Building a Unified Data Falsification Threat Landscape for Internet of Things/Cyberphysical Systems Applications

Shameek Bhattacharjee, Western Michigan University
Sajal K. Das, Missouri University of Science and Technology

We lay out a blueprint of a complete and parameterized threat landscape for data falsification/false data injection attacks on telemetry data collected from Internet of Things/cyberphysical systems applications under zero-trust assumptions, helping to enable better validation of anomaly-based attack detection methods.

Enormous amounts of telemetry data are being collected by smart devices, the so-called Internet of Things (IoT), which are the building blocks of cyberphysical systems (CPSs). Such wide area telemetry data drive decision making and operations in smart living IoT applications (for instance, smart energy, smart transportation, and so on) that improve civic well-being. Dependence on data analytics, the immediate civilian impact of wrong decisions, economic motivations, and vast attack surfaces (due to the community-scale IoT and CPS domains) make an extremely attractive target for data integrity attacks. In parallel, the past decade has seen unprecedented advanced persistent threats (for example, Stuxnet[22] and the Ukraine power grid attack) that can change various telemetry data, alter monitoring processes,

Digital Object Identifier 10.1109/MC.2022.3198599
Date of current version: 8 March 2023

when an adversary gains access privileges through a creative zero-day cyber, physical, or social engineering exploit.

Traditional cybersecurity is designed on the premise that everything within an enterprise/utility network is trustworthy as long as the perimeter is not breached. Therefore, the main focus is on protecting against outside threats, using cryptography, network traffic analysis, network segmentation, and fine-grained user access control. However, many agencies, such as Palo Alto Networks, the U.S. Department of Defense, and the National Institute of Standards and Technology, have formalized the need for a zero-trust architecture,[7] which recognizes that (static) trust in users and endpoint devices is a vulnerability. Once inside the network, adversaries and malicious insiders are free to move laterally and access and modify any data after gaining appropriate privileges. This creates data falsification attacks apart from the CP couplings that create data falsification attacks.

CONTRIBUTIONS

In this article, we unify a detailed threat landscape for data integrity attacks on telemetry data that target the operational accuracy of the smart living IoT. We first reveal why traditional cybersecurity practices are not enough and CP factors that make data integrity attacks a credible threat. Then, we propose four facets that characterize the data integrity threat landscape of IoT/CPS telemetry data. The four facets include 1) attack types, 2) the attack strength and aperture, 3) the attack scale, and 4) attack strategies. We provide a detailed exposition of different threat modeling aspects, attack emulation techniques, and a way to mathematically parameterize these facets such that all kinds of adversarial capabilities are implicitly accounted for. This, in turn, allows an unbiased evaluation of a defense framework in which the limits of the defense model can be tested. This is mainly because a defender never knows what kind of attack will be launched.

Our effort in the threat model specification is to discuss some pitfall assumptions that create asymmetry between reality and perception and lead to incomplete threat assessment and biased security performance evaluation. We offer a recipe to create unbiased attack simulations for other researchers working in industrial and smart living IoT applications, where the use of telemetry data is common. Furthermore, we lay out our threat model in a generic way but with some examples to help researchers get a common recipe to tailor our threat model for their needs. In the absence of labeled attack datasets, our threat modeling approach can be used to create a superset of many possible attack realizations. Even if labeled datasets of specific attacks are present, evaluating a defense framework is based on a specific instance of an attack. In contrast, our approach helps researchers generate a parameterized threat state-space universe, where the actual attack is an instance within this threat state space.

UNIFIED ABSTRACTION OF IoT/CPS

Regardless of the type of smart living IoT/CPS domain (such as the smart grid and smart transportation), a unified abstraction of the architecture and operations of smart living IoT domain is possible as shown in Figure 1. Under the umbrella of such a unified view, an effective unified characterization of a threat landscape of telemetry data in the sensing loop is enabled that applies across various IoT/CPS domains.

Smart grid

A smart grid is a large domain consisting of multiple functional units, such as advanced metering infrastructure (AMI)[18] and phasor measurement unit (PMU) infrastructure.[3] Each functional unit typically has a controller that runs specific services (for example, demand response and automated billing by AMI and voltage sag detection by PMUs) based on telemetry data collected from IoT endpoint devices (such as smart meters and PMUs). For simplicity, we call the field-area IoT endpoints as *IoT devices*.

In the AMI, smart meters collect energy consumption and generation data from smart home appliances (customer loads) and renewable energy sources. In the hierarchical architecture shown in Figure 1, multiple smart meters connect to a neighborhood area network (NAN) device that forwards all the meters' data to a field area network (FAN) gateway (fog node) for local area analytics. Multiple FAN gateways connect to a cloud controller hosting computations for wide area analytics (for example, billing and load profiling) and decisions (signaling the remote disconnect of an appliance), which is communicated to the demand response switches (actuators) at the customer site. The AMI also applies to water distribution monitoring in a similar way.[10]

Similarly, the distribution/transmission layers in smart grids have multiple PMUs acting as IoT devices that collect sample voltages, current amplitudes, and angles between the voltage and current in each of the three phases. Multiple PMUs transmit such telemetry data to a phasor data concentrator (PDC). A PDC forwards data from multiple PMUs to a local link controller (LLC). Multiple LLCs then connect to a wide area controller determining

events, the state of the grid, and control decisions (for instance, islanding and load balancing).

Smart transportation

In this domain, vehicle-to-infrastructure (V2I)[8] and vehicle-to-vehicle[17] units run telemetry data-driven services to control traffic congestion, vehicular rerouting, platooning, and incident response. Future V2I systems will collect vehicular data (such as speed, road segment, velocity, and direction) via dedicated short-range radios hosted on smart cars. In other implementations, transport measurement channel (TMC)[8] equipment automatically senses such data as vehicles pass along a road segment. Such data are forwarded to IoT devices known as *roadside units* (*RSUs*). These RSUs collect data from smart cars and TMCs and then forward the information to the respective fog and cloud servers for local and wide area analytics via the Internet. The fog/cloud server can issue control commands to vehicles and driving apps (for example, rerouting and speed recommendations) and traffic signals (signal switching information) for traffic management. The actuators are humans taking actions and signaling logic in traffic signals.

