Harvard Data Science Review • Issue 5.1, Winter 2023

Statistics and Data Science for Cybersecurity

Alfred Hero[1] Soummya Kar[2] Jose Moura[2] Joshua Neil[3]
H. Vincent Poor[4] Melissa Turcotte[5] Bowei Xi[6]

[1]University of Michigan, Ann Arbor, Michigan, United States of America,
[2]Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America,
[3]Securonix Inc., Addison, Texas, United States of America,
[4]Princeton University, Princeton, New Jersey, United States of America,
[5]Secureworks Inc., Atlanta, Georgia, United States of America,
[6]Purdue University, West Lafayette, Indiana, United States of America

Published on: Jan 26, 2023

DOI: https://doi.org/10.1162/99608f92.a42024d0

ABSTRACT

Cybersecurity is an ever-important aspect of our interconnected world, but security defenses lag behind the adversaries who with increasing sophistication seek to disrupt cybersystems. The emergence of massively distributed systems such as the Internet of Things (IoT) has opened up new vulnerabilities that go beyond traditional protective measures such as firewalls, password protection, and single point-of-attack responses. To address these emerging vulnerabilities, data science has much to contribute, including methods of distributed statistical inference, data fusion, anomaly detection, and adversarial machine learning.

Keywords: Adversarial machine learning, enterprise cybersecurity, resilient distributed networks, information theoretic security

Media Summary

Cybersecurity is an ever-important aspect of our interconnected world, but security defenses lag behind the adversaries who with increasing sophistication seek to disrupt cybersystems. The emergence of massively distributed systems such as the Internet of Things (IoT) has opened up new vulnerabilities that go beyond traditional protective measures such as firewalls, password protection, and single point-of-attack responses. To address these emerging vulnerabilities, data science has much to contribute, including methods of distributed statistical inference, data fusion, anomaly detection, and adversarial machine learning.

1. Introduction

Interconnected data services continue to grow in size and complexity, exposing new vulnerabilities to cyberattacks and outages. Services such as cellphone, texting, media streaming, cloud computing, and data storage increasingly underpin our economic, social, and community support structure. Interruption or disruption of such essential services can create great inconvenience to people and organizations. Massive distributed sensing and control systems, forming the emerging Internet of Things (IoT), are also vulnerable to malicious attacks with potentially grave consequences, for ex- ample, a deadly attack on a connected autonomous vehicle (CAV) network. However, cybersecurity software has not evolved sufficiently quickly to cope with the increasing sophistication of attackers and the systems that they target. As methods of network measurement and data science mature, there is potential to learn attacker behaviors, mitigate attacks, and make networked systems less vulnerable. This creates opportunities for statisticians and data scientists to develop better methods that span the spectrum of data collection, data analysis, behavioral modeling, and information security.

In acknowledging these challenges and to assess the possibilities of using data science, a half day symposium entitled Statistics and Data Science for Cybersecurity was organized by the US National Academies Committee on Applied and Theoretical Statistics (CATS) in June 2018. This article represents the perspectives and points of view of the authors, who participated at the meeting as organizers and speakers. The article is divided into four topical areas of data science for cybersecurity: the emerging role of data science in cybersecurity (Section 2); data-driven cybersecurity for enterprise systems (Section 3); cybersecurity data collection and analysis (Section 3); data-driven cybersecurity for the IoT (Section 4); ensuring information privacy in the IoT (Section 5).

A major theme that emerged is that current security methods adopted for enterprise and ad hoc networks can be improved by applying advanced statistical and machine learning (ML) techniques that adaptively and quickly detect attacks. Machine learning has had an impact on user-level security services, for example, Gmail and Yahoo mail spam filters incorporate ML methods that achieve very high detection accuracy (Dada et al., 2019). This suggests that ML can be used to similarly improve upon traditional cybersecurity methods that focus on rule-based approaches such as cryptography, authentication, filtering and other data-agnostic mechanisms. As systems become more complex and as attackers become more sophisticated, these traditional methods have not kept up. On the other hand, networked systems are increasingly being instrumented with data monitoring, both at the edge and within-network, which has created a wealth of data traces with which statistical machine learning approaches can integrate diverse information to better detect, track, and understand attacks. Several machine learning approaches have been emerging for related security applications, including: adversarial and game theoretic machine learning models for spam filtering (C. Wang et al., 2021); statistical anomaly detection methods that can detect subtle changes in network data (Lazarevic et al., 2003); transfer learning for detecting novel network attacks (Zhao et al., 2019); and semisupervised methods that can learn from partially labeled network data (Hou et al., 2018).

More generally, there is a need for cybersecurity defenses that go beyond the use of reactive models, for example, based on signature detection or rule-based decisions, toward predictive models. Predictive models will use statistically based anomaly detection principles to detect and control novel malware and evolving software exploitation strategies. Such models will exploit data collected within and at the periphery of organizations to continuously update and improve defenses. Most importantly they will account for attack behavior, called the ’kill chain,’ that starts with initial penetration of the network and progresses to exfiltration, where stolen data is transmitted back to the adversary. Building effective defenses will involve data-driven mechanisms combining stochastic modeling, dynamic graphs, and concepts from statistical control (Neil et al., 2013; Turcotte et al., 2018).

Decentralized networks such as the IoT pose special challenges for cybersecurity, and they require different solution strategies than do centralized enterprise networks. Lacking central administration, it is unrealistic to expect that all devices on the IoT can be equally protected from attacks, and properties of network

infrastructure may similarly be outside of any central control. Furthermore, the high diversity of the devices, which can include access points, actuators, and sensors, make traditional approaches to device-level protection more difficult. This creates new opportunities for defense: as IoT devices become more powerful, data-driven security can be implemented at the edge and device levels. Furthermore the communication links of the network can themselves be protected by using recently proposed secure multiparty communication protocols developed using multiuser communications and information theory. Such protection can ensure the privacy of information shared across the network, depriving the attacker of an important access point for discovering network vulnerabilities.

We emphasize that this article focuses on some of the opportunities for statistics, machine learning, and information theory to positively impact network security. It does not offer a comprehensive review of the field of cybersecurity nor does it provide a general purpose survey of statistical methods for this field. The article's contributions are summarized below.

Section 2 discusses the emerging role of data science in cybersecurity. It lays out the dichotomy between centrally administered enterprise networks and decentralized ad hoc networks in terms of their particular cybersecurity challenges and how data-driven methods can be used to overcome some of these challenges. The subsequent sections provide a deeper dive into some of the specific ways that statistical machine learning and information theory can make a difference.

Section 3 discusses opportunities to incorporate data science into cybersecure enterprise systems. The section discusses the potential of recent statistical methods for adversarial machine learning. It has become apparent that machine learning algorithms are quite vulnerable to poisoning and evasion attacks during the training and test phases, respectively (Goodfellow et al., 2015), (Y. Zhou et al., 2012). High-dimensional deep neural network classifiers are especially sensitive. For example, convolutional neural network (CNN) classifiers can be compromised by attackers who apply simple black-box 'zero-th order optimization' (ZOO) methods (P.-Y. Chen et al., 2017) to defeat the classifier. and continues with a discussion of advances in statistical methods of anomaly detection that have the potential to improve enterprise system security. Debilitating cyberattacks on the network infrastructure of many enterprise organizations are on the rise and are becoming increasingly sophisticated and threatening (Sanger, 2018).

Sections 4 and 5 turn to the topic of physical layer cybersecurity and resilient decision-making algorithms for distributed sensor, actuator, and transmission networks characteristic of the emerging Internet of Things. Such distributed networks can involve millions of interconnected sensors and actuators that communicate with each other, requiring different approaches to cybersecurity than for centrally administered networks like enterprise systems. As contrasted to the application layer approaches discussed in Section 3, physical layer approaches to cybersecurity exploit properties of the physical transmission medium, for example, strengthening the defenders wireless network using cooperative relays or degrading the attacker's wireless access to the network by insertion of jamming or fading into the channel (Dong et al., 2010; Hero, 2003).

In Section 4, we describe challenges in making the data collected by decentralized IoT networks less sensitive to attacks. In contrast to a centrally administered network where all devices share their information with a central server, in decentralized IoT networks devices share information locally and must collectively come to a consensus on decisions. This makes decentralized networks especially vulnerable to attacks since an adversary can exploit the lack of central administration, targeting a small number of devices that are critical in the data sharing or decision paths. Tools for quantifying network sensitivity to different types of attacks are needed so that designers can meet the challenge of designing more resilient systems.

In Section 5, we turn to information theoretic approaches to cybersecurity, specifically privacy and security, for IoT networks. Formulated differently than algorithmic differential privacy (Dwork & Roth, 2014; Dwork & Smith, 2009), the information theoretic approach to privacy benefits from more than 60 years of progress in information theory (Shannon, 1949), the enabler of the digital revolution creating digital communications, data compression, and cryptography. In particular, as discussed in Section 5, information theory provides a general framework for understanding differential privacy in terms of fundamental rate-distortion trade-offs, degraded communications, and wiretap channels, that is especially useful for ad hoc IoT networks.

Finally, in Section 6 we include a broader discussion of cross-cutting themes that arose at the workshop and chart a course for future research directions.

2. The Emerging Role of Data Science in Cybersecurity

Cybersecurity for networked systems has been a research focus more than the past three decades and is relevant to a broad range of network applications (Jang-Jaccard & Nepal, 2014). Such systems are vulnerable to attacks that seek to disrupt or damage system functionality and attacks that seek to compromise information security. Distributed denial of service and link jamming are examples of attacks that disrupt system functionality. Intrusion and eavesdropping are examples of attacks that compromise information security, gaining unauthorized access to proprietary system data or private user data and compromising confidentiality, integrity, and availability (CIA) (Samonas & Coss, 2014). Such attacks can occur in sequence or in parallel, for example, a ransom attack that steals confidential information to break into a device, hijacks the operating system, and uses it to launch attacks on other network devices. Two types of networks are of particular interest and have different vulnerabilities: centrally administered enterprise networks and non-centrally administered networks such as the ad hoc IoT. While enterprise and ad hoc IoT systems may pose different security challenges and vary in terms of applicable security solutions, they share some important traits, notably, heterogeneity of the network, their large scale, and the multiplicity of vulnerabilities to attack. They may both have complex interconnections of diverse sensors, actuators, and devices, ranging from cloud and edge resources to end-user devices (cameras, microphones, appliances, etc.) having different characteristics and capabilities (McKinsey & Company, 2019). In addition to device diversity, these systems are multilayered, both in terms of functionality and network administration. The main difference between enterprise networks

and ad hoc networks is that the former usually has centralized administration while in the latter, administration is often decentralized, for example, peer-to-peer protocols.

The wide diversity and distribution of access points in both types of networks creates security challenges since it opens the system up to a large number of potential attack points, each protected at different levels of sophistication and resources, for example, passwords, authentication, or firewalls. A resourceful adversary may deploy diverse tools to inflict different types of attacks, for example, a coordinated intrusion or a distributed denial of service (DDOS) attack on enterprise infrastructure (Farina et al., 2016; Mansfield-Devine, 2015), in addition to man-in-the middle (MITM) and phishing attacks on end-users. The latter often involve a third party that intercepts sensitive communications between an organization and its clients; the breaches on DigiNotar and Equifax provide real-world examples (Veracode, 2022). Still other attackers could aim at corrupting application-specific data, for example, false data injection (FDI) attacks on sensors and actuators. Such attacks are often aimed at industrial systems to spoof the underlying control and automation processes; real-world examples include attacks such as the Stuxnet and the Maroochy Shire sewage control system incidents (Liang et al., 2017).

