# Building a Unified Data Falsification Threat Landscape for IoT/CPS Applications

Shameek Bhattacharjee* and Sajal K. Das**
*Department of Computer Science, Western Michigan University, Kalamazoo, USA
**Department of Computer Science, Missouri University of Science and Technology, Rolla, USA
shameek.bhattacharjee@wmich.edu; sdas@mst.edu

***Abstract*—We layout a blue-print of a complete threat-landscape for data falsification/false data injection (FDI) attacks on telemetry data collected from IoT/CPS applications under zero-trust assumptions. Our motivation stems from the often-criticized lack of formal threat and realistic model for data falsification attacks on IoT and CPS telemetry data in the sensing loop. We first offer perspectives on why and when IoT/CPS telemetry data is vulnerable to false data injection attacks and related adversarial motivations. Thereafter, we enlist characteristic facets that should characterize the threat model for false data injection in IoT/CPS. Specifically, we discuss implementation issues for developing an advanced FDI attack simulation. Our approach helps researchers generate a parameterized threat state space universe, where many actual attacks are an instance within that threat state space. This will enable better validation of anomaly based attack detection methods that detect such attacks.**

***Index Terms*—Data falsification attack; CPS security; IoT security; Threat Landscape; Attack Simulation Datasets; False Data Injection**

## I. Introduction

Enormous amounts of telemetry data are being collected by smart devices, the so-called *Internet of Things* (IoT), which are the building blocks of cyber-physical systems (CPS). Such wide area telemetry data drive decision making and operations in smart living IoT applications (viz., smart energy, smart transportation etc.) that improve civic well-being.

Dependence on data analytics, immediate civilian impact of wrong decisions, economic motivations, vast attack surfaces (due to the community scale IoT and CPS domains), make an extremely attractive target of data integrity attacks.

Parallelly, the last decade has seen unprecedented advanced persistent threats (e.g., Stuxnet [1], Ukraine power grid attack) that can change various telemetry data, alter monitoring processes, when an adversary gains access privileges through a creative zero-day cyber, physical or social engineering exploit.

Traditional cyber security 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 from outside using cryptography, network traffic analysis, network segmentation and fine grained user-access control. However, many agencies such as Palo Alto Networks, US Department of Defense, and NIST, have formalized the need for Zero Trust Architecture [2], which recognizes that (static) trust on users and end point 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 cyber physical 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 smart living IoT. We first reveal why traditional cybersecurity practices are not enough and cyber physical factors that make data integrity attacks a credible threat. Then, we propose four facets that characterize the data integrity threat landscape of a IoT/CPS telemetry data. The four facets include: (1) Attack Types, (2) Attack Strength and Aperture, (3) Attack Scale, and (4) Attack Strategies. We provide 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 where 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 creates asymmetry between reality and perception which leads 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 use of telemetry data is common. Furthermore, we layout 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 dataset 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.

## II. Unified Abstraction of IoT/CPS

Regardless of the type of smart living IoT/CPS domain (viz., smart grid, smart transportation), a unified abstraction of the architecture and operations in smart living IoT domains is possible. 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) [3] and phasor measurement unit infrastructure (PMU) [4]. Each functional unit typically has a controller that runs specific services (e.g., demand response and automated billing by AMI, voltage sag detection by PMUs) based on the telemetry data collected from IoT endpoint devices (e.g., smart meters, phasor measurement units). For simplicity, we call the field area IoT end-points as just 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 meter's data, to a FAN gateway (fog node) for local area analytics. Multiple FANs gateways connect to a cloud controller hosting computations for wide area analytics (e.g., billing, load profiling) and decisions (signalling 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 [5].

Similarly, the distribution/transmission layers in smart grid have multiple PMUs acting as IoT devices that collect sample voltage, current amplitudes, angles between voltage and current on 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 connected to a wide area controller determining the events, state of the grid and control decisions (e.g., islanding, load balancing).

**Smart Transportation:** In this domain, vehicle-to-infrastructure (V2I) [6] and vehicle-to-vehicle (V2V) [7] units run telemetry data driven services to control traffic congestion, vehicular re-routing, platooning and incident response. Future V2I systems will collect vehicular data (e.g., speeds, road segment, velocity, direction) via dedicated short range radios hosted on smart cars. In other implementations, a Transport Measurement Channel (TMC) [6] equipment autosenses such data as vehicles pass by a road segment. Such data is forwarded to IoT devices known as road side units (RSU). These RSUs collect data from smart cars or TMCs and then forwards it to the respective fog and cloud servers for local and wide area analytics via internet. The fog/cloud server can issue control commands to vehicles or driving apps (e.g., rerouting and speed recommendations), or traffic signals (the signal switching information) for traffic management. The actuators are humans taking actions or signalling logic in traffic signals.