ANATOMY OF DATA INTEGRITY EXPLOITS IN IoT

We know that cryptography-based approaches to encryption, such as digital signatures, can offer protection against adversaries accessing and modifying data, thus reducing the chances of a data falsification attack. Naturally, a question arises: Why does the research community still need to worry about data integrity attacks in IoT telemetry data? In this section, we provide practical reasons why data falsification attacks are a credible threat in IoT/CPS domains.

Lack of cryptoagility in IoT

The secrecy of the Rivest–Shamir–Adleman (RSA) algorithm (a commonly used public-key cryptography method) depends on the inability of an adversary to factorize the two randomly chosen prime numbers used to derive RSA algorithms' public keys. Hence, if the prime factors are discovered, the adversary can rederive an RSA's private key. Following this, the adversary can decrypt and modify data from devices. To prevent this, two key requirements need to be fulfilled: high randomness in prime numbers and enough computing power to transform input data into a strong key. Random number generators in digital systems rely on physical nondeterministic inputs/measurements that are sourced from device hardware (for example, mouse pointer movements, keystroke patterns, clock signals, phase noise, and so on). Computers/smartphones have hardware that allows the collection of nondeterministic inputs, which creates randomness.

In contrast, IoT devices lack sources of randomness, due to limitations in the attached hardware (for instance, the absence of keystroke patterns and mouse movements). Keys generated by lightweight IoT devices are therefore at risk of not being sufficiently random. This increases the chance that two keys share a factor, allowing the keys to be broken. Heninger et al.[15] found that most of the keys that were broken,

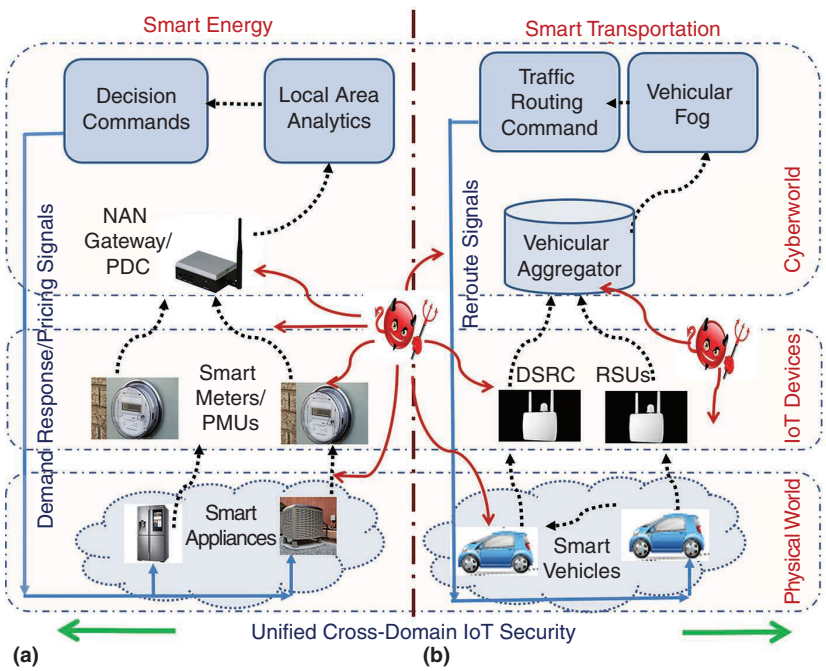

FIGURE 1. An example of the cross-domain layered IoT architecture in the (a) smart energy and (b) smart transportation CPS domains. NAN: neighborhood area network; PDC phase data concentrator; DSRC: dedicated short-range communication; RSU: roadside unit.

and Hastings et al.[13] reported that one in 172 devices' digital certificates were compromised. Preloading keys on an IoT device during manufacturing will open up devices to supply chain attacks, where an untrustworthy manufacturer or logistics company tampers with the keys en route.[12] Also, Cardenas et al.[16] noted that many industrial IoT/CPS systems cannot afford authentication and encryption altogether, due to hard real-time operational requirements.

Physical data manipulation via transduction attacks

IoT devices are vulnerable to transduction attacks[19] that disturb analog signals sensed by the devices as well as physical tampering with firmware[21] such that the accurate conversion of the analog-to-digital output of the telemetry data is altered. This results in false telemetry data reports from the devices. A transduction attack exploits a vulnerability in the physics of how sensor/hardware processes input analog signals, as surveyed in Yan et al.,[14] by directing some malicious electromagnetic signal as interference onto an IoT device. Sensors translate physical analog signals into electrical signals. Thereafter, software in the firmware (or a remote app) interprets and reads binary representations rather than direct physical and electrical quantities. An adversary can vary the intensity of the malicious signal on target IoT devices, changing the extent to which data are falsified. In Liu et al.,[17] a transduction attack on a self-driving car was shown. Security practices, such as static analysis, fuzz testing, and signed software updates, do not offer detection of a sensor delivering false data.[19] Similarly, cryptography/network intrusion detection is unable to detect and prevent such attacks.

Insider threats and social engineering exploits

A set of detailed case studies at Cybersecurity Insiders[11] finds 68% of cybersecurity breaches have some involvement of at least one or more utility/enterprise insiders in key positions of knowledge, even when the main threat actor is external to an organization. Furthermore, various social engineering exploits are used (such as phishing emails) on employees of an organization to extract privileges that bypass traditional cybersecurity and give control of data and devices to an adversary, causing data integrity attacks.

VARIOUS SOCIAL ENGINEERING EXPLOITS ARE USED (SUCH AS PHISHING EMAILS) ON EMPLOYEES OF AN ORGANIZATION TO EXTRACT PRIVILEGES THAT BYPASS TRADITIONAL CYBERSECURITY AND GIVE CONTROL OF DATA AND DEVICES TO AN ADVERSARY, CAUSING DATA INTEGRITY ATTACKS.

