-axis is BAG inference error as the KL-divergence measure. The *x*-axis is BAG nodes under attack.)

measure, and (e) three nodes at the same height but with different in-degree centrality measures. As BAG node noise injection criteria, (statistically exhaustive) noise was introduced into adversary-targeted BAG nodes using *light-tailed* (Normal with mean 0 variance 1) and *heavy-tailed* (Pareto with α set to 1 indicating finite mean and infinite variance) distributions. Large scale Monte Carlo simulations, conducted with 1000 randomized instances per method, generated distributions of possible probabilistic inference error outcomes under these perturbations. The results (outcomes) were analyzed using the standard *percentage error* and *KL-divergence* metrics to quantify deviations from the non-noisy baseline. As an example, perturbing the top three in-degree BAG nodes under Pareto noise was compared in inference error performance with normal (i.e., no noise) conditions, using the KL-divergence measure (metric). Monte Carlo simulation results are plotted in Figure 4, and sample complexity plots are shown in Figure 3.

5 SIMULATION RESULTS AND ANALYSIS WITH ICS CRM ACTION ITEMS

It is evident from Figure 3 that SC is sufficiently high for gaining high statistical accuracy and confidence. Even if we tolerate certain practically viable error percentages, the number of non-noisy samples to correctly infer with statistical confidence, even with small/moderate BAG node sizes is quite high. Hence, **BAG inference is likely not a suitable ICS CRM methodology for small data driven CPTs.** In practice, sample size far less than SC usually gives fairly good inference accuracy for detecting known incidents due to strong domain expertise knowledge. Challenges with small/noisy data arise when BAG size is large and managers infer unseen (e.g., zero day) cyber incidents. Samples on 'strategic' BAG nodes then demand sample sizes near SC limits. For noisy CPT settings, *we observe (in Figure 4) a wide statistical variance of BAG inference errors even with a low amount of injected adversarial noise.* From all the node selection methods, we see that applying noise functions on lower height nodes (near the top of the BAG) has more influence on independent ancestors, with larger error ranges for nodes with influential ancestors. For instance, *<ssh, 1, 2>* connects multiple paths leading to the root, making it critical in the propagation of information within the network and is reflected by the KL-divergence range being roughly double of the *<Dos, 0, 1>* and *<Exec, 0, 1>* nodes. Any disturbance here is more likely to affect the overall inference accuracy. In the BAG, as we move down towards the root, nodes may accumulate adversarial noise effects

from multiple parent nodes, leading to greater variability in error. Conversely, narrower ranges indicate lower average response spread or sensitivity to noise, suggesting that nodes higher up in the hierarchy (like *<Dos, 0, 1>*) are less affected. Hence, **as a managerial action item, ICS CRM must ensure enough (near SC benchmark) non-noisy CPT entries for nodes higher in the BAG**, as inference is most sensitive to adversarial noise on such nodes. Another thing worthy of note is *if CPT entries are higher for relatively high in-degree/betweenness centrality nodes prior to noise injection, it indicates an increased adversary attention on these nodes - noise injected at these nodes lead to more significant shifts in predicted BAG inference outcomes, when compared to noise injection on similar centrality but low parameter (CPT entry) value nodes.* Hence, **ICS CRM managers must ensure a high fraction (near SC limit) of non-noisy CPT entries for such centrally-located nodes in the BAG, to reduce the quantity of false alarms.**

6 PAPER SUMMARY

We arrived at an interesting and surprising result that while Bayesian AI is a viable tool for anomaly detection applications in IT enterprise security management, it is not recommended for IT/OT convergent ICSs due to the lack of sufficient non-noisy data on the CPT parameters related to the cyber risk variables of a cyber attack graph characteristic of ICS environments. To derive our result we developed a novel graphical sensitivity analysis simulation framework intersecting noisy data Bayesian network inference. Monte Carlo simulations showed that managers should collect sufficient non-noisy AI parameter data on system cyber risk variables for quality adversarial intrusion detection as part of CRM optimization.