## III. Anatomy of Data Integrity Exploits in IoT

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

### A. Lack of Crypto-Agility in IoT

The secrecy of the RSA algorithm (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's algorithms' public key. Hence, if the prime factors are discovered, the adversary can re-derive the RSA's private key. Following this, the adversary can decrypt, modify the data from devices. To prevent the above, two key requirements need to be fulfilled: (R1) high randomness in prime numbers; (R2) enough computing power to transform input data into a strong key.

Random number generators in digital systems, rely on physical non-deterministic inputs/measurements that are sourced from the device hardware (e.g., mouse pointer movements, keystroke patterns, clock signals, phase noise, etc.). Computers/smartphones have the hardware that allows the collection of non-deterministic inputs, which creates randomness.

In contrast, IoT devices lack sources of randomness due to limitations in the attached hardware (e.g. absence of keystroke patterns, 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 and allowing the keys to be broken. The [8] found that most of the keys were broken and [9] reported that every 1 in 172 devices' digital certificates were compromised.

Pre-loading of keys on the 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 [10]. Also [11], noted that many industrial IoT/CPS systems cannot afford authentication and encryption altogether due to hard real-time operational requirements.

### B. Physical Data Manipulation via Transduction Attacks

IoT devices are vulnerable to transduction attacks [12] that disturb analog signals sensed by the device or physical tampering of firmware [13] such that it alters accurate conversion of analog to digital output of the telemetry data. This results in false data reported from the device. A transduction attack exploits a vulnerability in the physics of how sensor/hardware processes input analog signals as surveyed in [14] by directing some malicious electromagnetic signal as interference onto the IoT device. Sensors translate physical analog signals into electrical signals. Thereafter, the software in the firmware (or a remote App) interprets and reads the binary representations rather than the direct physical or electrical quantities. As adversary vary the intensity of the malicious signal on the target IoT devices, the extent of falsified data may change. In [7], a transduction attack was shown over a self-driving car.

Security practices such as static analysis, fuzz testing, and signed software updates do not offer detection of a sensor delivering false data [12]. Similarly, cryptography/network intrusion detection is unable to detect or prevent such attacks.