Heterogeneity in IoT market network interfaces

While some hardware security against transduction attacks exists, it requires all manufacturers to adhere to a common hardware security standard. Additionally, a single network contains devices assembled from parts from tens to hundreds of distinct manufacturers. Most importantly, the sheer community scale of IoT devices for smart living makes embedded and secure hardware an expensive solution, as reasoned in Bhattacharjee and Das.[2] Thus, the detection of data integrity attacks from physical exploits is going to be a challenge for the community-scale IoT and the hardware in IoT markets. Furthermore, IoT devices in smart cities are getting a FAN/edge interfacing layer using 5G, software radios, and shared spectrum technologies, which use highly programmable wired and wireless networking components.[9] The software-defined nature allows FAN devices to be compromised, enabling the implementation of advanced attack strategies. This is further elaborated in the "Attack Scale" and "Attack Strategies" sections.

Conclusion

The reasons for IoT devices being attractive targets are 1) easy physical access causes transduction/physical attacks, 2) the devices have limited hardware/memory capabilities, 3) large-scale deployment makes deploying in situ embedded security very expensive and impractical for utility providers, 4) the devices have cyberconnectivity to the Internet, and 5) data integrity attacks in smart living have both an immediate civilian and an immediate economic impact, depending on the attack type.

UNIFIED DATA INTEGRITY THREAT LANDSCAPE

In this section, we list a set of characteristics that specify a detailed and realistic threat model in smart living IoT telemetry data. The IoT domain can be a target for the following categories of organized adversaries: 1) rival nations, 2) insiders, 3) cybercriminals, 4) rogue/selfish customers, and 5) business competitors. Depending on an adversary's capability, the actual parameters will vary, but there is no way of predicting what they might be. Instead of snapshot attacks, our contribution provides a recipe to generate a superset of threat landscapes that, in turn, allows unbiased evaluation of defense methods.

> MANY RESEARCH WORKS DO NOT PARAMETERIZE THE ATTACK SCALE AS A VARIABLE IN THE THREAT MODEL, OR THEY ASSUME ONLY LOWER ATTACK SCALES TO BE "REALISTIC."

Attack scale

The attack scale refers to the number of compromised IoT telemetry devices and telemetry data streams. Redundancy is often present in the IoT telemetry endpoint layer to achieve wide area monitoring. Most works assume a fixed fraction of compromised IoT devices constrained by an attack budget. Also, many research works do not parameterize the attack scale as a variable in the threat model, or they assume only lower attack scales to be "realistic." These factors result in the following pitfalls.

First, the effective fraction of compromised devices depends on the size of the network and is unrelated to an adversary's attack budget. Hence, smaller-sized networks will have a large fraction of compromised IoT devices, even with a seemingly smaller budget. Examples are smart meters in microgrids and TMC sensors for decentralized traffic monitoring. Second, the effective scale of an attack increases if an attacker compromises the intermediate data aggregator that is present in most wide area IoT/CPS applications (for example, NAN gateways,[18] PDCs,[3] and RSUs[8]) since the aggregator can manipulate data streams from multiple IoT devices at once. This is a realistic possibility since most edge aggregator devices in smart cities are planned to be wireless and programmable Universal Software Radio Peripheral radios, which contain a USB port and are is easily physically accessible and programmable over the air.[9]

Third, the cheapness of the exploit should be taken into account. If the exploit is cheap, a high number of compromised IoT devices is possible, even with a small attack budget. As an example, in 2009–2010, an attack on Puerto Rico's smart grid[21] metering utility (the Puerto Rico Electric Power Authority) was carried out by utility insiders and maintenance personnel, who tampered with thousands of smart meters to launch data falsification by using a portable optical laser probe tool kit that cost just US$400. For proper scientific treatment, it is necessary that the research community parameterizes the attack scale, from very small to large values, and then tests breakdown points to validate proposed defense models.

Attack strength and aperture

The attack strength denotes the average margin of falsified data from each compromised IoT device. Our preliminary research on smart meters revealed that many works rarely identify compromised meters that launch margins of false data below 450 W. Similarly, we observed, in Roy et al.,[4] that existing works on PMU data integrity attacks rarely parameterize the average margin of falsification of current magnitude values.

In our recent work Bhattacharjee et al.,[1] we showed that since the standard deviation of the data is high in smart living IoT/CPS systems, due to human and environmental randomness, the average margin of false data is lower than the standard deviation, making classical statistics-based detection ineffective. Furthermore, popular known information-theoretic detection approaches are bypassed under such low margins. At the same time, we showed that the attack impact on a utility under low attack strengths is significant. Thus, large dynamic variations in the data of smart living IoT systems enable valid low attack strengths to hide behind this randomness. Similarly, in Roy et al.,[4] we showed that current data measured at PMUs installed at the distribution level show high fluctuations.

For modeling, we denote δ_{avg} as a strategic attack strength variable, which is the mean of the sample perturbations δ_t (over the attack lifetime),

which gives the average extent of the falsification from original data values. The specific distribution of perturbations is dictated by the attack strategy used, as detailed in the "Attack Strategies" section. The perturbations δ_t are sampled from a strategic interval $[\delta_{min}, \delta_{max}]$. We name the width of this interval $|\delta_{max} - \delta_{min}|$ the aperture of attack strength. The aperture for attack simulation should be such that it does not raise obvious suspicions and violate the physical bounds of legitimate operation. The aperture affects the shape of the falsified data distribution and the ease of detection by artificial intelligence (AI)-based attack detectors.

Finally, a mathematical function quantifying the attack impact as a function of the attack strength is crucial for realistic attack and defense performance evaluation. Furthermore, a practical time horizon within which an adversary wants its attack budget investment to be accrued via the attack impact break-even time should be taken into consideration. A way to calculate this was shown in our previous work Bhattacharjee and Das.[2] The attack budget depends on the exploit used, which could very cheap. Therefore, lower attack strengths may have a quick break-even time and still offer tangible benefits to the adversary.

Data integrity attack types

Attack types specify the way data are falsified, depending on the goal. We give some examples of how different attack objectives manifest as attack types. Following this, one can model the attack type according to how any application works.