By exploiting vulnerabilities at resource-limited networked entities and endpoints, sophisticated adversaries may overwhelm classical security mechanisms, which often have limited capabilities to adapt to changing adversarial attack scenarios. To counter such devious attacks, future systems will need to incorporate proactive data-driven approaches to cybersecurity, specifically those using statistical and artificial intelligence (AI)-based approaches. Recent advances in these areas hold much promise for both complementing traditional security measures and, if properly designed, exploiting the rich high-dimensional and multimodal data generated by these systems. In enterprise systems a network operations center (NOC) can monitor, integrate, and analyze large amounts of diverse data coming from network flows, web proxy logs, DNS logs, and other sources discussed in Section 2.1. Critical cyber-physical infrastructures, such as the smart grid, financial networks, health monitoring networks, and transportation networks must be closely monitored at the device, transport, and application levels at both interior and edge points in the network. There is thus a need for statistical data-driven approaches to analyzing the complex data generated in multilevel network monitoring (Collins, 2017; F. Zhang et al., 2019).

Some of the attributes of enterprise system security deserve mention as they determine the type of data of highest value for detection of attacks. Data-driven security approaches may be broadly distinguished as falling into three categories: network, endpoint, and application-centric. Network security focuses on preventing intrusions at the network-wide level and ensuring overall system reliability, while endpoint security techniques are primarily designed to protect and secure endpoint devices and users. Application security mechanisms, on the other hand, are typically designed to ensure reliability of specific applications. Due to the complexity of attacks in today's enterprise networked systems, multistack approaches that use information from different layers of the system are gaining traction and provide resilience to larger classes of attacks than the

compartmentalized protective mechanisms of the past. Indeed, the cross-layer design of enterprise networks offers a key advantage to data-driven security solutions that can integrate data from different physical and application components of the system. Network intrusion and anomaly detection algorithms of this kind are discussed in Section 3.

On the other end of the spectrum are ad hoc IoT networks, increasingly used in low-cost local wireless sensor network (WSN) applications and mobile ad hoc networks (MANETS). These networks are frequently deployed in an ad hoc fashion in environments without much security infrastructure. They often feature plug-and-play entities that dynamically leave and join the network, in addition to new entities, services, and devices that have the ability to seamlessly integrate into the network. These services and devices are often application-centric, having the goal of reliably carrying out a small number of sensing, computing, and inference tasks without the need for network infrastructure. Examples include distributed camera or chemical sensor networks and ad hoc mobile media delivery systems (TrellisWare Technologies Inc., 2017).

Unlike centrally administered enterprise systems where network security is often designed into the system to protect the infrastructure as a whole, by their decentralized nature existing ad hoc networks are often unable to achieve adequate network-level security. In Section 4 and Section 5, we discuss the potential of methods of secure multiparty communication and computation for improving network-level security and privacy in decentralized networks.

We next describe more concisely some of the unique characteristics of statistical and AI-based security techniques in enterprise and ad hoc network settings.

2.1. Data-Driven Automation of Enterprise Cybersecurity

Today an increasingly diverse group of malicious actors, ranging from individual hackers to sophisticated nation-state groups, are targeting enterprise cybersystems using a wide variety of ever-changing attack techniques (Verizon, 2022) Attackers have two clear advantages. First, the defense must be comprehensive, while the offense can be targeted. Defenders need to protect every asset and attack vector of a network, while attackers need only identify one vulnerability in these vast networks to gain a foothold and conduct their attacks. Even a single operating system may have many vulnerabilities, and once networked, connected to the internet, and driven by human beings, the defensive task appears daunting. Second, it is better for the defense to be prospective and proactive rather than retrospective and reactive. For example, the first antivirus technologies used ‘signatures,’ really cryptographic hashes of byte strings, found in certain malicious software, to identify other examples of this malware (Nachenberg, 1997).This practice pervaded much of the technology produced by cybersecurity companies up until recently. Other classic defense mechanisms are ‘rule-based’ and can be thought of as a decision tree with hard-coded parameters, often extracted through trial and error and expert knowledge, and taken from real-world examples. Clearly, both of these methods are retrospective, and adversaries are quick to innovate around these pinpoint behaviors to avoid detection. This process has resulted

in the poor state of cybersecurity today: breaches are common, companies are losing valuable intellectual property, and people are losing control of their personal information.

To meet this threat, cybersecurity technology has been advancing dramatically in recent years. Advances in collection, storage, and big data infrastructure have led to the ability to capture and analyze high-resolution data in near real time from each computer, virtual machine, and network tap in even the largest global computer networks. Where the data of yesterday were measured at the perimeter of a cybersystem and were only aggregated summaries, now we have excellent visibility within the operating network, resolved down to individual assets reporting almost continuously. These data can be summarized in a wide variety of logs and meta-data, as well as low-level operating system and network telemetry, and can be generated and stored in petabyte storage facilities.

A crucial source of data are the operating systems of the machines on the network. Recently, the cybersecurity industry has introduced new capabilities to capture these data, known as Endpoint Detection and Response (EDR) technology (Hassan et al., 2020). EDR involves the addition of monitoring software executing on the operating systems that measures various telemetry of security interest, and uploads these data to either on-premise or cloud servers for analysis. Behaviors are manifold, but include registry and file system changes, command lines, process usage, and even network telemetry. Certainly, malicious execution through file-based malware can be detected from these data, in fact this is how antivirus software works. However, more subtle behavioral aspects stemming from adversaries using standard system tools in malicious ways, known as living-off-the-land techniques (Yasin, 2015), are detectable with these data, but typically require unsupervised statistical methods such as anomaly detection that can more reliably detect novel malicious behaviors and incorporate false-positive error control.

There are several challenges that need to be addressed before a researcher can begin using these data. Obtaining high-quality data can be very difficult, due to the distributed nature of the network, and the enormous volumes of data that a large cybersystem can produce. Infrastructure for collection and storage is significant, and in the case of EDR, one needs additional software installed on each computer that is being monitored. In addition, once collected, many of the sources have limited use in their raw form. With the exception of EDR data, these data were originally intended for network performance or usage monitoring, not for cybersecurity purposes. They are typically not well adapted to data-driven methods of detection due to the lack of industry-wide standards for data annotation and logging. This can be problematic when trying to apply automated large-scale analytical methods to the data. Finally, in order to maximize the potential value of a single data source, it often has to be combined with other data sources, such as host configuration, and link traffic data among those sources listed at the beginning of Section 3. Standardizing and indexing such a large variety of data to enable analysis requires significant engineering effort. Industry or regulatory standardization of protocols and procedures for data collection, logging and curation of integrated cybersecurity databases would facilitate development of network monitoring tools.

Despite decades of research in statistics and data analytics for the large and expanding area of cyber defense, there remains a lack of consensus on best statistical practices. The growth over recent years of internet-connected devices and operational technologies has caused a dramatic increase in the amount of data collected, making the need for automated and scalable statistical methods more acute. In addition, there is a rise in the number of new measurement sensors being placed on networks, and with it new data sources capturing network behaviors that previously were not being monitored. This drastic rise in the amount of data and the ever-shifting attack landscape has created an abundance of new research opportunities in data fusion and machine learning for security applications.

Machine learning and even deep learning have been applied to fuse and take advantage of the multiplicity of data sources (Sarker et al., 2020), for example, for malware detection and prevention, (Dua & Du, 2011; Mahmood & Afzal, 2013). Many of these algorithms are supervised, and are used to extract features about the data that allow them to identify new instances of malware due to similarities in the underlying feature space. This has led to a reduction in the impact of first-seen (known as 0-day) malware that behaves similarly to samples already observed. However, the success of supervised methods crucially depend upon their training sets being representative of future attacks. In the malware space, machine learning methods have been a successful generalization from simple signature matching, enlarging the space of detectable malware. However, malware authors have proven adept at finding new ways of exploiting vulnerabilities in cybersecurity software, in particular, by launching attacks that differ slightly from those represented in the training sets.

Malware is not the only piece to the security puzzle. Adversaries have developed many ways of gaining access to computer networks, not all of which require the exploitation of software. In systems without dual authentication, one only needs a valid user's credential (password) to access the account of that user. From there, privilege escalation may not involve exploitation and thus we are seeing malware-less attacks become more common (FireEye, 2018). In many cases, given the wide variety of options available to the attacker, truly representative training sets simply do not exist. Even if such training sets could be generated, one can expect an adversary to be able change behavior rapidly, rendering these training sets inaccurate. Unfortunately, much of today's ML and AI software only works when there is clean training data with little noise and no malicious contamination, and when future test data has identical properties as the training data. Although existing software has the capability to process large amounts of data generated by cybersystems in real time, the underlying assumptions of clean training data and identical training and test distributions do not hold. Indeed, sophisticated adversaries often purposely generate adversarial samples at the training and the test stage, causing the software to make crucial mistakes.

Data-driven AI methods have recently been touted as the solution to the deficiencies of rules-based security methods. However, traditional AI is a double-edged sword (Taddeo et al., 2019) since AI can fail in unpredictable ways, and this vulnerability can be exploited by an attacker. For example, the huge number of

degrees of freedom of deep neural networks can enable very high predictive accuracy when the training data is representative of the test sample. However, a carefully chosen perturbation of a test sample can severely degrade the prediction accuracy, resulting in vulnerability to adversarial attacks. The translation of recent advances, such as adversarial machine learning and transfer learning, can improve predictor robustness. However, to date these have not yet matured sufficiently to be widely incorporated into enterprise security software but there has been a recent push in these directions, going beyond the traditional security solutions (Sarker et al., 2020). Indeed, cybersecurity tasks require higher degrees of robustness and security assurance than traditional tasks performed by AI. Improving AI in complex data environments, under the constraints of limited training data containing malicious samples and corrupted test samples during operations, will enhance the security of the entire cybersystem. Furthermore, we need to consider all the existing ML techniques, not just the most popular ones such as deep learning, to address the full spectrum of challenges in processing and analyzing data in distributed and complex systems.

Another important challenge in improving enterprise security is anomaly detection. Specifically, it will be important to develop online streaming anomaly detection methods that can accurately discriminate between anomalous deviations due to outliers versus those due to attackers (Akoglu et al., 2015; Patcha & Park, 2007; Ranshous et al., 2015). This will likely involve stochastic modeling, the use of dynamic graphs, and applying concepts of statistical control. While much of cyber research is centered around anomaly detection, there is still much work to be done in real-time detection and quantification of anomalies within complex cybersystem data sets. Building models that capture normal behavior across multiple networking layers are needed. This will include modeling of normal device behavior, behavior of users, and pairwise interactions (computer-to-computer), operating system behaviors that will make it easier to detect deviations introduced by the attacker. A major challenge is modeling anomalous data since, by definition, anomalies are extremely rare. Machine learning methods such as few-shot learning that learn anomalous behaviors from very few labeled examples seem promising (X. Zhou et al., 2020). Classical statistical approaches to quickest change point detection in multivariate time series (Killick & Eckley, 2014; Lavielle & Teyssiere, 2006) are also promising.