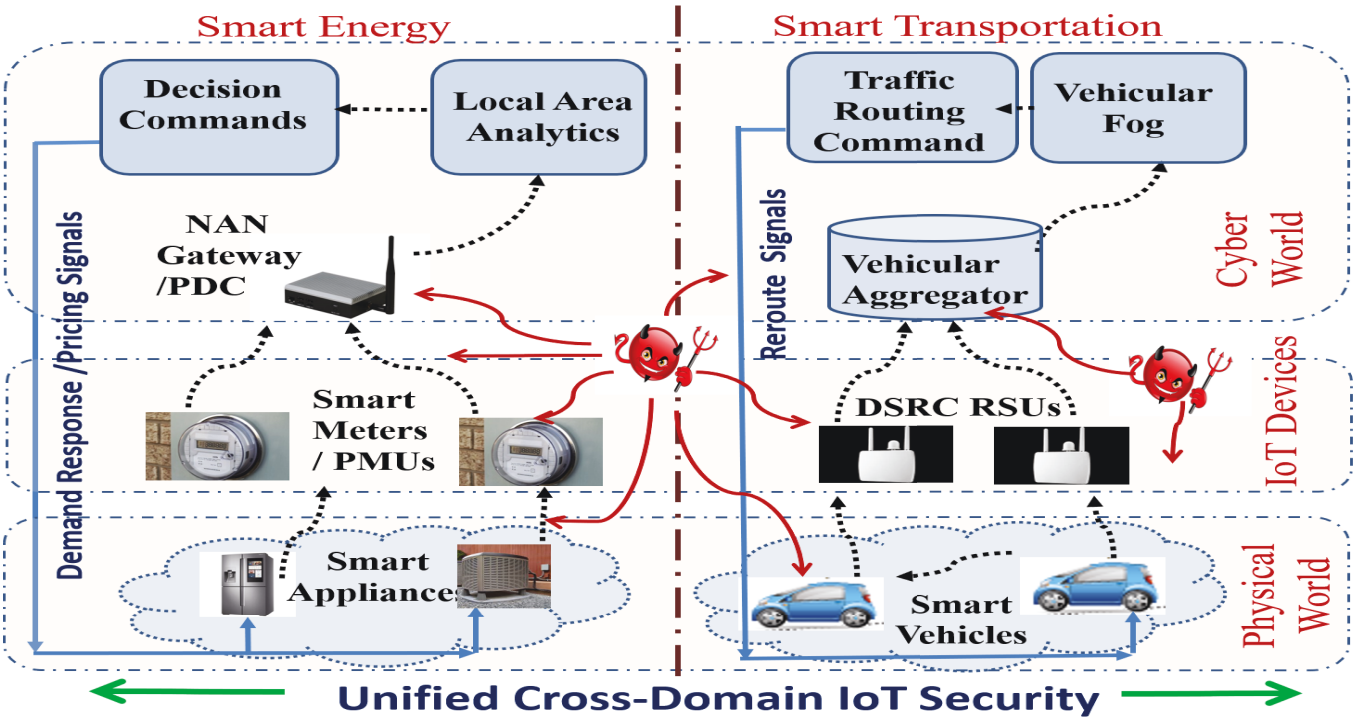



Fig. 1. **An example of Cross-Domain Layered IoT Architecture in CPS domains**

### *C. Insider Threats and Social Engineering Exploits*

A set of detailed case studies done in [15] finds 68% of the cybersecurity breaches have some involvement of at least one or more utility/enterprise insider in key positions of knowledge, even when the main threat actor is external to the organization. Furthermore, various social engineering exploits are used (e.g. phishing emails) on employees of an organization to extract privileges that bypass the traditional cybersecurity, and give control of data and devices to the adversary, causing data integrity attacks.

### *D. Heterogeneity in IoT Markets and Network Interfaces*

While some hardware security to transduction attacks exist, it requires all manufacturers to adhere to a common hardware security standard. Additionally, a single network contain devices assembled from tens to hundreds of distinct manufacturers. Most importantly, the sheer community scale of the IoT devices for smart living makes embedded and secure hardware an expensive solution as reasoned in [16]. Thus, detection of data integrity attacks from physical exploits is going to be a challenge in community-scale IoT and the status quo on commodity hardware in IoT markets.

Furthermore, IoT devices in smart cities are getting connected to the utility WAN via a field area network/edge interfacing layer that uses 5G, software radios, and shared spectrum technologies, which use highly programmable wired and wireless networking components [17]. The software-defined nature allows field area network devices to be compromised, enabling implementation of advanced attack strategies. This is further elaborated in Sec. IV-A and Sec. IV-D.

**Conclusion** The reasons for IoT devices being an attractive target is: (i) ease of physical access causes transduction/physical attacks, (ii) limited hardware/memory capabilities causing lack of cryptoagility, (iii) large scale deployment makes deploying in-situ embedded security in IoT devices very expensive and impractical for a utility provider, (iv) cyber connectivity of IoT devices to the internet, (v) data integrity attacks in smart living have both an immediate civilian as well as non-immediate economic impact based on the attack type.

## IV. 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 from the following categories of organized adversaries: (i) rival nation, (ii) insiders (iii) cyber criminals, (iv) rogue/selfish customers v) business competitors. Depending the capability, the actual parameters will vary but there is no way of predicting what it might be. Instead of snapshot attacks, our contribution provides a recipe to generate a superset of threat landscape that in turn allows unbiased evaluation of defense methods.

### *A. Attack Scale*

It refers to the number of compromised IoT telemetry devices or telemetry data streams. Redundancy is often present in the IoT telemetry end-point 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 assume only lower attack scales as 'realistic'. The above results in the following pitfalls:

First, the effective fraction of compromised devices depends on the *size of the network* and is unrelated to adversaries' 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 micro-grids and TMC sensors for decentralized traffic monitoring.

Second, the effective scale of attack, increases if an attacker compromises the intermediate data aggregator that is present in most wide area IoT/CPS applications (e.g., NAN gateway [3], PDC [4], RSU [6]) since it can manipulate *data streams* from multiple IoT devices at once. This is realistic possibility, since most edge aggregator devices in smart cities are planned to be wireless and programmable USRP radios which contain an USB port, easily accessible physically, and programmable over the air [17].

Third, the cheapness of the exploit should be taken into account. If the exploit is cheap, a high number of compromised IoT devices are possible, even with a small attack budget that is not impractically high. As an example, in 2009/10, an attack on Puerto-Rico's smart grid [13] metering utility's (PREPA), was carried out by utility insiders and maintenance personnel who tampered the thousands of smart meters to launch data falsification, using a portable optical laser probe toolkit that cost just $\$400$.

For proper scientific treatment, it is necessary that the research community parameterizes the attack scale from very small to large values and then test breakdown points to validate proposed defense models.