Additive. The adversary adds some strategic values to the original data point stream of an IoT endpoint such that the reported value $P^i_{rep}(t) = P^i_{act}(t) + \delta_t$, where δ_t is a sample from a strategic distribution whose mean converges to δ_{avg}. The goal of an additive attack can vary according to the IoT application and should be understood from the perspective of how the application is using the data. In the AMI context, it is achieved by a load altering exploit that causes smart meters to sense more than the actual power consumption, causing increased bills and unduly increased power generation.[2]

In the transportation context, an additive attack will prevent congestion detection. For this, the adversary needs to make sure that legitimately decreasing speeds from vehicles (due to real congestion and accidents) in a neighborhood are not visible to the traffic control application. To do that, the attacker has to add a strategic amount to the true speed values. Therefore, the adversary is effectively doing an additive perturbation to the original IoT data stream. In the PMU context, the additive attack type introduced on the current data stream by organized criminals/rival nations makes the control center believe in a sudden increase in the load that will lead to load shedding in that particular phase in a three-phase system. Therefore, for additive falsification, the modified attack sample is $I^i_t(t) = I^i_t(\text{act}) + I_{\delta_t}$ from a compromised PMU.

Deductive. From compromised IoT data streams, the adversary reduces the original data points such that the reported value $P^i_{rep}(t) = P^i_{act}(t) - \delta_t$. In smart metering, this is the most widely seen attack type, where the adversary's goal is to inflict losses for a utility or an equivalent gain for a large set of customers through low bills. In the transportation context, this attack type will fake traffic jams by reducing the values of speed data.[8] When launched in strategic areas, it will lead to traffic being rerouted from that area to create a strategic void that may be used for criminal activity and to create congestion elsewhere. In the PMU context, a reduced current implies a lower load from the customer side[4] that would falsely indicate that the control center should draw less power from generators and energy producers.

LARGE DYNAMIC VARIATIONS IN THE DATA OF SMART LIVING IOT SYSTEMS ENABLE VALID LOW ATTACK STRENGTHS TO HIDE BEHIND THIS RANDOMNESS.

Camouflage. The adversary divides the total attack scale into two groups, then launches additive attacks from one group and deductive attacks from the other group, maintaining the same attack strength for each group at the same time.[2] This guarantees a minimal difference in the mean and bypasses many common statistical detectors. From a stealth point of view, this attack type keeps the mean of the data the same at any point of time from all IoT devices bypassing common statistical

detectors. From an operational impact perspective, this attack is practical in the following ways. In smart metering, the deductive group of customers benefits at the expense of the additive group, while no suspicion is raised in the detectors that use sanity checks that measure the total inflow and outflow at the junction meters to detect evidence of deductive attacks.

In smart transportation (V2I), the network is divided into zones (clusters) for traffic monitoring. Within a zone, one subregion's TMC sensors[8] orchestrate deductive perturbation to fake a jam, while in another region, TMCs orchestrate additive perturbation in an area that actually is facing congestion. In this way, a decentralized zone-level anomaly detector has less suspicion due to an unchanged mean in the aggregate data reported from this traffic zone. At the same time, a microtraffic jam is faked, and a real traffic jam is missed, causing more traffic to avoid the area with the fake jam and instead go into the additive-perturbed TMC area where the congestion is real. This will worsen the original congestion in the area where the TMCs were additively perturbed.

Alternating switching. While camouflage attacks have additive and deductive attacks at the same time from different IoT devices, alternating switching involves an individual IoT device stream alternating between additive and deductive attacks over the attack lifetime (a time slice with equal parts additive and deductive attacks of the same δ_{avg}).[1] From a stealth perspective, the benefit is circumventing device-level diagnostics and detectors that use coarse-grained autoregression and time averages. From an operational impact perspective, the benefit is to exploit features, such as demand-based changes in smart living IoT applications. In Bhattacharjee et al.,[1] we show how alternating switching in smart meters separately alternates between additive and deductive attacks over the time domain, with the same margin of false data. In Roy et al.,[4] we showed that it makes sense for PMUs to alternate between high and low current values to create instability in the control systems that use such data.

Replay. A replay attack involves an adversary replaying older (believable) data to mask a change point related to an emergency event, which, when missed by the control center, will lead to an unsafe condition.[25] In this attack type, the adversary's goal is to prevent the system from detecting certain emergency conditions. The adversary usually remains silent for most of the time, waiting for an emergency or a special event to occur. As soon as this happens (detected via the data pattern), the adversary replays older readings that do not accurately reflect the altered state of the emergency. For example, in a winter polar vortex, smart meter data can show a spike due to the increased load from heating appliances. However, if the adversary replays older data points, the sudden spike in consumption data is hidden. Hence, the event will be missed, and appropriate countermeasures, such as increased generation and islanding, will not be taken. Similarly, this can happen with speed measurements in transportation networks using traffic incident detection applications. Similarly, in water distribution systems, a replay attack was shown in Palleti et al.[25] to mask the detection of a water leak. For this attack, the main simulation consideration is the effective attack lifetime that is equal to the required time between the attack start and when the intended damage is done. The attack lifetime could vary between applications and goals. The adversary needs to keep a copy of old values over a time span that is equal to the attack lifetime.

Mirroring. A variation of data replay is the mirroring attack,[4] which replays old data, instead of current data, like a mirror image, where the most recent old data are replayed first and the oldest data are replayed last. This type of attack is effective in creating instability by masking change points in an IoT network. Since the most recent data points are the first ones to be replayed, followed by older points, there is less change in change point and time series detectors. Therefore, less suspicion is raised compared to just replaying an old set of data values in the sequence in which they were recorded. The attack lifetime depends on the period through which the actual incident lasts.

Conflict. In this slight variation of the camouflage attack, additive and deductive attack groups do not have the same attack scale and strength.[6] A practical scenario for this is when an IoT application has been compromised by two different adversaries with conflicting goals; one adversary launches an additive attack, and the other launches a deductive attack, with different goals and stealth levels. The attack parameter here is the attack scale of the additive and deductive groups and their attack strengths.

Saturation. This type of attack occurs when sensed data from an IoT/CPS sensor get stuck at one value[20] and the sensor is unable to sense and send the real physical quantity. The exploit is an electromagnetic interference that causes "sensor saturation." Sensor

hardware has a well-defined "operating region" based on the expected range of the strength of input electromagnetic stimuli. If the input stimuli are within this range, sensors produce a linear digital output value proportional to the changes in the input stimuli. However, when the input stimuli strength is higher than the upper bound of the operating range, the sensor output gets saturated (that is, it gets stuck at one value that is approximately equal to the saturation point). To simulate such an attack realistically, one needs to find out the saturation point of the sensor type used in an IoT device. In the context of the medical IoT, Krebs on Security[20] showed how drop sensors stopped sensing the exact amount of fluid flowing through an infusion pump that controlled the amount of medicine injected into a patient's body.