Challenges in Data-Centric Cybersecurity

- Lack of industry-wide standards for data annotation and logging
- Learning from complex data that is contaminated, heterogeneous, and dynamic
- Training anomaly detectors with very few labeled samples

---

3. Data-Driven Cybersecurity for Enterprise Systems

Network data collection and analysis is critical to a good cybersecurity posture. Many different types of relevant data can be captured from sensors both on the on-premise network and more recently in enterprise cloud environments. The level of detail provided by such sensors can range from full packet capture (PCAP)

data to aggregated network traffic data to detailed logs of activity happening on network endpoints. Such data can enable anomaly detection, risk characterization, and overall enhanced network awareness. If data were collected continuously, such functions might be accomplished in real time, providing early warning to network administrators and users.

Many existing data sources are currently available for cybersecurity, including the following in the area of internet security: 1). Network Flow Data (NetFlow): A summary of unidirectional flow of network IP packets providing a high-level summary of network connection events between pairs of devices; 2). Web proxy logs: A meta-level summary of requests from clients to servers outside the network; 3). Domain Name System (DNS) logs: DNS maps fully qualified domain names (FQDN) to IP addresses, which when combined with other cyber data, like NetFlow, can improve anomaly detection and tracking; 4) E-mail logs: E-mail logs can be collected from mail servers, packet capture, or other sources and provide meta-level information about e-mails.

There exist many public data sets containing data from these sources and these have been useful for development at the interface between network cybersecurity and data analysis (Canadian Institute for Cybersecurity, 2019; Cyber Systems and Technology Group, 2019; Glasser & Lindauer, 2013; Kent, 2014, 2015; J. Ma, 2019; Malware_traffic, 2019; Turcotte et al., 2018; US Department of Homeland Security, 2019). Some of these represent data collected from real operational environments, while others are simulated or pseudo-generated.

3.1. Anomaly Detection

Until the last decade, commercial cybersecurity tools for detecting anomalous events was dominated by signature-based methods (A. Z. Li & Barton, 2019). Next generation anomaly detection tools employ either supervised anomaly detection, which uses previously observed anomalous data in the training, or unsupervised anomaly detection that requires only nonanomalous training data and equates anomalies with rare events in the training data. Even though unsupervised anomaly detection methods are capable of higher sensitivity, especially for detecting novel attacks, a principal reason that they have not been more widely adopted in practice is their inherently higher false positive rate relative to supervised methods (Patcha & Park, 2007). Furthermore, no matter how good the model, many rare events in the data are benign, and the differences between malicious and benign data are subtle and require context to resolve them (Hayes & Capretz, 2015), for example, using additional historical data to narrow down the peculiarities of a cyberattack.

When a large number of data streams can be used for potential anomaly detection, it is common to independently train multiple anomaly detectors on each stream and use Boolean combining rules to aggregate the outputs of the detectors. For example, for the OR aggregation rule an anomaly is declared when at least one of the detectors raises an alarm. For a fixed anomaly detection threshold, as the number of false alarms increases in the number of streams. Conversely, if the thresholds are adjusted to provide a desirable level of family-wise false positive error rate, the detection power decreases with number of streams, that is, the size of

the anomaly must increase to be reliably detectable. This is the problem of multiple statistical comparisons (Miller, 1981) and it is a fundamental limitation for anomaly detection. Hence, research into new approaches to anomaly detection that can integrate multiple data streams without substantial loss of detection power is a worthwhile area for computer scientists and data scientists to pursue. An interesting approach that uses analyst feedback to automatically tune anomaly detection procedures (Das et al., 2017) is an alternative.

Another promising approach to handling the dichotomy between benign and malicious anomalies is to use the Bayesian theory of evidence (Berger, 2013). This could entail specifying a subjective prior on the benign and malicious signatures, which is then used to marginalize the likelihood function, obtaining the respective likelihoods of the data under benign and malicious hypotheses, called the evidence. The likelihood ratio could then be used to discriminate between these two types of anomalies. A related approach was adopted in (Hou et al., 2018) to distinguish between low- and high-utility anomalies when there is a utility prior.

3.2. Adversarial Machine Learning

Cybersecurity is by nature adversarial. It is frequently referred to as an arms race (Taddeo & Floridi, 2018) where more powerful security measures are overcome by malicious adversaries who learn and exploit system vulnerabilities. As discussed above, traditional AI/ML is not robust to adversarial attack because it is not designed to be. Adversarial ML is a growing field within AI that seeks to develop principles for robustifying machine learning methods against adversaries. Out of the many possible adversarial ML formulations, the most relevant to cybersecurity is the game theory formulation.

Adversarial attack strategies can be broadly categorized as poisoning attacks or evasion attacks. Poisoning attacks attempt to contaminate the training process (Biggio et al., 2012; Xi, 2020) by manipulating the training data to degrade the learned classification regions of the classifier. Learning systems that are periodically retrained, such as network intrusion detection systems, are especially vulnerable to poisoning attacks. An early example of poisoning attacks were attacks on email spam filters that injected incorrectly labeled spam emails during training (Nelson et al., 2008). Evasion attacks degrade classifier performance after training (Biggio et al., 2013) by manipulating test samples to exploit classifier vulnerabilities. Deep neural networks (DNN) are especially vulnerable to evasion attacks due to the very high dimension and complexity of their learned class decision regions. Gradient-based white box attacks (e.g., (Athalye et al., 2018; Carlini & Wagner, 2017; Sharif et al., 2016) assume adversaries have complete knowledge of the trained DNNs (training set, architecture, weights, and activations). When adversaries do not have complete knowledge of a DNN, they can launch a black box attack that uses repeated probes to learn successful adversarial samples (Papernot et al., 2017). In intrusion attacks, black box attack strategies have been applied to generate new malware that are able to avoid detection (Anderson et al., 2017).

Game theoretic approaches. Game theory can be used to model the interaction between effective attackers and defenders of a network. Such models can then be used to put up defenses that are resilient to even the strongest

attackers. Several game theoretic models have been proposed in the literature on adversarial learning in cybersecurity applications. (Brückner & Scheffer, 2009; Dalvi et al., 2004; Do et al., 2017; Pawlick et al., 2019) used a two-player static noncooperative zero-sum game formulation to model various adversarial learning scenarios, where one player is a classifier/learner and the other player represents the group of adversaries. The minimax solution to this game, characterized by the Nash equilibrium, suggests constructing attack-resilient classifiers by minimizing worst-case misclassification error (Dekel et al., 2010; El Ghaoui et al., 2003; Globerson & Roweis, 2006; Lanckriet et al., 2002; Teo et al., 2007).

The formulation of the cybersecurity problem as a zero-sum and static (simultaneous play) game can be criticized as unrealistic Fielder et al., 2014, 2016. The zero-sum assumption is unrealistic since attackers and defenders generally have different notions of damage inflicted, for example, the attacker may be seeking to maximize inconvenience to the users while the defender may be seeking to minimize the financial cost of detecting or mitigating the attacker. The static formulation is unrealistic since real cybersecurity plays are not simultaneous. Rather, they are sequential and thus dynamic; a defender takes an action only in response to an attacker's action. An alternative approach is the Stackelberg game that allows players to take sequential actions. In a Stackelberg game, the player who makes the first move is the leader and the other player is a follower, choosing his own action only after observing the leader's action. Every player maximizes its own utility function, allowing for non–zero-sum rewards. The game's equilibrium strategy then leads to a more realistic cybersecurity defense model compared with static zero-sum models (Brückner & Scheffer, 2011; Fielder et al., 2014; Kantarcıoğlu et al., 2011; Liu & Chawla, 2010).

3.3. Practical Considerations and Future Directions

A principal bottleneck in large-scale data analysis is data preprocessing, which may include data parsing, data reduction, data cleaning, or data integration (García et al., 2016). Data preprocessing must extract the most informative features to enable effective performance of analysis methods described in previous subsections. An additional challenge in cybersecurity is that adversaries can attack vulnerabilities in the data preprocessing step itself. For example, in anomaly detection it is common to preprocess high-dimensional data streams by applying data-reduction methods like principal component analysis (PCA). Such methods project the data onto a data-determined lower dimensional subspace, which can be attacked by an adversary by perturbing entries in the high-dimensional data matrix (F. Li et al., 2020). Statistical methods for robustifying PCA and other data reduction technical against adversaries can be effective (Braverman et al., 2021). However, there remains much more work to do to, both in terms of practical implementation and on certifying robustness for cybersecurity applications. Statistical models are especially challenging to develop when the data is nonstationary, as in network traffic, and the provenance of statistical changes difficult to interpret (Heard et al., 2014; Price-Williams et al., 2017).

An important challenge is the lack of sufficient quantities of representative data sets that can be used to build and test statistical modeling for cybersecurity. While there are some existing data sets, as described at the

beginning of this section, there is limited publicly available data representing both baseline data and adversarial attack behaviors. Thus, a central repository of data in a standardized format would be an invaluable resource for researchers. Central data repositories had significant positive impact on research progress in other scientific domains, for example, the NCBI-GEO database (Barrett et al., 2012) for genomics, and would similarly enhance the progress of cybersecurity research.

Challenges Cybersecurity Data Analysis

- Multistream anomaly detection and the multiple comparisons problem
- Nonstationary data and evolving adversarial attack strategies
- Finding realistic public data sets for model building and testing

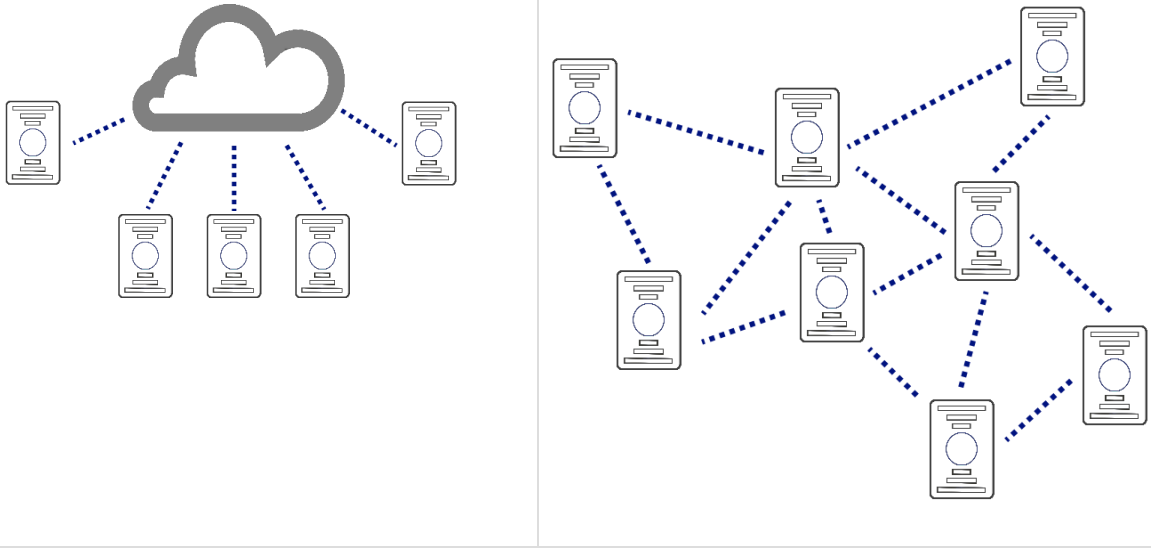
---

4. Data-Driven Cybersecurity for the IoT

The Internet of Things (IoT) is defined as a network of devices that interface with the physical world with at least one sensor or actuator and interface with the digital world with at least one network interface, for example, Ethernet, Wi-Fi or Bluetooth (IoT Act of US Congress, 2020). National security standards for IoT are emerging (Fagen et al., 2021), but they focus on protecting individual devices in centrally administered network, for example, consumer IoT home networks and telehealth monitoring networks. As the IoT expands, its large scale and heterogeneity will make central administration impractical and ad hoc peer-to-peer network administration will become the norm. Such non-centrally administered networks present cybersecurity challenges different from those encountered in centrally administered systems. In a decentralized network, devices act as independent agents who communicate with each other and collectively come to consensus without central control or administration. This creates more points of vulnerability that can be exploited by an adversary; who can attack the data-sharing infrastructure, the consensus process, or the devices themselves. Distributed denial of service (DDOS), man in the middle, or jamming attacks can be especially damaging because the lack of centralized infrastructure makes the damage from a localized attack difficult to contain. An attack on data (measurement attack), an attack on devices (Byzantine attack), or an attack on data links (jamming attack) are vulnerabilities specific to the distributed IoT. Decentralized IoT will benefit from practical distributed computing and distributed inference strategies to detect, identify, and mitigate such attacks.