ACKNOWLEDGEMENT

This study has been supported by funding from Cybersecurity at MIT Sloan (CAMS).

REFERENCES

Ammann, P., D. Wijesekera, and S. Kaushik. 2002. "Scalable, graph-based network vulnerability analysis". In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, 217–224.

Chen, G. and Z. Ge. 2020. "Robust Bayesian networks for low-quality data modeling and process monitoring applications". *Control Engineering Practice* 97:104344.

Cooper, G. F. 1990. "The computational complexity of probabilistic inference using Bayesian belief networks". *Artificial intelligence* 42(2-3):393–405.

Dagum, P. and M. Luby. 1993. "Approximating probabilistic inference in Bayesian belief networks is NP-hard". *Artificial intelligence* 60(1):141–153.

Dasgupta, S. 1997. "The sample complexity of learning fixed-structure Bayesian networks". *Machine Learning* 29:165–180.

Dewri, R., N. Poolsappasit, I. Ray, and D. Whitley. 2007. "Optimal security hardening using multi-objective optimization on attack tree models of networks". In *Proceedings of the 14th ACM conference on Computer and communications security*, 204–213.

Frigault, M. and L. Wang. 2008. "Measuring network security using bayesian network-based attack graphs". In *2008 32nd Annual IEEE International Computer Software and Applications Conference*, 698–703. IEEE.

Hou, Y., E. Zheng, W. Guo, Q. Xiao and Z. Xu. 2020. "Learning Bayesian network parameters with small data set: A parameter extension under constraints method". *IEEE Access* 8:24979–24989.

Khan, S. and S. Madnick. 2021. "Cybersafety: A system-theoretic approach to identify cyber-vulnerabilities & mitigation requirements in industrial control systems". *IEEE Transactions on Dependable and Secure Computing* 19(5):3312–3328.

Koller, D. 2009. "Probabilistic Graphical Models: Principles and Techniques". The MIT Press.

Leveson, N., M. Daouk, N. Dulac, and K. Marais. 2003. "Applying STAMP in accident analysis". In *NASA Conference Publication*, 177–198. NASA; 1998.

Li, Y., J. Chen, and L. Feng. 2012. "Dealing with uncertainty: A survey of theories and practices". *IEEE Transactions on Knowledge and Data Engineering* 25(11):2463–2482.

Maccarone, L., D. Buede, S. Bowman, C. Burdick, M. Bracken, J. Jones *et al*. 2022. "Development of a Bayesian Network to Model Malicious Cyber-Activity in Operational Technology Environments.". Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Nzoukou, W., L. Wang, S. Jajodia, and A. Singhal. 2013. "A unified framework for measuring a network's mean time-to-compromise". In *2013 IEEE 32nd International Symposium on Reliable Distributed Systems*, 215–224. IEEE.

Pal, R., P. Liu, T. Lu, and E. Hua. 2023. "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 (TCPS)* 6(4):1–31.

Pal, R., T. Lu, P. Liu, and X. Yin. 2021. "Cyber (re-) insurance policy writing is NP-hard in IoT societies". In *2021 Winter Simulation Conference (WSC)*, 1–12. IEEE.

Pal, R., R. Sequeira, and S. Zeijlemaker. 2024. "How Hard is it to Estimate Systemic Enterprise Cyber-Risk?". In *2024 Winter Simulation Conference (WSC)*.

Pamula, J., S. Jajodia, P. Ammann, and V. Swarup. 2006. "A weakest-adversary security metric for network configuration security analysis". In *Proceedings of the 2nd ACM workshop on Quality of protection*, 31–38.

Pearl, J. 1988. "Evidential reasoning under uncertainty". In *Exploring Artificial Intelligence*, 381–418. Elsevier.

Perusquía, J. A., J. E. Griffin, and C. Villa. 2022. "Bayesian models applied to cyber security anomaly detection problems". *International Statistical Review* 90(1):78–99.

Pfleeger, S. and R. Cunningham. 2010. "Why measuring security is hard". *IEEE Security & Privacy* 8(4):46–54.