### B. Attack Strength and Aperture

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 when the margins of false data are below 450W. Similarly, we observed in [18] that existing works on PMU data integrity attacks rarely parameterize the average margin of falsification of current magnitude values.

In our recent work [19], we showed that since the standard deviation of the data is high in smart living IoT/CPS systems due human and environmental randomness, the average margin of false data if lower than the standard deviation, makes classical statistics-based detection ineffective. Furthermore, popular information-theoretic detection approaches are bypassed under such low margins [19]. At the same time, we showed that the attack impact on the utility under low attack strengths is significant. Thus, large dynamic variations in the data of smart living IoT enable valid low attack strengths to hide behind this randomness. Similarly, in [18] we showed that current data measured at PMUs installed at the distribution level show high fluctuations.

For modeling, we denote $\delta_{avg}$ as a strategic *attack strength* variable, which is the mean of the sample perturbations $\delta_t$ (over the attack lifetime), which gives the average extent of falsification from original data values. The specific distribution of perturbations is dictated by the attack strategy used as detailed in Sec IV-D.

The perturbations $\delta_t$, are sampled from a strategic interval $[\delta_{min}, \delta_{max}]$. We name the width of this interval $|\delta_{max} - \delta_{min}|$ as *the aperture of attack strength*. The aperture for attack simulation should be such that it does not raise obvious suspicions or violate the physical bounds of legitimate operation. The aperture affects the shape of the falsified data distribution and the ease of detection by AI-based attack detectors.

Finally, a mathematical function quantifying the *attack impact* as a function attack strength is crucial for realistic attack and defense performance evaluation. Furthermore, a practical time horizon within which the 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 [16]. The attack budget depends on the exploit used which could very cheap. Therefore, smaller attack strengths may have quick breakeven time and still offer tangible benefits to the adversary.

### C. Data Integrity Attack Types

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

*1) Additive:* The adversary *adds* some strategic values to the original data stream of an IoT end-point, such that the reported value $P^i_{rep}(t) = P^i_{act}(t) + \delta_t$, where $\delta_t$ a sample from a strategic distribution whose mean converges to $\delta_{avg}$. The goal of an additive attack can vary according to the IoT application, 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 the smart meters to sense more than actual power consumption, causing increased bills and undue increase power generation [16].

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 or accident) in a neighborhood, are not visible to the traffic control application. To do that, it 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 load that will lead to load shedding in that particular phase in a 3-phase. Therefore, for additive falsification, the modified attack sample is $I^i_t = I^i_t(act) + I_{\delta_t}$ from a compromised PMU.

*2) 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 the utility or an equivalent gain for a large set of customers with low bills.

In the transportation context, this attack type will fake traffic jams by reducing the values of speed data [6]. When launched in strategic areas, it will lead to traffic re-routed from that area to create a strategic void that may be used for criminal activity or create congestion elsewhere.

In the PMU context, a reduced current implies a lower load from the customer side [18] that would falsely indicate to the control center to draw lesser power from the generator or energy producers.