The functionality and security of IoT systems depends on the underlying system architecture, either (centrally) coordinated or ad hoc with no central coordination. In a coordinated architecture, individual devices forward their local data to a fusion center, possibly in the cloud, which is then responsible for data processing and decision-making (see Figure 1a). As IoT devices become more intelligent, they may carry a greater portion of the computational burden. For example, in Fog and Edge computing setups, a significant amount of computing is done locally at the device level instead of relying solely on a cloud fusion center (Chiang & Zhang, 2016). In a fully distributed uncoordinated architecture, the end devices cooperate with each other over a peer-to-peer

communication network to achieve a common objective (see Figure 1b). The networked team of devices carries all of the computational burden and processes their local data without the aid of a cloud fusion center.



(a) Architecture with Central Processor

(b) Fully Distributed Ad Hoc Architecture

Figure 1. (A) In centralized architectures, devices transmit their data to a fusion center, and there is no device-to-device communication. (B) In distributed (uncoordinated) architectures, devices cooperate with each other over a peer-to-peer communication network without the aid of a fusion center.

Decentralized architectures, leveraging Fog and Edge computing, scale more easily than centralized architectures because they take advantage of the computational power of the individual end devices themselves (Chiang & Zhang, 2016). They are also more suitable than centralized architectures for applications that may be latency sensitive or for applications in which it is difficult to establish a central coordinator (Y. Chen et al., 2018a). For example, convoys of autonomous vehicles may cooperate to coordinate traffic at an intersection without the need for traffic signals or to collectively navigate through hazardous weather conditions (Lu et al., 2014; Uhlemann, 2018). As a consequence of real-time processing requirements of controlling autonomous vehicles, the computations need to be performed at the vehicle, instead of in the cloud.

The security of the network and the trustworthiness of the data loom as a prevalent issue in decentralized IoT architectures. A malicious adversary may launch a Byzantine agent attack that hijacks a subset of the devices, manipulates their local data streams, and controls their behavior. For example, once an adversary gains device control he might spoof LIDAR on autonomous vehicles to create false, artificial obstacles (Harris, 2015), compromise sensors or actuators in modern automobiles or robotic platforms (Franchetti et al., 2017), or control UAVs by falsifying their onboard sensor measurements (Davidson et al., 2016). As IoT applications continue to scale up, it becomes increasingly difficult to prevent security intrusions on all of the devices, and it becomes more likely that at least some of the devices will fall under adversarial attack and control (Y. Chen et al., 2018a). With a sufficient number of hijacked devices, the attacker can effectively cripple the collective

inference and decision-making process of the network, in emulation of the attack strategy in the Byzantine Generals problem (Lamport et al., 1982).

The principal challenges in dealing with Byzantine attacks on the IoT are to detect and identify adversarial hijacked devices; and to design consensus algorithms that are resilient to a small number of adversarial hijacked devices. To meet these challenges, computation and inference methods must necessarily be distributed and secure. Distributed methods have been proposed to detect and identify adversarial devices (Pasqualetti et al., 2012; Pasqualetti et al., 2013), but these are impractical for large networks since they require each individual agent to know the entire network's topology. It has been shown that the correct consensus decision can be guaranteed if and only if fewer than a third of the devices are compromised (Lamport et al., 1982). Resilient consensus algorithms can be used to attain this guarantee under certain restrictions on the network topology (LeBlanc & Hassan, 2014; LeBlanc et al., 2015). However, such methods only provide asymptotic guarantees as the number of message passing steps goes to infinity, that is, incurring infinite delay. Methods providing guarantees in practical systems with bounded delay are lacking.

While the Byzantine attacker gains control of a few devices in an attempt to disrupt the collective decision-making process, the measurement attack attempts to manipulate the primary data streams at the sensors without necessarily taking control of the devices themselves (Y. Chen et al., 2018b; LeBlanc & Hassan, 2014; LeBlanc et al., 2015; Sundaram & Hadjicostis, 2011). Attackers can inject noise or interference into the sensor data in an attempt to contaminate the entire data-processing pipeline across the network. Similarly to the Byzantine attack, the main challenges are to detect manipulation of the data streams; and to design data collection and aggregation protocols that are resilient to moderate amounts of contamination. Many methods for defending against measurement attack are based on robustifying the statistical estimator that derives its estimates from the contaminated data. An example is the Saturating Adaptive Gain Estimator (SAGE) that uses a locally weighted nearest neighbor method with an adaptive gain (Kar & Moura, 2011, 2013). Similarly to other robust statistical procedures, for example, the trimmed mean for outlier resistance, robustness comes at the expense of reduced accuracy when no attacker is present.

4.1. Practical Considerations and Future Directions

From early work on decentralized multiparty computations in Byzantine adversarial scenarios (Dolev et al., 1986; Lamport et al., 1982) to more recent developments, there has been substantial interest in secure and resilient inference and optimization in several types of decentralized settings. For example, (Kailkhura et al., 2015; Vempaty et al., 2013; J. Zhang et al., 2017; J. Zhang et al., 2015) focus on fusion center–based architectures, while recently federated decentralized server-client architectures have received significant attention (Kairouz et al., 2021), So et al., 2020. In federated architectures, different strategies are needed to achieve security goals including robust aggregation and order statistical approaches (Pillutla et al., 2022; So et al., 2020). However, most strategies require centralized coordination and are therefore not directly applicable to the decentralized networked setting. Decentralized alternatives, such as the SAGE approach described in

earlier subsections, are often based on consensus or distributed optimization, which will leverage on advances in distributed approaches to machine learning, signal processing, and statistics Nedic, 2020, Jordan et al., 2018. While a detailed analysis of existing security approaches in wireless, potentially ad hoc networks of the type described in earlier subsections is beyond the scope of the current article, we point the reader to a survey (Z. Yang et al., 2020) for in-depth discussions.

There are several worthwhile future research directions. For instance, most of the current work in distributed network security focus on statistical guarantees, such as consistency in the presence of adversaries. However, formal optimality or efficiency guarantees are generally lacking. To understand in fuller generality the factors influencing convergence in secure decentralized ad hoc networks, a worthwhile future direction of research will be to obtain algorithm-agnostic convergence bounds. Such bounds have proved useful for distributed machine learning without security guarantees (Nedich, 2015) and these could likely be extended. Another worthwhile direction would be the development of unified approaches to design of resilient stochastic distributed optimization with heterogeneous agents and cost functions.

We end this section with a discussion on the challenges of implementing security measures in the distributed IoT context. The nodes in a typical IoT network are resource constrained (cost, size, and power) and it is critical to design schemes for distributed learning and inference that are not only resilient to adversarial attacks but are modest in terms of communication and computing requirements. The complexity of distributed processing tasks can range from relatively simple computation tasks, such as regression (as discussed in this article) (Fodor et al., 2021), to complex machine learning tasks such as training deep neural networks (Lian et al., 2017). Several architectures have been proposed that are adapted to different node computation and communication capabilities. There has been an increasing push to implement on-device model training for fairly complex ML tasks, such as natural language processing (NLP) in IoT type platforms. For instance, prototypes of federated learning (Kairouz et al., 2021) have been implemented on standard smart home devices such as Google Home and Amazon Alexa (H. Zhang & Kim, 2021). Furthermore, distributed consensus–based ML and inference algorithms, with fairly complex tasks, of the type discussed here have been implemented in for wireless IoT networks using nodes equipped with Rasp- berry Pi devices (Upton & Halfacree, 2014). Additionally, lightweight inter-device communication protocols to enable agent cooperation have been prototyped using the message passing interfacing (MPI) standard. While somewhat out of scope of the current article, we point the interested reader to (Kar et al., 2020) and its references for descriptions of Raspberry Pi and MPI-based implementations of fully distributed consensus+innovations procedures for certain classes of optimization and inference tasks. Secure and resilient variants of these procedures can be readily integrated into these implementation frameworks, for example, by suitable modification of the innovation terms for simple tasks like regression. For more complex tasks we refer the reader to discussions of distributed consensus–based training for distributed machine learning (Lian et al., 2017).

Cybersecurity Challenges for Decentralized IoT

- Ensuring security and trustworthiness of data in decentralized ad hoc networks
- Distributed detection and localization of attacks on devices and links
- Achieving consensus among devices in an adversarial environment

---

5. Ensuring Information Privacy in the IoT

In applications where data products are constructed from sources containing sensitive or confidential information, the confidentiality and privacy of this information must be protected. The most basic protection measure is to remove sensitive data fields from the data products. However, this may be insufficient because a resourceful adversary may be able recover these fields by other means, for example, correlating the data product to external data sources. Certification that a data product has a guaranteed level of privacy protection is a challenging problem, especially when data is collected, aggregated, and transmitted over a distributed network like the IoT. Such certification would need to account for the integrity of the entire data pipeline; from data collection and summarization at the device level to wireless data transmission between devices, to data aggregation and summarization into the data product used by cybersecurity analysts or anomaly detection.

At a fundamental level, privacy certification is a problem of information theory. Assume that someone proposes a putative privacy protecting mechanism that maps a data set (input) to a data product (output), that is, a possibly noisy summarization of the data. The principal objective is to theoretically quantify the trade-off between leakage of sensitive information and the accuracy of the summarization. The general framework of differential privacy (Dwork, 2008) captures this trade-off by determining the statistical indistinguishability between the probability distributions of two outputs corresponding to two neighboring inputs. Algorithmic differential privacy is concerned with finding algorithms that achieve a desired trade-off between privacy and accuracy (Dwork & Roth, 2014). Mutual information differential privacy (W. Wang et al., 2016) models the input-output mechanism as a channel and uses rate-distortion theory to establish fundamental limits on the privacy vs. accuracy trade-off. Algorithmic and information theoretic approaches to differential privacy are complementary and, under certain simplifying assumptions, they are equivalent (Cuff & Yu, 2016; Makhdoumi et al., 2014).