Ru, X., X. Gao, Y. Wang, and X. Liu. 2023. "Bayesian network parameter learning using constraint-based data extension method". *Applied Intelligence* 53(9):9958–9977.

Sahu, A. and K. Davis. 2022. "Inter-domain fusion for enhanced intrusion detection in power systems: An evidence theoretic and meta-heuristic approach". *Sensors* 22(6):2100.

Sheyner, O., J. Haines, S. Jha, R. Lippmann and J. M. Wing. 2002. "Automated generation and analysis of attack graphs". In *Proceedings 2002 IEEE Symposium on Security and Privacy*, 273–284. IEEE.

Sommestad, T., M. Ekstedt, and L. Nordstrom. 2009. "Modeling security of power communication systems using defense graphs and influence diagrams". *IEEE Transactions on Power Delivery* 24(4):1801–1808.

Stamp, J., A. McIntyre, and B. Ricardson. 2009. "Reliability impacts from cyber attack on electric power systems". In *2009 IEEE/PES Power Systems Conference and Exposition*, 1–8. IEEE.

Sun, X., J. Dai, P. Liu, A. Singhal and J. Yen. 2018. "Using Bayesian networks for probabilistic identification of zero-day attack paths". *IEEE Transactions on Information Forensics and Security* 13(10):2506–2521.

Ten, C.-W., J. Hong, and C.-C. Liu. 2011. "Anomaly detection for cybersecurity of the substations". *IEEE Transactions on Smart Grid* 2(4):865–873.

Ten, C.-W., C.-C. Liu, and M. Govindarasu. 2007. "Vulnerability assessment of cybersecurity for SCADA systems using attack trees". In *2007 IEEE Power Engineering Society General Meeting*, 1–8. IEEE.

Verma, T. and J. Pearl. 1990. "Causal networks: Semantics and expressiveness". In *Machine intelligence and pattern recognition*, Volume 9, 69–76. Elsevier.

Wang, L., S. Jajodia, A. Singhal, P. Cheng and S. Noel. 2013. "k-zero day safety: A network security metric for measuring the risk of unknown vulnerabilities". *IEEE Transactions on Dependable and Secure Computing* 11(1):30–44.

Wang, L., S. Jajodia, A. Singhal, M. Frigault, L. Wang, S. Jajodia *et al*. 2017. "Measuring the overall network security by combining cvss scores based on attack graphs and bayesian networks". *Network Security Metrics*:1–23.

Yang, B., M. Hoffman, and N. Brown. 2023. "Bayesian Networks for Interpretable Cyberattack Detection". In *2013 56th Hawaii International Conference on System Sciences (HICSS)*.

Young, W. and N. Leveson. 2013. "Systems thinking for safety and security". In *Proceedings of the 29th annual computer security applications conference*, 1–8.

Zhang, Y., L. Wang, Y. Xiang, and C.-W. Ten. 2015. "Power system reliability evaluation with SCADA cybersecurity considerations". *IEEE Transactions on Smart Grid* 6(4):1707–1721.

AUTHOR BIOGRAPHIES

YAPHET LEMIESA is a student in the Mechanical Engineering and EECS departments at MIT. He is also a researcher with Cybersecurity at MIT Sloan (CAMS) at the MIT Sloan School of Management. His primary research interest lies in cyber risk management using AI/ML ops for critical infrastructures and business networks. His email address is yaphkl75@mit.edu.

RANJAN PAL is a Research Scientist with the MIT Sloan School of Management, and an invited working group member of the World Economic Forum. His primary research interests lie in cyber risk and resilience management using interdisciplinary methods. He serves as an Associate Editor of the ACM Transactions on MIS. His email address is ranjanp@mit.edu.

MICHAEL SIEGEL is a Principal Research Scientist with the MIT Sloan School of Management. His primary research interest lies in cybersecurity management of information systems. He is the founding co-Director of the Cybersecurity at MIT Sloan (CAMS) center within the MIT Sloan School of Management. His email is msiegel@mit.edu.