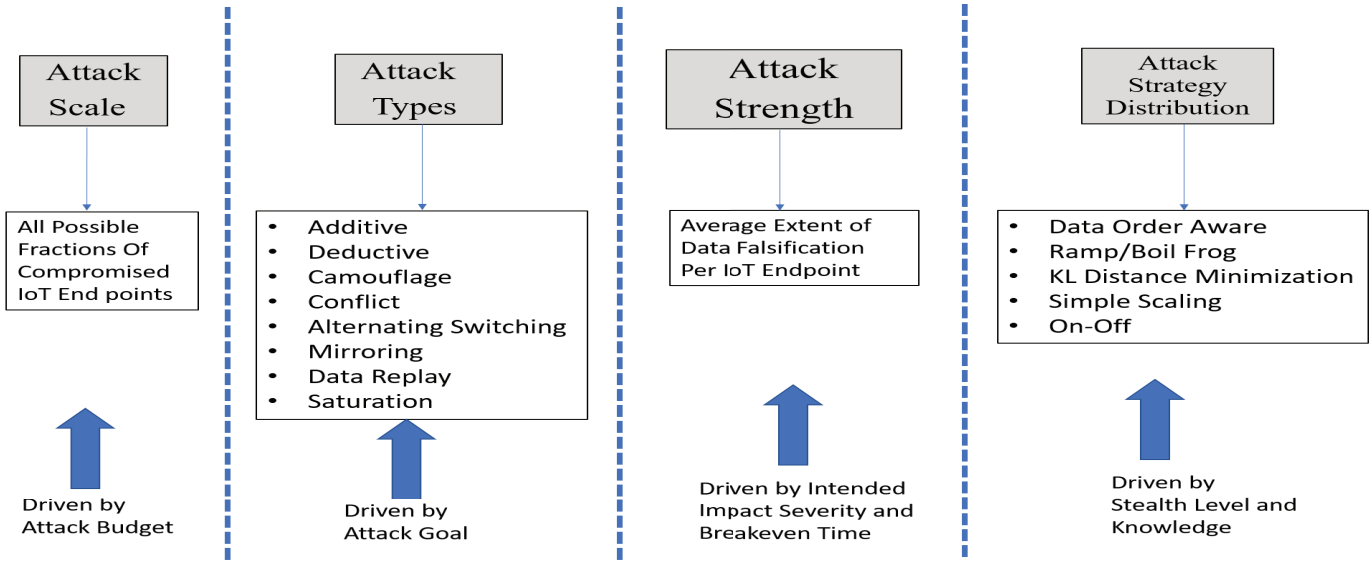



Fig. 2. **The Unified Data Integrity Threat Landscape in Smart Living IoT Telemetry Data**

*3) 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 [16]. This guarantees 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 and the expense of the additive group while no suspicion is raised in the detectors that use sanity checks that measure 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 sub-region's TMC sensors [6] orchestrate deductive perturbation to fake a jam while another region's TMCs additive perturbation in an area that actually is facing congestion. In this way, a decentralized zone level anomaly detector while has lesser suspicion due to an unchanged mean in the aggregate data reported from this traffic zone. At the same time, a micro traffic jam is faked and a real traffic jam is missed causing more traffic to avoid the area with a 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 TMCs were additively perturbed.

*4) Alternating switching:* While camouflage attacks have additive and deductive attacks at the same time from different IoT devices, the *alternating switching*, involve an individual IoT device stream alternating between additive and deductive attacks over the attack lifetime (a time slice with equal parts of additive and deductive attacks of same $\delta_{avg}$) [19].

From a stealth perspective, the benefit is circumventing device-level diagnostics or detectors that use coarse-grained auto-regression or time averages. From an operational impact perspective, the benefit is to exploit features such as demand-based changes in smart living IoT applications.

In [19], we show how alternating switching in smart meter separately alternates between additive and deductive attacks over the time domain with the same margin of false data. In [18], 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.

*5) Replay: 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 [20]. 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, the smart meter data can show a spike due to increased load from heating appliances. However, if the adversary replays older data points not, the sudden spike in consumption data is hidden. Hence, the event will be missed and appropriate countermeasures such as increased generation or islanding will not be taken. Similarly, this can happen on the speed measurements in the transportation network on traffic incident detection applications. Similarly, in water distribution systems, a replay attack was shown in [20] to mask the detection of water leak. For this attack, the main simulation consideration is the effective attack lifetime, that is equal to the required time between attack start and until 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.

*6) Mirroring:* A variation of the data replay is mirroring attack [18], which replays old data instead of the current data like a mirror image; where the most recent old data is replayed first and the oldest data is replayed last. This type of attack is effective in creating instability by masking change points in a IoT network. Since the most recent data points are the first

ones to be replayed followed by older points, there is a less change in a change point and time series detectors. Therefore, less suspicion is raised compared to just replaying an old set of data values in the sequence it was recorded. The *attack lifetime* depends on the period in the actual incident lasts.

*7) Conflict:* In this slight variation of camouflage attack, with the difference that additive and deductive attack groups do not have the same attack scale and strength [21]. A practical scenario for this is when an IoT application has been compromised by two different adversaries with conflicting goals; one adversary launching an additive attack and the other launching a deductive attack uncoordinated with different goals and stealth levels. The attack parameter here is the attack scale of the additive and deductive group and their attack strengths.

*8) Saturation:* This type of attack occurs when the sensed data from a IoT/CPS sensor get stuck at one value [22] and the sensor is unable to sense and send the real physical quantity. The exploit is a 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 is within this range, the 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 (i.e., it gets stuck at one value) that is approximately equal to the saturation point. To simulate such attack realistically, one needs to find out the saturation point of the sensor type used in the IoT device. In the context of medical IoT, the authors in [22] showed how the drop sensors stopped sensing the exact amount of fluid flowing through an infusion pump that controls amount of medicine injected in a patient's body.