Attack strategies

Attack strategies specify the way false data are injected into the space/time distribution of authentic data. Strategies are influenced by the level of prior knowledge and access. Prior knowledge could be further categorized into no prior knowledge, partial knowledge, and complete knowledge. Partial prior knowledge is realistic between the other two extremes. These include knowledge of data distributions and knowledge of state-of-the-art approaches to data integrity attack detection. Accordingly, the following possible falsification strategies can happen (see Figure 2): 1) data order aware, 2) on–off, 3) incremental ramp (or boil frog), and 4) Kullback–Leibler (KL) distance minimization.

These strategies can be easily launched from NAN, PDC, and RSU components for AMI, PMU, and V2I applications, respectively, which have visibility of multiple telemetry device data flows at once, as well as from a botnet that receives data from multiple IoT devices and implements these strategies with a certain attack type and strength. The attack scale will be how many device data flows are being manipulated.

Data order-aware strategy. In a data order-aware strategy, the adversary

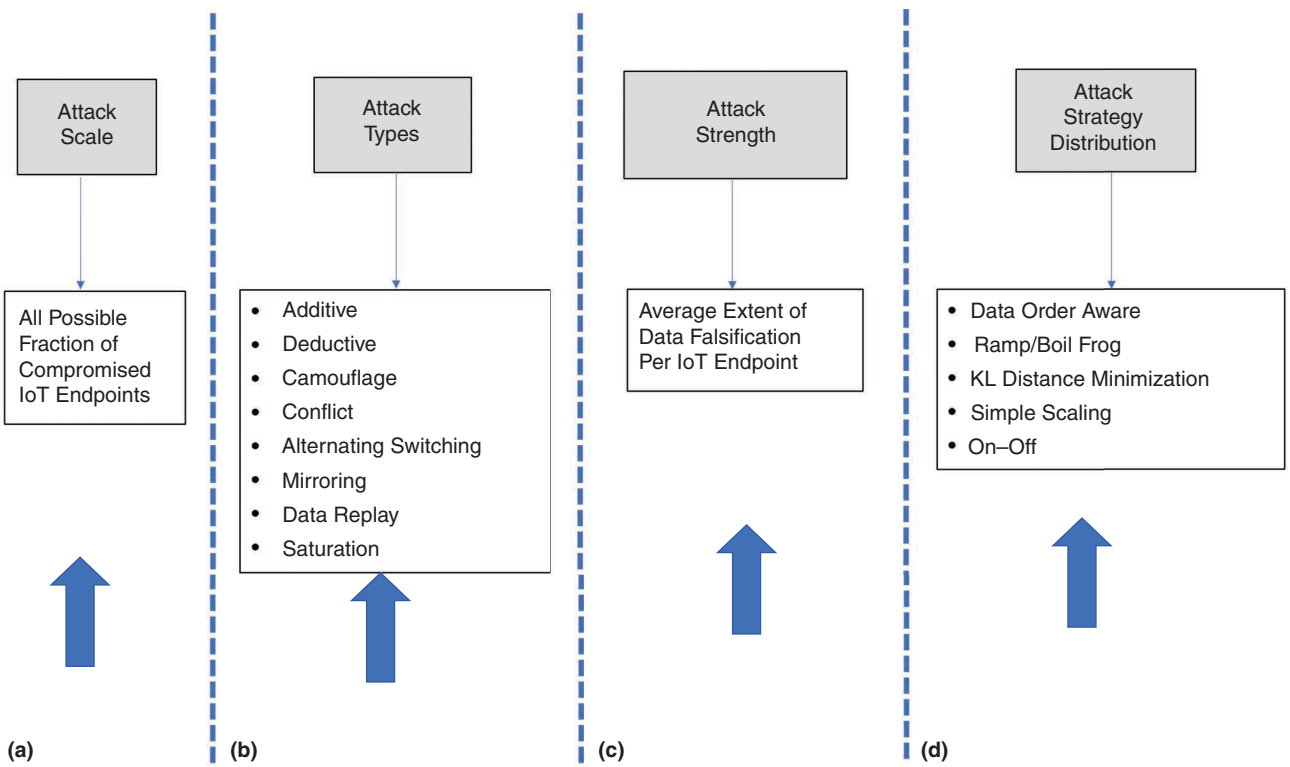

FIGURE 2. The unified data integrity threat landscape in smart living IoT telemetry data. Adversarial actions driven by the (a) attack budget, (b) attack goal, (c) intended severity and break-even time, and (d) stealth level and knowledge.

injects perturbations with just the knowledge of the extreme points of whatever is observed in any time slot. This strategy works as follows. The adversary intercepts the actual data from the set of M compromised (out of a total N) IoT devices/streams such that $P_t^{(1)}(\text{act}) \leq, \ldots, P_t^{(m)}(\text{act}) \leq P_t^{(M)}(\text{act})$. This may happen by compromising an aggregator or controlling multiple IoT sensing endpoints like a botnet. Subsequently, M random numbers are generated by the adversary for δ_t, sorted as $\delta_t^{\min} \leq, \ldots, \leq \delta_t^{\max}$.

For an additive attack, the lowest observed data are changed with the highest $\delta_t^{\max}$, while the highest observed power consumption data are modified with the lowest $\delta_t^{\min}$ and so on such that $P_t^{(1)}(\text{act}) + \delta_t^{\max}, \ldots, P_t^{(M)}(\text{act}) + \delta_t^{\min}$. For a deductive attack, the highest observed data are changed with the highest $\delta_t^{\max}$, while the lowest observed power consumption data are changed with the lowest $\delta_t^{\min}$. Hence, $P_t^{(1)}(\text{act}) - \delta_t^{\min}, \ldots, P_t^{(M)}(\text{act}) - \delta_t^{\max}$.