Guaranteeing privacy in the IoT networks is more challenging than in centrally administered networks like the internet. IoT networks are used primarily for data gathering, inference, and control rather than for telephony, texting, or media distribution. The heterogeneity of device capabilities, the diverse data types, and the differing degrees of data sensitivity prohibit one-size-fits-all approaches to privacy protection. Furthermore, low latency and high throughput are especially important for real-time IoT applications involving feedback control like production automation or autonomous driving. Most importantly, there are differences in the data transport layer as interdevice communication is typically in the form of short packets, rather than long packets as you would have in media distribution. The IoT infrastructure often depends on ad hoc wireless protocols for medium access control and dynamic resource allocation. This can create additional privacy vulnerabilities, for

example, susceptibility to eavesdroppers who attempt to intercept data during transmission. As a consequence, protection of privacy in the IoT must take into account vulnerability of the communication channels between devices. Thus, communication and information theory will play a crucial role in implementation and certification of privacy and confidentiality in the IoT,

5.1. Physical Layer Security

IoT networks are characterized by a layered hierarchy of data flow. The IoT consists of three principal layers: the *application layer* where the application protocols are running, for example, software for data analysis; the *network layer*, which is responsible for moving messages around in the network; and the *physical layer*, which is where the actual data transport takes place, that is, through the physical medium that supports network operation. For the wireless parts of the IoT this medium is the 'ether' through which radio signal propagate.

The main method for securing data is encryption at the application layer, which can be very complex and add delay. Low latency will often be a system requirement in IoT systems that support real-time applications, such as autonomous driving, and delay caused by decryption can pose problems. Moreover, if the network is comprised of low-complexity terminals, the complexity of public key cryptography can be too high. Furthermore, without infrastructure to handle key management and authentication, it is generally difficult to implement encryption.

An alternative way of approaching data confidentiality in the IoT is to examine the security of the physical layer. The fundamental approach to addressing this issue is *information theoretic security*, which originated with (Shannon, 1949) who addressed the security of cipher systems such as that illustrated in Figure 2. The cipher system consists of a transmitter, Alice, and a legitimate receiver, Bob, and Alice wants to send a message, the data M, to Bob reliably, while keeping it secret from an eavesdropper, Eve. An information theoretic guarantee, called perfect secrecy, can be achieved if Bob and Alice share a key that has complexity, as measured by Shannon entropy, at least as high as the message that is being sent. Perfect secrecy would completely protect the data from an eavesdropper (Eve) but it is impractical due to the need to communicate the high-complexity key in addition to the data. Most encryption systems rely instead on the computational difficulty of decryption to achieve a practical, rather than absolute, degree of secrecy.

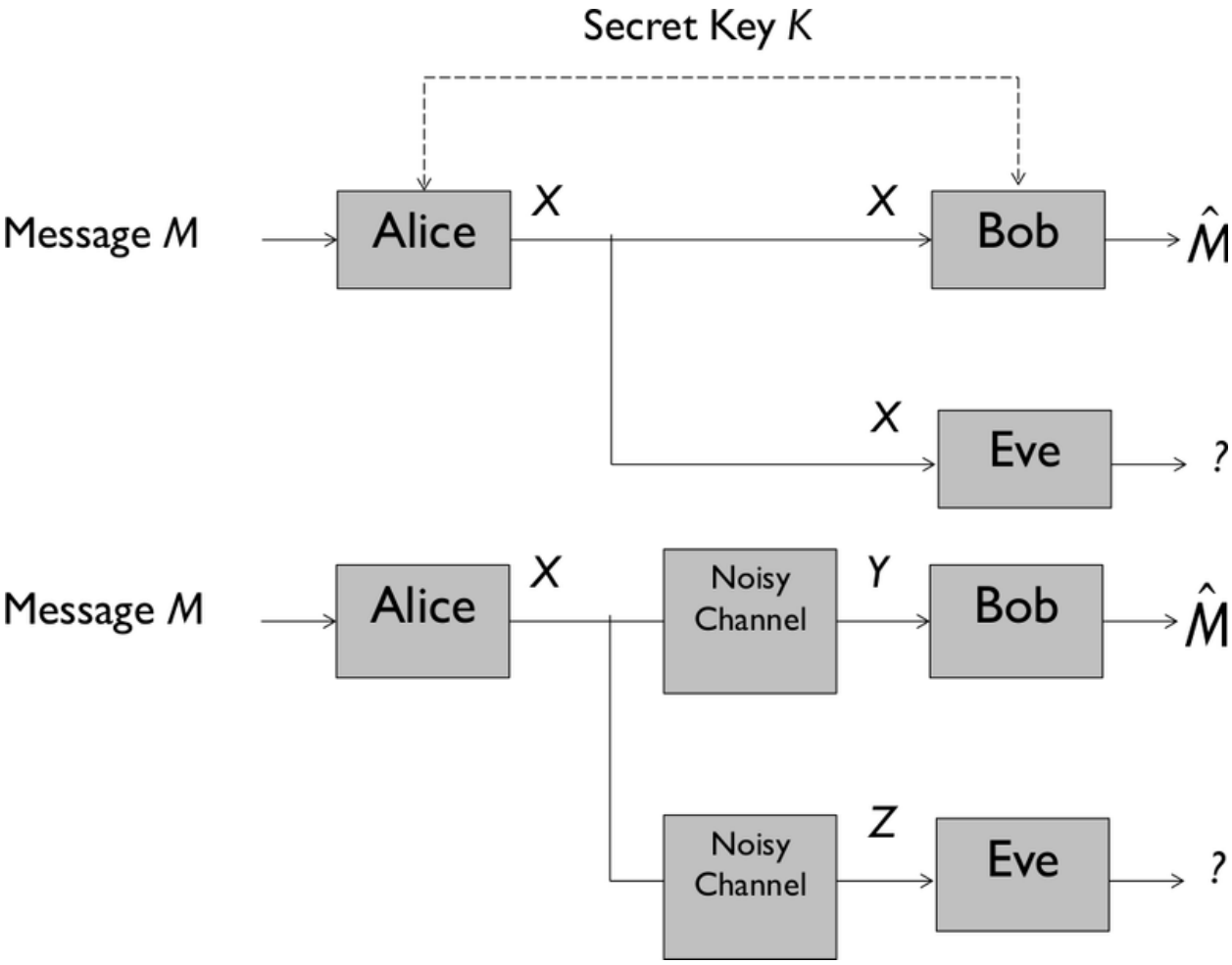


Figure 2. Left: Shannon's model of a cipher system. Right: Wyner's wiretap channel. These provide models for establishing closeness to optimality of methods for achieving physical layer security in distributed networked systems like IoT.

Another information-theoretic way of looking at transmission of data with confidentiality is due to (Wyner, 1975), in which he looked at the same model of Alice, Bob, and Eve, but eliminated the need for a shared key, recognizing instead that the noisiness of communication channels can be exploited to provide information security guarantees, as illustrated in the right panel of Figure 2. Wyner addressed the issue of whether one can transmit in secrecy over such a channel using only the differences in the noisy channels between Alice and the two receivers to provide confidentiality. In particular, he looked at the trade-off between the reliable rate at which the message could be transmitted to Bob and the so-called *equivocation* about the message in the signal received by Eve. The equivocation here is the conditional entropy of the message M given the signal received by Eve, which is a measure of how much randomness is left in the message after Eve sees her received signal. To preserve the secrecy of the message, this quantity should be large, and if the equivocation is the same as the entropy of the transmitted message, then no information about the source can be understood by Eve.

In this context, Wyner defined what he called the *secrecy capacity* of a channel, which is the maximum rate at which a message can be sent reliably and in perfect secrecy from Alice to Bob. He showed that the secrecy capacity can be positive, but if and only if the signal received at Eve is *degraded* in a certain sense relative to that at Bob; essentially this means that the channel to Eve is somehow worse for data transmission than the channel to Bob.

The concept of secrecy capacity is especially meaningful for networks of data collecting devices in the IoT. From Shannon's and Wyner's work, it can be seen that generally what is needed for secure data transmission is for the legitimate receiver to have some kind of advantage over potential eavesdroppers. This can be achieved by sharing a secret key, as in Shannon's model, or using a better channel for the link, as in Wyner's model. For the latter, physical properties of the transmission channels in the network can be used to provide this kind of advantage. For wireless links, three of these properties are: channel fading, interference, and channel dimension. Fading occurs when the transmitted radio waves undergo multiple reflections along the propagation path, producing deep fades over time that degrade the eavesdropper's channel. Interference occurs when structured noise degrades the eavesdropper's channel and can be accomplished through jamming. The dimension of the channel can be increased to make it more difficult for the eavesdropper to intercept messages, for example, through the use of directional antennas or relays, increasing the *secrecy degrees of freedom*. Exploitation of these and other channel properties can provide physical layer security and privacy protection in large networks like the IoT (Poor & Schaefer, 2017). Physical layer security has a significant advantage in the setting of IoT, where devices are of much lower complexity than those in cellular communication networks. This is because using the channel to secure messages is of similar complexity to that of error-control coding, which is much lower than that of typical cryptographic algorithms.

5.2. Privacy-Utility Trade-Offs

There is no free lunch in the endeavor to protect data privacy, which is governed by a *privacy vs utility trade-off.* Specifically, sensor systems in the IoT generate a lot of electronic data, the utility of which depends on its accessibility to the users of the data. But that accessibility then endangers privacy. To achieve a desired trade-off one must balance two extremes. Completely accessible data is as useful as possible, while completely inaccessible data is as private as possible. Hence, as in the trade-off between adversarial robustness and accuracy discussed in Section 4, there is trade-off between privacy and utility, which can be characterized using methods of information theory. The degree of privacy is quantified by information leakage, measured by the equivocation (conditional entropy), while utility is quantified as fidelity of the accessible data relative to the original data. With this view of the privacy vs. utility trade-off, if we are given a statistical model for the data, we can create a region of achievable pairs of utility and privacy where the outer boundary of that region is the efficient frontier of the trade-off; that is, if we want to demand a certain amount of utility, this boundary tells us the minimum amount of privacy that must be sacrificed, or vice-versa. This type of analysis has been applied to the IoT (Poor, 2018) and to smart metering in the electricity grid (Giaconi et al., 2020).

5.3. Practical Considerations and Future Directions

Information theory can provide fundamental limits and insights on protecting information privacy and security in the IoT. However, information theory was not conceived to provide practical solutions directly, but rather to indicate how close practical methods are to such limits. Information theory has played a similar role in the development of telecommunications, networking, and information systems more generally. However, information theory often points to practicable ways of enhancing security or privacy, for example, wiretap codes and secure compression. However, methods based on information theoretic ideas have for the most part not been implemented in practical systems, and thus one of the challenges is to introduce these into practical settings for providing security or privacy guarantees in settings where the traditional privacy protection methods may not be practical, for example, in the IoT. This challenge is currently being addressed by standards working groups working toward the 6th generation (6G) of wireless networks, which, among other features, are expected to support mas- sive connectivity of low-complexity devices. It should be noted that classical information theory is an asymptotic theory, and thus one challenge in applying it to practical systems is to develop nonasymptotic, finite-block-length characterization of fundamental limits. Progress in this direction has been made (W. Yang et al., 2019) although there is still much work to be done.

A relatively recent approach to addressing issues of privacy in machine learning and data science applications over networks is federated learning, which is a distributed learning paradigm in which multiple end-user devices, or clients, work together to collaboratively train a model. They do so by each training a local version of the model based on its local data set, and then sending these local models to an aggregator, or server, to be combined into a stronger model. This aggregated model is then returned to the clients for further iteration (Nguyen et al., 2021). Such a model provides a degree of privacy protection of the client data since it remains on the end-user devices, and only the model parameters are shared with the aggregator. However, privacy can still be compromised by sharing the models, and thus further privacy-protection steps have to be taken. This gives rise to another trade-off, which is the trade-off between client data privacy and performance of the machine learning algorithm (C. Ma et al., 2020) to perform, for example, global network-wide anomaly detection. This area is still in its relative infancy for wireless networks, and thus also represents an opportunity for further innovation, again for 6G, which is expected to incorporate AI-at-the-edge into many aspects of its operations and services. Information-theoretic methods will likely have a role in understanding the fundamental limits (Yagli et al., 2020).