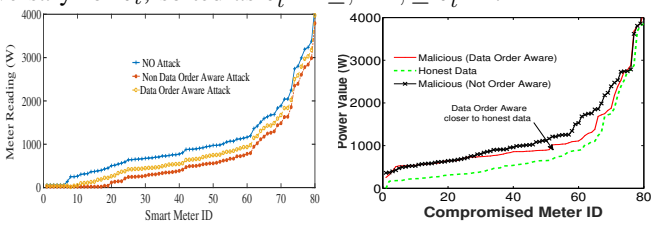
### D. Attack Strategies

Attack strategies specify the way the false data is injected into the space/time distribution of the authentic data. Strategies are influenced by 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: (1) Data order aware (2) On-off strategy (3) Incremental ramp or boil frog strategies (4) KL distance minimization strategy. The last two strategies (5) step and (6) scaling are added here for the sake of completion and are no knowledge attacks.

These strategies can be easily launched from a NaN, PDC, or RSU components for AMI, PMU, and V2I applications respectively, which have visibility of multiple telemetry device data flows at once or 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 intercepted.

*1) Data Order Aware Strategy:* In a data order aware strategy, the adversary injects perturbations with just the knowledge of the extreme points of whatever it observes any time slot. This strategy works as follows: The adversary intercepts the actual data from the set of $M$ compromised (out of total $N$) IoT devices/streams such that $P_t^{(1)}(act) \leq, \cdots, P_t^{(m)}(act), \leq P_t^{(M)}(act)$. 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 $\delta_t$, sorted as $\delta_t^{min} \leq, \cdots, \leq \delta_t^{max}$.




(a) (b)

Fig. 3. Data Order Aware Strategy (a) Deductive (b) Additive

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

For a deductive attack, the highest observed data is changed with the highest $\delta_t^{max}$, while the lowest observed power consumption data is changed with the lowest $\delta_t^{min}$. Hence, $P_t^{(1)}(act) - \delta_t^{min}, \cdots, P_t^{(M)}(act) - \delta_t^{max}$.

For a camouflage attack, the sorted $P_t^{(1)}(act) \leq, \cdots, \leq P_t^{(M)}(act)$ is divided into two parts, and corresponding portions are changed accordingly. This strategy aims to minimize the sample distance between actual and the falsified data, while keeping the same $\delta_{avg}$.

The data order aware strategy is a nonoptimal but a real-time and simple way of minimizing the pointwise distance between realizations of the original and attack distributions, and we reported the attack in our previous works [16], [19]. The evidence is shown in Figs. 3(a) and 4(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 the attack.

*2) On-Off strategy:* 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 on-off strategy.

The application agnostic benefit of this is that such attacks can delay convergence of ML and AI classifiers used in the identification of compromised IoT devices. An example using smart metering was shown in our recent work [21]. 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 the detection under such strategies.

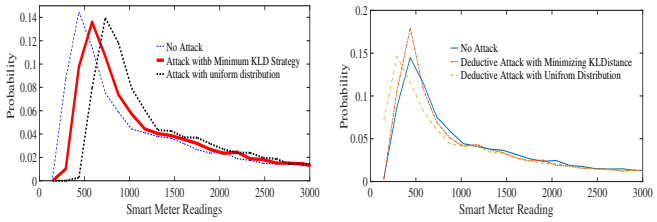
The application specific benefit of on-off in AMI and PMU arise from the dynamic demand based pricing of electricity that fluctuate throughput the day. Similarly, in transportation systems, the traffic volumes are not uniform throughout all times An adversary can be only interested to attack under certain occasions of high or low prices or traffic demands.

For parameterized modeling, one needs to vary *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 life time* of the attack, that depends on how many ON and OFF periods and the length of each ON and OFF period.

*3) Incremental and Ramp Strategy::* The incremental/ramp strategy are different names of the idea that involves a very slow increment in effective $\delta_{avg}$ bias over multiple time slices, until the intended $\delta_{avg}$ is attained [16], [19]. The goal of this is to ensure that time series update metrics record very small changes and fool them into thinking these changes as 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 $\delta_{avg}$ after some delay. Similarly, if and when an adversary decides to start or stop attacking in order to mimic a leak in water distribution system, this strategy would make sense because the leak grows over time. When the adversary intends to stop attacks (note any strategy can be combined with on-off), the $\delta_{avg}$ can gradually decrement to prevent another obvious change point [23].