For a camouflage attack, the sorted $P_t^{(1)}(\text{act}) \leq, \ldots, \leq P_t^{(M)}(\text{act})$ is divided into two parts, and corresponding portions are changed accordingly. This strategy aims to minimize the sample distance between actual and falsified data, while keeping the same δ_{avg}. The data order-aware strategy is a nonoptimal but real-time and simple way of minimizing the pointwise distance between realizations of the original and attack distributions, and we reported the attack in previous work.[1,2] The evidence is shown in Figures 3(a) and (b), where the data order-aware line is closer to the original data distribution. Divergence-based detectors have a lesser probability of detection without sacrificing the operational impact of an attack.

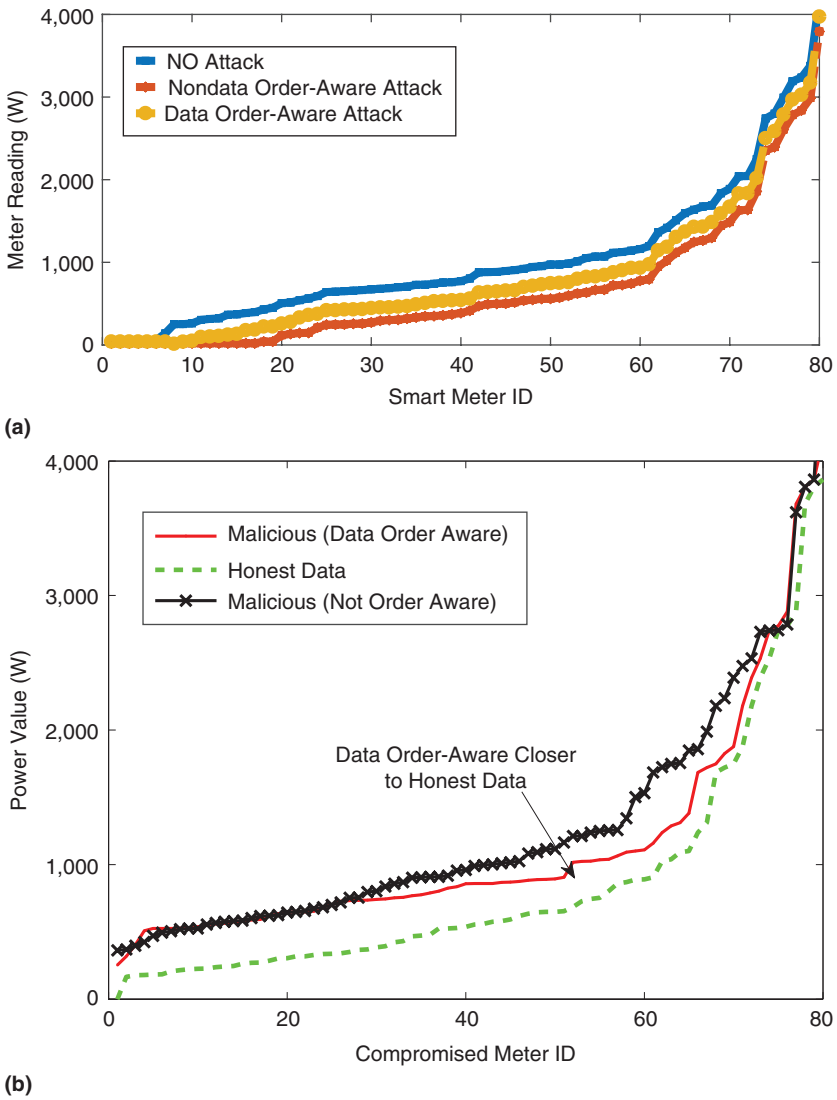

FIGURE 3. The data order-aware strategy: (a) deductive and (b) additive.

On–off strategy. The on–off strategy alternates between no attacks and attack periods; perturbations are sporadically distributed over the time domain. There are application-specific and application-agnostic benefits of the on–off strategy. The application-agnostic benefit of this is that such attacks can delay the convergence of machine learning (ML) and AI classifiers used in the identification of compromised IoT devices. An example using smart metering was shown in our recent work Bhattacharjee et al.[6] The delayed identification is caused because attacked data are sporadically hidden within large periods of no attacks. Due this imbalance, the anomaly is detected after a long time

horizon. Therefore, appropriate modifications in AI-based defenses are necessary to speed up detection under such strategies.

The application-specific benefit of on–off in AMI and PMUs arises from the dynamic demand-based pricing of electricity, which fluctuates throughput the day. Similarly, in transportation systems, traffic volumes are not uniform throughout all times. An adversary can be interested to attack only under certain occasions of high or low prices or traffic demands. For parameterized modeling, one needs to vary the length of each on and off period within realistic bounds. The second thing to vary is the on-to-off ratio in the total attack lifetime, which depends on how many on and off periods there are and the length of each on and off period.

Incremental/ramp strategy. The incremental/ramp strategy involves a very slow increment in the effective δ_{avg} bias over multiple time slices until the intended δ_{avg} is attained.[1,2] The goal of this is to ensure that time series update metrics record very small changes and fool them into thinking these changes are noise. This has a benefit in terms of not raising a sudden alarm in a change point detector, but the attacker eventually achieves its application-specific benefit by reaching the intended δ_{avg} after some delay. Similarly, if and when an adversary decides to start or stop attacking to mimic a leak in a water distribution system, this strategy would make sense because the leak grows over time. When the adversary intends to stop attacks (note that any strategy can be combined with on–off), the δ_{avg} can gradually decrease to prevent another obvious change point.[5]

For attack modeling and simulation, we need to consider two parameters: 1) the step difference variable $\Delta_i = \left|\delta_{\text{avg}}^{(i)} - \delta_{\text{avg}}^{(i+1)}\right|$ that dictates how much the attack strength changes between successive occurrences (i represents iteration number) and 2) the dwell time interval between successive increments $t_{(i+1)-i}$ that dictates the time gap between successive increments of the attack strength.