Information Privacy Challenges for the IoT

- Implementing privacy mechanisms across heterogeneous devices and links
- Balancing information security against computation and latency
- Balancing local data privacy against global anomaly detection

---

6. Discussion

The previous sections have described some of the challenges in cybersecurity for centrally ad- ministered enterprise systems and distributed IoT-like systems. While the technologies for assuring cybersecurity for these types of systems may differ in the details of implementation, they will both benefit from progress in data collection, statistical analysis, and mathematical modeling. As pointed out in Section 3, the quality of the training data is the primary bottleneck in developing effective strategies to thwart adversaries. Furthermore, classical supervised machine learning approaches will be inadequate as gaps in training data collected in the past will always be exploited by future adversaries. For such a cat-and-mouse game, as respectively discussed in Sections 3, 4, and 5, game theory, distributed computational theory, and information theory will enable us to both understand fundamental limits and to craft more effective countermeasures. Perhaps paralleling the success of information and communication theory in guiding the telecommunications industry over the past 60 years, data science and modeling may lead to advances in cybersecure networks. This is a great opportunity for technologists, engineers, and computer scientists to work together with data scientists and mathematicians to develop the next generation cybersecurity systems.

There are other challenges that go beyond what has been discussed in this article. The authors believe that there are major hurdles that will require going beyond purely technological solutions. An IoT world with millions if not billions of devices instrumenting our physical world will require fundamentally new approaches to model and detect attacks and anomalies. To effectively secure systems against attacks, designers must account for the motivations of people, including both the users–why do they click on a phishing email?–and the intruders–what motivates them to act that way? Accounting for and modeling such motivations are among the most challenging questions for cybersecurity and will benefit from the insights of psychological and behavioral scientists. Involvement of such scientists in developing models for attackers' and users' behavior is arguably essential for keeping up with the human-driven cybersecurity arms race.

In the future, companies will benefit from creating a longer priority list of anomaly scores that goes beyond what is reported by standard intrusion-detection software. These anomaly scores should be computed based on more than one metric in order to capture diverse types of deviation of a signal from its baseline. Even though it may increase false positives, active and iterative searching through the network for threats will be a much more effective active defence technique. This technique is referred to as 'cyber threat hunting' (Milajerdi et al., 2019) and will allow companies to detect the rarest events, that may also be the most fatal, even if it happens but once a year or every couple of years. Companies must be willing to suffer through many false positives to make sure they catch the black swan. Operators may need to relax their tolerances for false positives and keep looking at more data to catch those extremely damaging events. As discussed in Section 3, statistics and machine learning will be needed to reduce the multiple testing bias associated with multistream anomaly detection and formulate more effective approaches to benign vs. malicious anomaly classification.

As alluded to in Section 4, the prevailing concept of the IoT is that it is a network of devices having limited computational ability. Future technological advances may overcome this limitation, for example, via breakthroughs in distributed computing or quantum computing, providing cheap powerful devices with essentially unbounded computational resources to counter cyberattacks. However, such advances are not likely to settle the debate on advantages and disadvantages of distributed vs. centralized network control architectures. Centralized control and coordination represents a single point of failure, whereas in a distributed setting, failure of a few nodes may still leave many other connected nodes operational in sufficient number to launch a defense against the attack. On the other hand, a centralized architecture has immediate access to all available data, while access to distributed data will be limited by link capacity, privacy concerns, and latency. Distributed architectures will continue to present challenges for diagnosis and countermeasures since subnet operators will need to piece together all data in order to detect and respond to an attack. A more practical alternative may be a compromise between a centralized and distributed cyber-architecture, switching between the two as threat levels change.

7. Conclusion

This article highlighted research questions at the forefront of cybersecurity lying at the inter- section of statistics, machine learning, and information theory. Fundamental advances are needed to address emerging applications domains, including adversarial machine learning, attack detection in enterprise networks, and security and privacy in distributed networks, such as the Internet of Things (IoT). There are many opportunities for statisticians, engineers, and computer scientists to come together to develop the next generation of cybersecurity practices.

Acknowledgments

This paper would not have been possible without the support of the National Academies of Science, Engineering and Medicine and the US National Science Foundation (Division of Mathematical Sciences), Award number 1700083.

Author Contributions

All authors contributed equally to drafting the sections in this article. AH led the effort to integrate these sections into a coherent whole.

Disclosure Statement

Alfred Hero, Soummya Kar, Jose Moura, Joshua Neil, H. Vincent Poor, Melissa Turcotte, and Bowei Xi have no financial or non-financial disclosures to share for this article.

References

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, *29*(3), 626–688. https://doi.org/10.1007/s10618-014-0365-y

Anderson, H. S., Kharkar, A., Filar, B., & Roth, P. (2017). Evading machine learning malware detection. *Black Hat Conference*, 1–6. https://www.blackhat.com/docs/us-17/thursday/us-17-Anderson-Bot-Vs-Bot-Evading-Machine-Learning-Malware-Detection-wp.pdf

Athalye, A., Carlini, N., & Wagner, D. (2018). Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In J. Dy & A. Krause (Eds.), *Proceedings of the 35th international conference on machine learning* (pp. 274–283). http://proceedings.mlr.press/v80/athalye18a/athalye18a.pdf

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, N., Cynthia L and Serova, Davis, S., & Soboleva, A. (2012). Ncbi geo: Archive for functional genomics data sets—update. *Nucleic Acids Research*, *41*(D1), D991–D995. https://doi.org/10.1093/nar/gks1193

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer. https://doi.org/10.1007/978-1-4757-4286-2

Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., Giacinto, G., & Roli, F. (2013). Evasion attacks against machine learning at test time. In H. Blockeel, K. Kersting, S. Nijssen, & F. Železný (Eds.), *Lecture notes in computer science: Vol. 8190. Machine learning and knowledge discovery in databases.* (pp. 387–402). Springer. https://doi.org/10.1007/978-3-642-40994-3_25

Biggio, B., Nelson, B., & Laskov, P. (2012). Poisoning attacks against support vector machines. In J. Langford & J. Pineau (Eds.), *Proceedings of the 29th International Conference on Machine Learning* (pp. 1467–1474). http://icml.cc/2012/papers/880.pdf

Braverman, V., Hassidim, A., Matias, Y., Schain, M., Silwal, S., & Zhou, S. (2021). Adversarial robustness of streaming algorithms through importance sampling. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, & J. Wortman Vaughan *(Eds.)*, *Advances in Neural Information Processing Systems 34* (NeurIPS 2021) (pp. 3544–3557). Neural Information Processing Systems Foundation. https://papers.nips.cc/paper/2021/hash/1d01bd2e16f57892f0954902899f0692-Abstract.html

Brückner, M., & Scheffer, T. (2009). Nash equilibria of static prediction games. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (NIPS 2009) (pp. 171–179). Neural Information Processing Systems Foundation. https://papers.nips.cc/paper/2009/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Brückner, M., & Scheffer, T. (2011). Stackelberg games for adversarial prediction problems. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 547–555). Association for Computing Machinery. https://doi.org/10.1145/2020408.2020495

Canadian Institute for Cybersecurity. (2019). *Cybersecurity datasets*. University of New Brunswick. http://www.unb.ca/cic/research/datasets/index.html

Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. In 2017 *IEEE Symposium on Security and Privacy* (pp. 39–57). IEEE. https://doi.org/10.1109/SP.2017.49

Chen, P.-Y., Zhang, H., Sharma, Y., Yi, J., & Hsieh, C.-J. (2017). Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security* (pp. 15–26). Association for Computing Machinery. https://doi.org/10.1145/3128572.3140448

Chen, Y., Kar, S., & Moura, J. M. F. (2018a). The internet of things: Secure distributed inference. *IEEE Signal Processing Magazine*, *35*(5), 64–75. https://doi.org/10.1109/MSP.2018.2842097

Chen, Y., Kar, S., & Moura, J. M. F. (2018b). Resilient distributed estimation through adversary detection. *IEEE Transactions on Signal Processing*, *66*(9), 2455–2469. https://doi.org/10.1109/TSP.2018.2813330

Chiang, M., & Zhang, T. (2016). Fog and IoT: An overview of research opportunities. *IEEE Internet of Things Journal*, *3*(6), 854–864. https://doi.org/10.1109/JIOT.2016.2584538

Collins, M. (2017). *Network security through data analysis: From data to action (2nd ed.)*. O'Reilly Media.

Cuff, P., & Yu, L. (2016). Differential privacy as a mutual information constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 43–54). Association for Computing Machinery. https://doi.org/10.1145/2976749.2978308

Cyber Systems and Technology Group. (2019). DARPA intrusion detection datasets [https://www.ll.mit.edu/r-d/datasets].

Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S. M., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: Review, approaches and open research problems. *Heliyon*, *5*(6), Article e01802. https://doi.org/10.1016/j.heliyon.2019.e01802

Dalvi, N., Domingos, P., Mausam, Sanghai, S., & Verma, D. (2004). Adversarial classification. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 99–108). Association for Computing Machinery. https://doi.org/10.1145/1014052.1014066

Das, S., Wong, W.-K., Fern, A., Dietterich, T. G., & Siddiqui, M. A. (2017). Incorporating feedback into tree-based anomaly detection. In *Proceedings of KDD 2017 Workshop on Interactive Data Exploration and Analytics (IDEA'17)*. Association for Computing Machinery.

Davidson, D., Wu, H., Jellinek, R., Ristenpart, T., & Singh, V. (2016). Controlling UAVs with sensor input spoofing attacks. In *Proceedings of the 10th USENIX Conference on Offensive Technologies* (pp. 221–231).