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




(a) (b)

Fig. 4. KL Distance Minimization Strategy (a) Additive (b) Deductive

*4) KLD Minimizing Strategy:* Kullback-Leibler Divergence is used as a key concept in many ML classifiers for attack detection. While false data is injected, one strategy could be inject a distribution that minimizes the Kullback-Leibler Divergence, while preserving the target $\delta_{avg}$. This ensures less obvious change in the classifiers. Fig. 4(a) depicts an illustration of KLD minimizing attack strategy for one smart meter device. The bold red line corresponds to the KLD minimization strategy. Note that the data distribution under KLD Minimization strategy closely to the true data distribution (blue line) than to simple scaling attack (gray line), even when the $\delta_{avg} = 200$ for both strategies.We implemented such an attack and discussed in stealth benefits in [19].

For attack simulation, we need to consider the following: whether the attacker's strategy should be mean seeking (Forward KL) or mode seeking (Reverse KL) [24]. If the defense mechanism is supervised approach then forward KL should be minimized and if its based on reinforcement learning 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.

*5) Step Strategy:* In this simple attack commonly reported in the CPS literature [25], 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 $i$-th device, although $\delta_t(T_a)$ can change in a different attack period. Thus the perturbation extent is mapped from a certain time context.

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

*6) 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 [16], [23]. The lower ($I_{\delta_{min}}$) and upper ($I_{\delta_{max}}$) 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.

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

To conclude we showed 6 attack strategies and gave application specific and application agnostic benefits of each of the possibilities and implementation considerations.

## V. Conclusion

In this article, we first explained why the data integrity attacks offer a more credible threat in the 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: i) attack scale, ii) attack strength, iii) attack type, and iv) attack strategies. Within each facet, we enumerated various possibilities of an attack state space. In this paper we explained the facets of 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 the attack simulation; how to encompass different adversarial goals and motivations behind inflicting an operational damage to the IoT/CPS utility; and how to assess the economic or service oriented disruption for the customers. Next, we unified the literature on different attacks and showed how they fall under the above facets. Following this recipe will ensure that the research in data integrity attacks on telemetry data for IoT/CPS sensing loop, embeds most possibilities of falsifying data to assess the operational impact and performance limits of various attack detection methods.

**Acknowledgements** The work is supported by NSF research grants SATC-2030611, SATC-2030624, OAC-2017289.

## References

[1] N. Falliere, L. O. Murchu, E. Chien, "W32 Stuxnet Dossier", *Symantec, version 1.4.*, 2011. [Online] Retrieved from https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf.

[2] Rose, S. , Borchert, O. , Mitchell, S., Connelly, S. "Zero Trust Architecture" *NIST, Special Publication-800-207*, 2020 [Online] https://doi.org/10.6028/NIST.SP.

[3] US department of Energy Report, "Advanced Metering Infrastructure and Customer Systems", Sept. 2016.

[4] A. Phadke, "Synchronized Phasor Measurements - A Historical Overview",*IEEE/PES Transmission and Distribution Conference and Exhibition*, Vol. 1. pp. 476479, 2002.

[5] [Online] Using Advanced Metering Infrastructure in a Water Quality Surveillance and Response System, Available at: https://www.epa.gov/sites/default/files/2021-03/documents/srs_ami_guidance_20210223_508_complete.pdf

[6] M. Wilbur, A. Dubey, B. Leao, and S. Bhattacharjee, "A Decentralized Approach for Real Time Anomaly Detection in Transportation Networks," *IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 274-282, 2019.

[7] J. Liu, C. Yan, W. Xu, "Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicles" *DEFCON 24*, Aug. 2016.

[8] N. Heninger, Z. Durumeric, E. Wustrow, J. Halderman, "Mining Your Ps and Qs: Detection of Widespread Weak Keys in Network Devices", *USENIX Security Symposium*, 2011.

[9] M. Hastings, J. Fried, N. Heninger, "Weak keys remain widespread in network devices *ACM Internet Measurement Conference*, pp, 49-63, 2016.

[10] https://info.keyfactor.com/factoring-rsa-keys-in-the-iot-era#introduction

[11] A. Cardenas, S. Amin, S. Sastry, "Research challenges for the security of control systems" *USENIX conference on Hot topics in security (HOTSEC'08)*, Berkeley, CA, USA.

[12] K. Fu, W. Xu, "Risks of Trusting the Physics of Sensors" *ACM Communications Magazine*, Vol. 61(2), pp. 20-23, Jan. 2018.

[13] [Online]https://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/

[14] C. Yan, H. Shin, C. Bolton, W. Xu, Y. Kim, Kevin Fu, "SoK: A Minimalist Approach to Formalizing Analog Sensor Security", *IEEE Security and Privacy Symposium*, 2020.