KL divergence-minimizing strategy. The KL divergence is used as a key concept in many ML classifiers for attack detection. While false data are injected, one strategy could be to inject a distribution that minimizes the KL divergence while preserving the target δ_{avg}. This ensures less obvious change in the classifiers. Figure 4(a) depicts the KL divergence-minimizing attack strategy for one smart meter device. The bold red line corresponds to the KL divergence-minimization strategy. Note that the data distribution under the KL divergence-minimization strategy tracks more closely to the true data distribution (the blue line) than to the simple scaling attack (the gray line), even when $\delta_{\text{avg}} = 200$ for both strategies. We implemented such an attack and discussed in stealth benefits in Bhattacharjee et al.[1]

For attack simulation, we need to consider the following: whether the

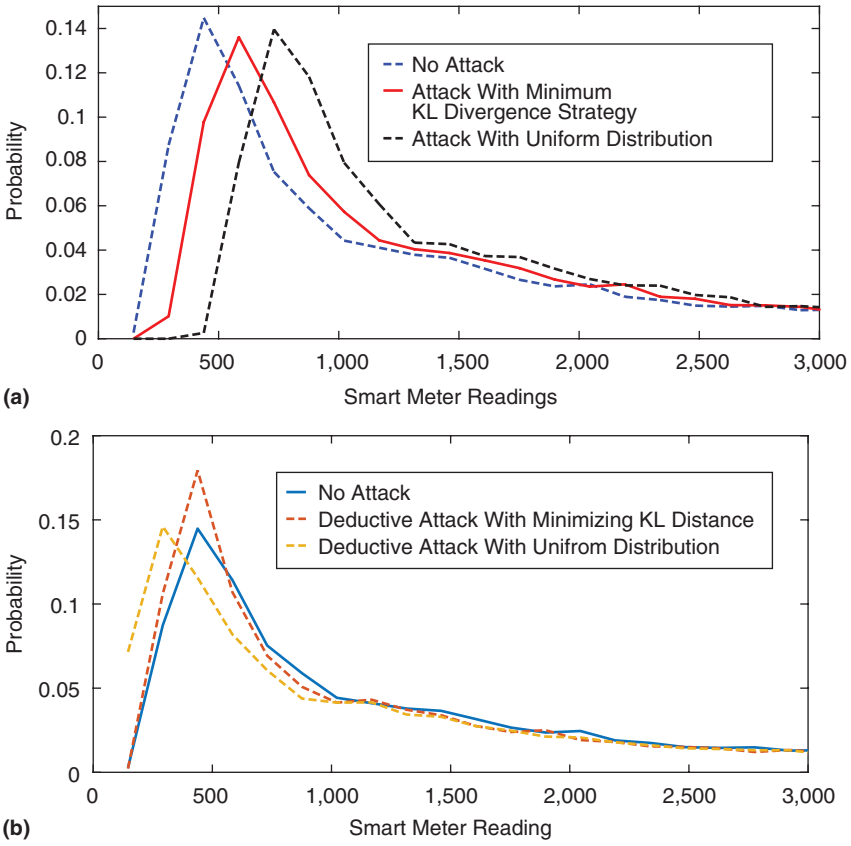

FIGURE 4. The KL distance-minimization strategy: (a) additive and (b) deductive.

attacker's strategy should be mean seeking (the forward KL) or mode seeking (the reverse KL).[24] If the defense mechanism is a supervised approach, then the forward KL should be minimized, and if it is based on reinforcement learning, the reverse KL should be minimized. However, this kind of attack strategy takes a longer time horizon to optimize and is practical only for delay-tolerant attack objectives.

Step strategy. In this simple attack commonly reported in the CPS literature,[23] the adversary modifies all samples to higher (additive) or lower (deductive) values by a constant $\delta_t(T_a)$ in a specified attack period T_a from the ith device, although $\delta_t(T_a)$ can change in a different attack period. Thus, the perturbation extent is mapped from a certain time context:

$$P_t^i = \begin{cases} P_t^i(\text{act}), & \text{if } t \notin T_a \\ P_t^i(\text{act}) + \delta_t(T_a), & \text{if } t \in T_a \end{cases}.$$

Scaling strategy. This attack involves the addition or subtraction of positive values (generated by a random function) to the actual measurements. It is the most commonly studied strategy.[2,5] The lower $\left(I_{\delta_{\min}}\right)$ and upper $\left(I_{\delta_{\max}}\right)$ bounds for selection are provided to the function as an input. While this is simple, it does not change the resultant shape of the load distribution drastically, making it a less obvious attack:

$$P_t^i = \begin{cases} P_t^i(\text{act}), & \text{if } t \notin \Delta_a \\ P_t^i(\text{act}) \pm \text{rand}(\delta_{\min}, \delta_{\max}), & \text{if } t \in \Delta_a \end{cases}.$$

To conclude, we showed six attack strategies and gave application-specific and application-agnostic benefits of each of the possibilities and implementation considerations.

ABOUT THE AUTHORS

SHAMEEK BHATTACHARJEE is with the Department of Computer Science, Western Michigan University, Kalamazoo, MI 49009 USA. His current research interests include information security in cyberphysical systems, the Internet of Things, wireless and social networks, particularly in topics such as anomaly detection, trust models, secure crowd-sensing, and dependable decision theory. Bhattacharjee received a Ph.D. from the University of Central Florida. He is a Member of IEEE. Contact him at shameek.bhattacharjee@wmich.edu.

SAJAL K. DAS is a professor of computer science and the Daniel St. Clair Endowed Chair at the Missouri University of Science and Technology, Rolla, MO 65409 USA. His research interests include cyberphysical systems, the Internet of Things, cybersecurity, pervasive and mobile computing, wireless sensor networks, and parallel computing. He is an associate editor of *IEEE Transactions on Dependable and Secure Computing, IEEE Transactions on Mobile Computing, IEEE/ACM Transactions on Networking*, and ACM *Transactions on Sensor Networks*. He is a Fellow of IEEE. Contact him at sdas@mst.edu.

In this article, we first explained why data integrity attacks offer a more credible threat in IoT/CPS systems, due to weaknesses arising in both cyber and physical domains. We also showed that the data falsification attack landscape should be specified by four main facets: 1) the attack scale, 2) the attack strength, 3) the attack type, and 4) attack strategies. Within each facet, we enumerated various possibilities of an attack state space. In this article, we explained the facets of the data falsification threat landscape in a way that allows a parameterized view of a threat model, rather than specific instances. This approach will enable researchers to understand what variables to introduce into attack simulations, how to encompass different adversarial goals and motivations behind inflicting operational damage to IoT/CPS utilities, and how to assess economic and service-oriented disruptions for customers. Next, we unified the literature on different attacks and showed how they fall under the preceding facets. Following this recipe will ensure that research in data integrity attacks on telemetry data for the IoT/CPS sensing loop embeds most possibilities of falsifying data to assess the operational impact and performance limits of various attack detection methods.