Dekel, O., Shamir, O., & Xiao, L. (2010). Learning to classify with missing and corrupted features. *Machine Learning*, *81(2)*, 149–178. https://doi.org/10.1007/s10994-009-5124-8

Do, C. T., Tran, N. H., Hong, C., Kamhoua, C. A., Kwiat, K. A., Blasch, E., Ren, S., Pissinou, N., & Iyengar, S. S. (2017). Game theory for cyber security and privacy. *ACM Computing Surveys (CSUR)*, *50*(2), Article 30. https://doi.org/10.1145/3057268

Dolev, D., Lynch, N. A., Pinter, S. S., Stark, E. W., & Weihl, W. E. (1986). Reaching approximate agreement in the presence of faults. *Journal of the ACM*, *33*(3), 499–516. https://doi.org/10.1145/5925.5931

Dong, L., Han, Z., Petropulu, A. P., & Poor, H. V. (2010). Improving wireless physical layer security via cooperating relays. *IEEE Transactions on Signal Processing*, *58*(3), 1875–1888. https://doi.org/10.1109/TSP.2009.2038412

Dua, S., & Du, X. (2011). *Data mining and machine learning in cybersecurity*. Auerbach Publications. https://doi.org/10.1201/b10867

Dwork, C. (2008). Differential privacy: A survey of results. In M. Agrawal, D. Du, Z. Duan, & A. Li (Eds.), *Lecture notes in computer science: Vol. 4978. Theory and Applications of Models of Computation* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-540-79228-4_1

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3–4), 211–407. https://doi.org/10.1561/0400000042

Dwork, C., & Smith, A. (2009). Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality*, *1*(2), 135–154. https://doi.org/10.29012/jpc.v1i2.570

El Ghaoui, L., Lanckriet, G. R. G., & Natsoulis, G. (2003). *Robust classification with interval data* (tech. rep. UCB/CSD-03-1279). EECS Department, University of California, Berkeley. http://www.eecs.berkeley.edu/Pubs/TechRpts/2003/5772.html

Fagen, M., Maronn, J., Brady, K., Cuthill, B., Megas, K., Herold, R., Lamire, D., & Hoen, B. (2021). IoT device cybersecurity guidance for the federal government. *National Institute of Standards and Technology (NIST) Special Publication 800-213*. U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800-213

Farina, P., Cambiaso, E., Papaleo, G., & Aiello, M. (2016). Are mobile botnets a possible threat? The case of SlowBot Net. *Computers & Security*, *58*, 268–283. https://doi.org/10.1016/j.cose.2016.02.005

Fielder, A., Panaousis, E., Malacaria, P., Hankin, C., & Smeraldi, F. (2014). Game theory meets information security management. In N. Cuppens-Boulahia, F. Cuppens, S. Jajodia, A. Abou El Kalam, & T. Sans (Eds.), IFIP Advances in Information and Communication Technology: Vol. 428. *ICT Systems Security and Privacy Protection (pp.* 15–29). Springer. https://doi.org/10.1007/978-3-642-55415-5_2

Fielder, A., Panaousis, E., Malacaria, P., Hankin, C., & Smeraldi, F. (2016). Decision support approaches for cyber security investment. *Decision Support Systems*, *86*, 13–23. https://doi.org/10.1016/j.dss.2016.02.012

FireEye. (2018). *Get One Step Ahead of Email Threats*, *Email Threat Report for January-June 2018* (tech. rep.) https://www.fireeye.com/company/press- releases/2018/new-fireeye-email-threat-report-underlines-the-rise-in-malware-l.html

Fodor, L., Jakovetić, D., Krejić, N., Jerinkić, N. K., & Škrbić, S. (2021). Performance evaluation and analysis of distributed multi-agent optimization algorithms with sparsified directed communication. *EURASIP Journal on Advances in Signal Processing*, *2021*(1), 1–29. https://doi.org/10.1186/s13634-021-00736-4

Franchetti, F., Low, T. M., Mitsch, S., Mendoza, J. P., Gui, L., Phaosawasdi, A., Padua, D., Kar, S., Moura, J. M., Franusich, M., Johnson, J., Platzer, A., & Veloso, M. M. (2017). High-assurance spiral: End-to-end guarantees for robot and car control. *IEEE Control Systems Magazine*, *37*(2), 82–103. https://doi.org/10.1109/MCS.2016.2643244

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: Methods and prospects. *Big Data Analytics*, *1*(1), Article 9. https://doi.org/10.1186/s41044-016-0014-0

Giaconi, G., Gündüz, D., & Poor, H. V. (2020). Smart meter privacy. In A. Tajer, S. M. Perlaza, & H. V. Poor (Eds.), *Advanced data analytics for power systems (pp. 230–260)*. Cambridge University Press. https://doi.org/10.1017/9781108859806.014

Glasser, J., & Lindauer, B. (2013). Bridging the gap: A pragmatic approach to generating insider threat data. In *2013 IEEE Security and Privacy Workshops* (pp. 98–104). IEEE. https://doi.org/10.1109/SPW.2013.37

Globerson, A., & Roweis, S. (2006). Nightmare at test time: Robust learning by feature deletion. In W. Cohen & A. Moore (Eds.), *Proceedings of the 23rd International Conference on Machine Learning* (pp. 353–360). https://doi.org/10.1145/1143844.1143889

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015, May 9). *Explaining and harnessing adversarial examples*. Third *International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, United States. https://doi.org/10.48550/arXiv.1412.6572

Harris, M. (2015, September 4). Researcher hacks self-driving car sensors. *IEEE Spectrum*. https://spectrum.ieee.org/researcher-hacks-selfdriving-car-sensors

Hassan, W. U., Bates, A., & Marino, D. (2020). Tactical provenance analysis for endpoint detection and response systems. In *2020 IEEE Symposium on Security and Privacy* (pp. 1172–1189). IEEE. https://doi.org/10.1109/SP40000.2020.00096

Hayes, M. A., & Capretz, M. A. (2015). Contextual anomaly detection framework for big sensor data. *Journal of Big Data*, *2*(1), Article 2. https://doi.org/10.1186/s40537-014-0011-y

Heard, N., Rubin-Delanchy, P., & Lawson, D. J. (2014). Filtering automated polling traffic in computer network flow data. In *2014 IEEE Joint Intelligence and Security Informatics Conference* (pp. 268–271). IEEE. https://doi.org/10.1109/JISIC.2014.52

Hero, A. O. (2003). Secure space-time communication. *IEEE Transactions on Information Theory*, *49*(12), 3235–3249. https://doi.org/10.1109/TIT.2003.820010

Hou, E., Sricharan, K., & Hero, A. O. (2018). Latent Laplacian maximum entropy discrimination for detection of high-utility anomalies. *IEEE Transactions on Information Forensics and Security*, *13*(6), 1446–1459. https://doi.org/10.1109/TIFS.2018.2790580

IoT Act of US Congress. (2020). Internet of things cybersecurity improvement act of 2020. *Congressional Record, PUBLIC LAW 116–207—DEC. 4, 2020*. https://www.congress.gov/116/plaws/publ207/PLAW-116publ207.pdf

Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, *80*(5), 973–993. https://doi.org/10.1016/j.jcss.2014.02.005

Jordan, M. I., Lee, J. D., & Yang, Y. (2018). Communication-efficient distributed statistical inference. *Journal of the American Statistical Association*, *114*(526), 668–681. https://doi.org/10.1080/01621459.2018.1429274

Kailkhura, B., Han, Y. S., Brahma, S., & Varshney, P. K. (2015). Distributed Bayesian detection in the presence of Byzantine data. *IEEE Transactions on Signal Processing*, *63*(19), 5250– 5263. https://doi.org/10.1109/TSP.2015.2450191

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B. . . . Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, *14*(1–2), 1–210. https://doi.org/10.1561/2200000083

Kantarcıoğlu, M., Xi, B., & Clifton, C. (2011). Classifier evaluation and attribute selection against active adversaries. *Data Mining and Knowledge Discovery*, *22*(1–2), 291–335. https://doi.org/10.1007/s10618-010-0197-3

Kar, S., & Moura, J. M. F. (2011). Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs. *IEEE Journal of Selected Topics in Signal Processing*, *5*(4), 674–690. https://doi.org/10.1109/JSTSP.2011.2127446

Kar, S., & Moura, J. M. F. (2013). Consensus + innovations distributed inference over networks. *IEEE Signal Processing Magazine*, *30*(3), 99–109. https://doi.org/10.1109/MSP.2012.2235193

Kar, S., Moutis, P., & Whitacre, J. (2020). *Agent-based coordination scheme for pv integration (abc4pv)* (tech. rep.). Carnegie Mellon Univ., Pittsburgh, PA (United States).

Kent, A. D. (2014). *User-computer authentication associations in time*. Los Alamos National Laboratory. https://doi.org/10.11578/1160076

Kent, A. D. (2015). Cybersecurity Data Sources for Dynamic Network Research. *Dynamic Networks in Cybersecurity*. https://csr.lanl.gov/data/cyber1/

Killick, R., & Eckley, I. A. (2014). Changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, *58(3)*, 1–19. https://doi.org/10.18637/jss.v058.i03

Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, *4*(3), 382–401. https://doi.org/10.1145/357172.357176

Lanckriet, G. R. G., Ghaoui, L. E., Bhattacharyya, C., & I., J. M. (2002). A robust minimax approach to classification. *Journal of Machine Learning Research*, *3*, 555–582. https://doi.org/10.1162/153244303321897726

Lavielle, M., & Teyssiere, G. (2006). Detection of multiple change-points in multivariate time series. *Lithuanian Mathematical Journal*, *46*(3), 287–306. https://doi.org/10.1007/s10986-006-0028-9

Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., & Srivastava, J. (2003). A comparative study of anomaly detection schemes in network intrusion detection. In D. Barbara & C. Kamath *(Eds.)*, *Proceedings of the 2003 SIAM International Conference on Data Mining* (pp. 25–36). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611972733.3

LeBlanc, H. J., & Hassan, F. (2014). Resilient distributed parameter estimation in heterogeneous time-varying networks. In *Proceedings of 3rd International Conference on High Confidence Networked Systems (HiCoNS)* (pp. 19–28). Association for Computing Machinery. https://doi.org/10.1145/2566468.2566476

LeBlanc, H. J., Zhang, H., Koustsoukos, X., & Sundaram, S. (2015). Resilient asymptotic consensus in robust networks. *IEEE Journal of Selected Areas in Communications*, *31*(4), 766–781. https://doi.org/10.1109/JSAC.2013.130413

Li, A. Z., & Barton, D. (2019, November 14). A brief history of machine learning in cybersecurity. *Security Info Watch*. https://www.securityinfowatch.com/cybersecurity/article/21114214/a-brief-history-of-machine-learning-in-cybersecurity.

Li, F., Lai, L., & Cui, S. (2020). On the adversarial robustness of subspace learning. *IEEE Transactions on Signal Processing*, *68*, 1470–1483. https://doi.org/10.1109/TSP.2020.2974676

Lian, X., Zhang, C., Zhang, H., Hsieh, C.-J., Zhang, W., & Liu, J. (2017). Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett *(Eds.), Advances in Neural Information Processing Systems 30 (NIPS 2017) (pp. 5336–5346)*. Neural Information Processing Systems Foundation. https://papers.nips.cc/paper/2017/hash/f75526659f31040afeb61cb7133e4e6d-Abstract.html

Liang, G., Zhao, J., Luo, F., Weller, S. R., & Dong, Z. Y. (2017). A review of false data injection attacks against modern power systems. *IEEE Transactions on Smart Grid*, *8*(4), 1630– 1638. https://doi.org/10.1109/TSG.2015.2495133

Liu, W., & Chawla, S. (2010). Mining adversarial patterns via regularized loss minimization. *Machine Learning*, *81*(1), 69–83. https://doi.org/10.1007/s10994-010-5199-2

Lu, N., Cheng, N., Zhang, N., Shen, X., & Mark., J. W. (2014). Connected vehicles: Solutions and challenges. *IEEE Internet of Things Journal*, *1*(4), 289–299. https://doi.org/10.1109/JIOT.2014.2327587

Ma, C., Li, J., Ding, M., Yang, H. H., Shu, F., Quek, T. Q. S., & Poor, H. V. (2020). On safeguarding privacy and security in the framework of federated learning. *IEEE Network*, *24*(4), 242– 248. https://doi.org/10.1109/MNET.001.1900506

Ma, J. (2019, July 19). *Detecting malicious URLs*. http://www.sysnet.ucsd.edu/projects/url/.

Mahmood, T., & Afzal, U. (2013). Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools. In *2013 2nd National Conference on Information Assurance (NCIA)* (pp. 129–134). IEEE. https://doi.org/10.1109/NCIA.2013.6725337

Makhdoumi, A., Salamatian, S., Fawaz, N., & Médard, M. (2014). From the information bottleneck to the privacy funnel. In *Proceedings of the 2014 IEEE Information Theory Workshop* (pp. 501– 505). IEEE. https://doi.org/10.1109/ITW.2014.6970882

Malware_traffic. (2019). Malware and exploit kit traffic. *Malware-Traffic-Analysis.net*.http://malware-traffic-analysis.net/.