[15] https://www.cybersecurity-insiders.com/wp-content/uploads/2019/11/2020-Insider-Threat-Report-Gurucul.pdf

[16] S. Bhattacharjee, S. K. Das, "Detection and Forensics under Stealthy Data Falsification in Smart Metering Infrastructure", *IEEE Transactions on Dependable and Secure Computing* Vol. 16, Jan. 2021.

[17] https://powderwireless.net/

[18] P. Roy, S. Bhattacharjee, S.K. Das, "Real Time Stream Mining based Attack Detection in Distribution Level PMUs for Smart Grids", *IEEE Global Conference on Communications*, Dec. 2020.

[19] S.Bhattacharjee, P. Madhavarapu, S.K. Das, "A Diversity Index based Scoring Framework for Identifying Smart Meters Launching Stealthy Data Falsification Attacks", *ACM ASIA Conference on Computer and Communications Security*, June 2021.

[20] V. Palleti, V. Mishra, C. Ahmed, A. Mathur, "Can Replay Attacks Designed to Steal Water from Water Distribution Systems Remain Undetected?" *ACM Transactions on Cyber-Physical Systems* Vol.(5), Article 9, Jan. 2021.

[21] S. Bhattacharjee, V. Madhavarapu, S. Silvestri, S. K. Das "Attack Context Embedded Data Driven Trust Diagnostics in Smart Metering Infrastructure" *ACM Trans. Privacy and Security*, Vol. 24(2):9, pp. 1-36, 2021.

[22] Y. Park, Y. Son, H. Shin, D. Kim, Y. Kim, "This Ain't Your Dose: Sensor Spoofing Attack on Medical Infusion Pump", *USENIX Workshop on Offensive Technologies*, 2016.

[23] C. Ahmed, A. Mathur, M. Ochoa, "NoiSense Print: Detecting Data Integrity Attacks on Sensor Measurements Using Hardware-based Fingerprints" *ACM Transactions on Privacy and Security* Vol 24(1), Article 2, Feb. 2021.

[24] A. Malinin. M. Gales,"Reverse KL-Divergence Training of Prior Networks: Improved Uncertainty and Adversarial Robustness", *NuerIPS*, pp. 14547-14558, 2019.

[25] S. Pal, B. Sikdar and J. H. Chow, "Classification and Detection of PMU Data Manipulation Attacks Using Transmission Line Parameters," *IEEE Transactions on Smart Grid*, vol. 9(5), pp. 5057-5066, Sep. 2018.


**Shameek Bhattacharjee** received his PhD and MS degrees from the University of Central Florida, Orlando, in 2015 and 2011, respectively and BS from the West Bengal University of Technology, India, in 2009. He is currently an assistant professor with the Department of Computer Science at Western Michigan University. Between 2015-2018, he worked as a post-doctoral researcher with Missouri S & T at Rolla, Missouri. His current research interests include information security in cyber-physical systems, internet of things, wireless and social networks, particularly in topics such as anomaly detection, trust models, secure crowd-sensing, and dependable decision theory. Additionally, his secondary research interest is in the area of resource and energy efficient networking. He is a recipient of the Provost Fellowship and IEEE PIMRC Best Paper Award. He serves as a TPC member in various leading conferences such as IEEE ICDCS, IEEE SECON, IEEE/ACM Middleware, IEEE SMARTCOMP, IEEE WoWMoM, IEEE ICC, and a regular reviewer in leading journals and conferences such as IEEE Trans. on Mobile Computing, IEEE Trans. of Dependable and Secure Computing, IEEE J. of Selected Areas in Communication.

**Sajal K. Das** is a professor of Computer Science and the Daniel St. Clair Endowed Chair at the Missouri University of Science and Technology, where he was the Chair of Computer Science Department during 2013-2017. His research interests include cyber-physical systems, IoT, cybersecurity, pervasive and mobile computing, wireless sensor networks, and parallel computing, among others. He has made fundamental contributions to these areas and published extensively in high quality journals and peer-reviewed conference proceedings. He holds 5 US patents and coauthored 4 books, such as Handbook on Securing Cyber-Physical Critical Infrastructure: Foundations and Challenges, and Principles of Cyber-Physical Systems: An Interdisciplinary Approach. His h-index is 96 with more than 37,000 citations. He is a recipient of 12 Best Paper Awards at conferences like ACM MobiCom and IEEE PerCom, and numerous awards for teaching, mentoring and research including the IEEE Computer Societys Technical Achievement Award for pioneering contributions to sensor networks and mobile computing, and University of Missouri System Presidents Award for Sustained Career Excellence. Dr. Das serves as the founding Editor-in-Chief of Elseviers Pervasive and Mobile Computing Journal, and as Associate Editor of the 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 an IEEE Fellow.