ACKNOWLEDGMENT

The work is supported by National Science Foundation grants SATC 2030611, SATC-2030624, and OAC-2017289.

REFERENCES

1. S. Bhattacharjee, P. Madhavarapu, and S. K. Das, "A diversity index based scoring framework for identifying smart meters launching stealthy data falsification attacks," in *Proc. ACM ASIA Conf. Comput. Commun. Security*, Jun. 2021, pp. 26–39, doi: 10.1145/3433210.3437527.

2. S. Bhattacharjee and S. K. Das, "Detection and forensics under stealthy data falsification in smart metering infrastructure," *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 1, pp. 356–371, Jan. 2021, doi: 10.1109/TDSC.2018.2889729.
3. A. Phadke, "Synchronized phasor measurements: A historical overview," in *Proc. IEEE/PES Transmiss. Distrib. Conf. Exhib.*, 2002, vol. 1, pp. 476–479, doi: 10.1109/TDC.2002.1178427.
4. P. Roy, S. Bhattacharjee, and S. K. Das, "Real time stream mining based attack detection in distribution level PMUs for smart grids," in *Proc. IEEE Global Conf. Commun.*, Dec. 2020, pp. 1–6, doi: 10.1109/GLOBECOM42002.2020.9322072.
5. C. Ahmed, A. Mathur, and M. Ochoa, "NoiSense print: Detecting data integrity attacks on sensor measurements using hardware-based fingerprints," *ACM Trans. Privacy Secur.*, vol. 24, no. 1, pp. 1–35, Feb. 2021, doi: 10.1145/3410447.
6. S. Bhattacharjee, V. Madhavarapu, S. Silvestri, and S. K. Das, "Attack context embedded data driven trust diagnostics in smart metering infrastructure," *ACM Trans. Privacy Secur.*, vol. 24, no. 2, pp. 1–36, 2021, doi: 10.1145/3426739.
7. S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture," National Institute of Standards and Technology, Gaithersburg, MD, USA, Special Publication-800-207, 2020.
8. M. Wilbur, A. Dubey, B. Leao, and S. Bhattacharjee, "A decentralized approach for real time anomaly detection in transportation networks," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, 2019, pp. 274–282, doi: 10.1109/SMARTCOMP.2019.00063.
9. Powder. Accessed: Aug. 15, 2022. [Online]. Available: https://powderwireless.net/
10. "Using advanced metering infrastructure in a water quality surveillance and response system," United States Environmental Protection Agency, Washington, DC, USA, Mar. 2021. [Online]. Available: https://www.epa.gov/sites/default/files/2021-03/documents/srs_ami_guidance_20210223_508_complete.pdf
11. "Insider threat report." Cybersecurity Insiders. Accessed: Aug. 15, 2022. [Online]. Available: https://www.cybersecurity-insiders.com/wp-content/uploads/2019/11/2020-Insider-Threat-Report-Gurucul.pdf
12. "RSA certificate vulnerability factoring RSA keys in the IoT era." KeyFactor. Accessed: Aug. 15, 2022. [Online]. Available: https://info.keyfactor.com/factoring-rsa-keys-in-the-iot-era#introduction
13. M. Hastings, J. Fried, and N. Heninger, "Weak keys remain widespread in network devices," in *Proc. ACM Internet Meas. Conf.*, 2016, pp. 49–63, doi: 10.1145/2987443.2987486.
14. C. Yan, H. Shin, C. Bolton, W. Xu, Y. Kim, and K Fu, "SoK: A minimalist approach to formalizing analog sensor security," in *Proc. IEEE Security Privacy Symp.*, 2020, pp. 233–248, doi: 10.1109/SP40000.2020.00026.
15. N. Heninger, Z. Durumeric, E. Wustrow, and J. Halderman, "Mining your Ps and Qs: Detection of widespread weak keys in network devices," in *Proc. USENIX Security Symp.*, 2011, pp. 1–16.
16. A. Cardenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proc. USENIX Conf. Hot Topics Secur. (HOTSEC'08)*, Berkeley, CA, USA, 2008, pp. 1–6, doi: 10.5555/1496671.1496677.
17. J. Liu, C. Yan, and W. Xu, "Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicles," in *Proc. DEFCON 24*, Aug. 2016, pp. 1–13.
18. "Advanced metering infrastructure and customer systems," U.S. Department of Energy, Washington, DC, USA, Sep. 2016. [Online]. Available: https://www.energy.gov/sites/prod/files/2016/12/f34/AMI%20Summary%20Report_09-26-16.pdf
19. K. Fu and W. Xu, "Risks of trusting the physics of sensors," *ACM Commun. Mag.*, vol. 61, no. 2, pp. 20–23, Jan. 2018, doi: 10.1145/3176402.
20. Y. Park, Y. Son, H. Shin, D. Kim, and Y. Kim, "This ain't your dose: Sensor Spoofing attack on medical infusion pump," in *Proc. USENIX Workshop Offensive Technol.*, 2016, pp. 1–33.
21. "FBI: Smart meter hacks likely to spread." KrebsOnSecurity. Accessed: Aug. 15, 2022. [Online]. Available: https://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/
22. N. Falliere, L. O. Murchu, and E. Chien. "W32 Stuxnet Dossier." Symantec. Accessed: Aug. 15, 2022. [Online]. Available: https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf
23. S. Pal, B. Sikdar, and J. H. Chow, "Classification and detection of PMU data manipulation attacks using transmission line parameters," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5057–5066, Sep. 2018, doi: 10.1109/TSG.2017.2679122.
24. A. Malinin. M. Gales, "Reverse KL-divergence training of prior networks: Improved uncertainty and adversarial robustness," in *Proc. NuerIPS*, 2019, pp. 14,547–14,558.
25. V. Palleti, V. Mishra, C. Ahmed, and A. Mathur, "Can replay attacks designed to steal water from water distribution systems remain undetected?" *ACM Trans. Cyber-Phys. Syst.*, vol. 5, no. 1, Jan. 2021, Art. no. 9, doi: 10.1145/3406764.