Mansfield-Devine, S. (2015). The growth and evolution of DDoS. *Network Security*, *2015*(10), 13–20. https://doi.org/10.1016/S1353-4858(15)30092-1

McKinsey & Company. (2019). *Perspectives on transforming cybersecurity*. https://www.mckinsey.com/~/media/McKinsey/McKinsey%20Solutions/Cyber%20Solutions/Perspectives%20on%20transforming%20cybersecurity/Transforming%20cybersecurity_March2019.ashx.

Milajerdi, S. M., Eshete, B., Gjomemo, R., & Venkatakrishnan, V. (2019). POIROT: Aligning attack behavior with kernel audit records for cyber threat hunting. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1795–1812). Association for Computing Machinery. https://doi.org/10.1145/3319535.3363217

Miller, R. G. (1981). *Simultaneous statistical inference*. Springer. https://doi.org/10.1007/978-1-4613-8122-8

Nachenberg, C. (1997). Computer virus-antivirus coevolution. *Communications of the Association for Computing Machinery*, *40*(1), 46–51. https://doi.org/10.1145/242857.242869

Nedic, A. (2020). Distributed gradient methods for convex machine learning problems in networks: Distributed optimization. *IEEE Signal Processing Magazine*, *37*(3), 92–101. https://doi.org/10.1109/MSP.2020.2975210

Nedich, A. (2015). Convergence rate of distributed averaging dynamics and optimization in networks. *Foundations and Trends® in Systems and Control*, *2*(1), 1–100. https://doi.org/10.1561/2600000004

Neil, J., Uphoff, B., Hash, C., & Storlie, C. (2013). Towards improved detection of attackers in computer networks: New edges, fast updating, and host agents. In *2013 6th International Symposium on Resilient Control Systems (ISRCS)* (pp. 218–224). IEEE. https://doi.org/10.1109/ISRCS.2013.6623779

Nelson, B., Barreno, M., Chi, F. J., Joseph, A. D., Rubinstein, B. I., Saini, U., Sutton, C. A., Tygar, J. D., & Xia, K. (2008). Exploiting machine learning to subvert your spam filter. In F. Monrose *(Ed.)*, *Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats* (Article 7). USENIX Association. http://www.usenix.org/event/leet08/tech/full_papers/nelson/nelson.pdf

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., & Poor, H. V. (2021). Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(3), 1622–1658. https://doi.org/10.1109/COMST.2021.3075439

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia Conference on Computer and*

Communications Security (pp. 506–519). Association for Computing Machinery. https://doi.org/10.1145/3052973.3053009

Pasqualetti, F., Bicchi, A., & Bullo, F. (2012). Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, *57*(1), 90–104. https://doi.org/10.1109/TAC.2011.2158130

Pasqualetti, F., Dörfler, F., & Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, *58*(11), 2715–2729. https://doi.org/10.1109/TAC.2013.2266831

Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, *51*(12), 3448–3470. https://doi.org/10.1016/j.comnet.2007.02.001

Pawlick, J., Colbert, E., & Zhu, Q. (2019). A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. *ACM Computing Surveys (CSUR)*, *52*(4), Article 82. https://doi.org/10.1145/3337772

Pillutla, K., Kakade, S. M., & Harchaoui, Z. (2022). Robust aggregation for federated learning. *IEEE Transactions on Signal Processing*, *70*, 1142–1154. https://doi.org/10.1109/TSP.2022.3153135

Poor, H. V. (2018). Privacy in networks of interacting agents. In R. Tempo, S. Yurkovich, & P. Misra (Ed.), *Emerging applications of control and system theory* (pp. 259–268). Springer. https://doi.org/10.1007/978-3-319-67068-3_19

Poor, H. V., & Schaefer, R. F. (2017). Wireless physical layer security. *Proceedings of the National Academy of Sciences of the USA*, *114*(1), 19–26. https://doi.org/10.1073/pnas.1618130114

Price-Williams, M., Heard, N., & Turcotte, M. (2017). Detecting periodic subsequences in cyber security data. In J. Brynielsson *(Ed.), 2017 European Intelligence and Security Informatics Conference (EISIC)* (pp. 84–90). IEEE. https://doi.org/10.1109/EISIC.2017.40

Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C., & Samatova, N. F. (2015). Anomaly detection in dynamic networks: A survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, *7*(3), 223–247. https://doi.org/10.1002/wics.1347

Samonas, S., & Coss, D. (2014). The CIA strikes back: Redefining confidentiality, integrity and availability in security. *Journal of Information System Security*, *10*(3), 21–45.

Sanger, D. E. (2018). *War, sabotage, and fear in the cyber age*. Random House.

Sarker, I. H., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. *Journal of Big data*, *7*(1), Article 41.

https://doi.org/10.1186/s40537-020-00318-5

Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell System Technical Journal*, *28*(4), 656–715. https://doi.org/10.1002/j.1538-7305.1949.tb00928.x

Sharif, M., Bhagavatula, S., Bauer, L., & Reiter, M. K. (2016). Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1528–1540). Association for Computing Machinery. https://doi.org/10.1145/2976749.2978392

So, J., Güler, B., & Avestimehr, A. S. (2020). Byzantine-resilient secure federated learning. *IEEE Journal on Selected Areas in Communications*, *39*(7), 2168–2181. https://doi.org/10.1109/JSAC.2020.3041404

Sundaram, S., & Hadjicostis, C. N. (2011). Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Transactions on Automatic Control*, *56*(7), 1495–1508. https://doi.org/10.1109/TAC.2010.2088690

Taddeo, M., & Floridi, L. (2018). Regulate artificial intelligence to avert cyber arms race. *Nature*, *556(7701)*, 296–298. https://doi.org/10.1038/d41586-018-04602-6

Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, *1*(12), 557–560. https://doi.org/10.1038/s42256-019-0109-1

Teo, C. H., Globerson, A., Roweis, S. T., & Smola, A. J. (2007). Convex learning with invariances. In J. Platt, D. Koller, Y. Singer, & S. Roweis *(Eds.), Advances in Neural Information Processing Systems 20 (NIPS 2007) (pp. 1489–1496).* Neural Information Processing Systems Foundation. https://papers.nips.cc/paper/2007/hash/20b5e1cf8694af7a3c1ba4a87f073021-Abstract.html

TrellisWare Technologies Inc. (2017, October 29). *First responders*. https://www.trellisware.com/first-responders/.

Turcotte, M., Kent, A., & Hash, C. (2018). Unified host and network data set. In P. R.-D. Nick Heard Niall Adams & M. Turcotte (Eds.), *Data science for cyber-security* (pp. 1–22). World Scientific. https://doi.org/10.1142/9781786345646_001

Uhlemann, E. (2018). Time for autonomous vehicles to connect [connected vehicles]. *IEEE Vehicular Technology Magazine*, *13*(3), 10–13. https://doi.org/10.1109/MVT.2018.2848342

Upton, E., & Halfacree, G. (2014). *Raspberry pi user guide (3rd ed.)*. John Wiley & Sons.

US Department of Homeland Security. (2019). *Information marketplace for policy and analysis of cyber-risk and trust (IMPACT)*. Science and Technology Directorate. [https://www.dhs.gov/csd-impact].

Vempaty, A., Tong, L., & Varshney, P. K. (2013). Distributed inference with Byzantine data. *IEEE Signal Processing Magazine*, *30*(5), 65–75. https://doi.org/10.1109/MSP.2013.2262116

Veracode. (2022). *Man in the Middle (MITM) attack*. https://www.veracode.com/security/man-middle-attack

Verizon. (2022). *2022 Data breach investigations report* (tech. rep.). Verizon Enterprises. https://www.verizon.com/business/resources/reports/2022/dbir/2022-data-breach-investigations-report-dbir.pdf

Wang, C., Zhang, D., Huang, S., Li, X., & Ding, L. (2021). Crafting adversarial email content against machine learning based spam email detection. In *Proceedings of the 2021 International Symposium on Advanced Security on Software and Systems* (pp. 23–28). Association for Computing Machinery. https://doi.org/10.1145/3457340.3458302

Wang, W., Ying, L., & Zhang, J. (2016). On the relation between identifiability, differential privacy, and mutual-information privacy. *IEEE Transactions on Information Theory*, *62*(9), 5018– 5029. https://doi.org/10.1109/TIT.2016.2584610

Wyner, A. D. (1975). The wire-tap channel. *Bell System Technical Journal*, *54*(8), 1355–1387. https://doi.org/10.1002/j.1538-7305.1975.tb02040.x

Xi, B. (2020). Adversarial machine learning for cybersecurity and computer vision: Current developments and challenges. *Wiley Interdisciplinary Reviews: Computational Statistics*, *12*(5), Article e1511. https://doi.org/10.1002/wics.1511

Yagli, S., Dytso, A., & Poor, H. V. (2020). Information-theoretic bounds on the generalization error and privacy leakage in federated learning. In *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE.* https://doi.org/10.1109/SPAWC48557.2020.9154277

Yang, W., Schaefer, R. F., & Poor, H. V. (2019). Wiretap channels: Nonasymptotic fundamental limits. *IEEE Transactions on Information Theory*, *65*(7), 4069–4093. https://doi.org/10.1109/TIT.2019.2904500

Yang, Z., Gang, A., & Bajwa, W. U. (2020). Adversary-resilient distributed and decentralized statistical inference and machine learning: An overview of recent advances under the byzantine threat model. *IEEE Signal Processing Magazine*, *37*(3), 146–159. https://doi.org/10.1109/MSP.2020.2973345

Yasin, R. (2015). *Stealing data by 'living off the land'*. Dark Reading. https://www.darkreading.com/analytics/stealing-data-by-living-off-the-land/d/d-id/1322063

Zhang, F., Kodituwakku, H. A. D. E., Hines, J. W., & Coble, J. (2019). Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data. *IEEE Transactions on Industrial Informatics*, *15*(7), 4362–4369. https://doi.org/10.1109/TII.2019.2891261

Zhang, H., & Kim, J. (2021). *Towards a federated learning framework for heterogeneous devices of internet of things*. ArXiv. https://doi.org/10.48550/arXiv.2105.14675.

Zhang, J., Blum, R., Kaplan, L. M., & Lu, X. (2017). Functional forms of optimum spoofing attacks for vector parameter estimation in quantized sensor networks. *IEEE Transactions on Signal Processing*, *65*(3), 705–720. https://doi.org/10.1109/TSP.2016.2626258

Zhang, J., Blum, R., Lu, X., & Conus, D. (2015). Asymptotically optimum distributed estimation in the presence of attacks. *IEEE Transactions on Signal Processing*, *63*(5), 1086–1101. https://doi.org/10.1109/TSP.2014.2386281

Zhao, J., Shetty, S., Pan, J. W., Kamhoua, C., & Kwiat, K. (2019). Transfer learning for detecting unknown network attacks. E*URASIP Journal on Information Security*, *2019*(1), Article 1. https://doi.org/10.1186/s13635-019-0084-4

Zhou, X., Liang, W., Shimizu, S., Ma, J., & Jin, Q. (2020). Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems. *IEEE Transactions on Industrial Informatics*, *17*(8), 5790–5798. https://doi.org/10.1109/TII.2020.3047675

Zhou, Y., Kantarcioglu, M., Thuraisingham, B., & Xi, B. (2012). Adversarial support vector machine learning. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1059–1067). Association for Computing Machinery. https://doi.org/10.1145/2339530.2339697
