# Statistical Security Incident Forensics against Data Falsification in Smart Grid Advanced Metering Infrastructure

Shameek Bhattacharjee[1], Aditya Thakur[2], Simone Silvestri[1], and Sajal K. Das[1]
{shameek, astvd3, silvestris, sdas}@mst.edu
[1]Department of Computer Science, [2]Department of Electrical Engineering
Missouri University of Science and Technology, Rolla, MO, USA

## ABSTRACT

Compromised smart meters reporting false power consumption data in Advanced Metering Infrastructure (AMI) may have drastic consequences on a smart grid's operations. Most existing works only deal with electricity theft from customers. However, several other types of data falsification attacks are possible, when meters are compromised by *organized rivals*. In this paper, we first propose a taxonomy of possible data falsification strategies such as *additive, deductive, camouflage and conflict*, in AMI micro-grids. Then, we devise a statistical anomaly detection technique to identify the incidence of proposed attack types, by studying their impact on the observed data. Subsequently, a trust model based on Kullback-Leibler divergence is proposed to identify compromised smart meters for additive and deductive attacks. The resultant detection rates and false alarms are minimized through a robust aggregate measure that is calculated based on the detected attack type and successfully discriminating legitimate changes from malicious ones. For conflict and camouflage attacks, a generalized linear model and Weibull function based kernel trick is used over the trust score to facilitate more accurate classification. Using real data sets collected from AMI, we investigate several trade-offs that occur between attacker's revenue and costs, as well as the margin of false data and fraction of compromised nodes. Experimental results show that our model has a high true positive detection rate, while the average false alarm rate is just 8%, for most practical attack strategies, without depending on the expensive hardware based monitoring.

## Keywords

Statistical Anomaly Detection; Security Incident Forensics; Trust Models; Data Falsification; Information Theory; Supervised Learning; Smart Grid; Advanced Metering Infrastructure; Relative Entropy

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

*CODASPY'17, March 22-24, 2017, Scottsdale, AZ, USA*

© 2017 ACM. ISBN 978-1-4503-4523-1/17/03…$15.00

DOI: http://dx.doi.org/10.1145/3029806.3029833

## 1. INTRODUCTION

Advanced Metering Infrastructure (AMI) is one of the elementary units of the smart grid technology, which collects data on loads and consumer's power consumption [10], from Smart Meters installed on the customer site (see Fig. 1). Such data play a pivotal role in several critical tasks such as automated billing, demand response, load forecast and management [10].

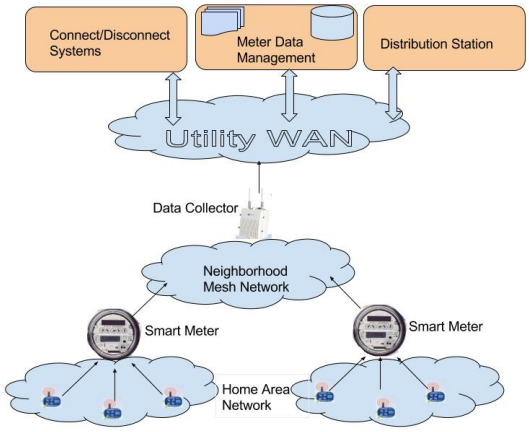



**Figure 1: Architecture of AMI**

Apart from automated billing, (already in use), strategic decisions are expected to be taken by future smart grids, based on the power consumption data. For example, these data will have implications on tasks such as daily and critical peak shifts [22]. When the consumption increases beyond a certain critical limit, emergency 'peaker plants' are currently used by most utilities for additional power generation to meet the demand. However, such peaker plants are extremely carbon as well as cost intensive. In the modern grid, the utility will also have the option for automated demand response where utilities pay customers to shut certain appliances temporarily (peak shifting) to obviate the need for additional generation [21]. In general, an accurate short or long term data on loads and consumption will aid in accurate demand response, load forecast and planned generation in the future smart grid. Hence, the integrity of the data generated from the AMI is of utmost importance.

Defense against falsification of power consumption data from AMIs, has largely focused on *electricity theft* [3, 7, 9, 17], where individual customers are primary adversaries who report lower than actual usage for lesser bills. Since isolated

smart meters belonging to rogue customers reduce the value of power consumption, we term such adversarial strategy as a *Deductive* mode of data falsification.

However, it has been widely acknowledged that given the cyber and interconnected nature of AMI, it could potentially be the target of powerful and organized adversaries such as rival nation states, utility insiders [20], organized cyber criminals and business competitors [4]. Such adversaries can compromise *several smart meters* and then *spoof* false power consumption data [7] from smart meters. Powerful and organized adversaries are more equipped to crack/leak cryptographic secrets, have a higher attack budget, and possess the ability to simultaneously attack other elements of the grid (e.g., audit logs, transformers meters) in order to avoid easy consistency checks on false data. Existing research does not focus on defense against such adversaries and is only restricted to attacks from isolated adversaries.

Additionally, the goals of these organized adversaries are not just restricted to monetary benefits on the customer side that result from electricity theft. As a recent real example, in Puerto Rico [20] a manufacturer and a utility insider colluded to install a large number of tampered smart meters that reported higher than actual power consumption. We term such an attack as *Additive* mode of data falsification. Conversely, an additive attack launched by a rival utility on its competing company's meters may induce loss of business confidence by the customers of the victim company, due to higher bills as reported in [19]. A class action lawsuit filed against a victim utility was reported in this case. If the utility participates in demand response, then utility may lose revenue from additive attacks for undue compensation payed to customers for induced peak shifts. Indirectly, additive attacks can be triggered when a load altering attack *(LAA)* [11] occurs on the individual appliances of a Home Area Network (HAN), thus increasing the net consumption sensed by the smart meter. It may also be noted that rival nations, businesses, or organized cyber criminals may orchestrate large scale deductive attacks to cripple the utility companies through revenue losses. Additive and deductive attacks are termed as 'simple' attack types.

Furthermore, we argue the possibility of mixed attack types on the smart meter data. For example, a balancing additive and deductive attack with the same margin of falsification of either type, could evade mean aggregate demand check/forecast models. We term such a strategy as a *Camouflage* attack, which may be motivated for generating lesser bills to one set of customers at the expense of the other set. Such attacks may stay undetected, without raising any suspicion because the total inflow and outflow of power measured at the transformer meters, and the total demand and reported usage remain unchanged. The attacker in such a case need not attack other elements in the grid (for e.g. transformer meters) to prevent easy consistency checks. In general, random additive and deductive attacks may simultaneously coexist in the same AMI network, when launched by different adversaries with conflicting goals. We term such a scenario as a *Conflict attack*, which is a mixed attack type with unequal margins of falsification for each underlying simple attack type. Existing literature cannot handle all of the above data falsification strategies.

In this paper, we first introduce a taxonomy of possible data falsification attacks launched by organized adversaries. Then we study the statistical properties of the distribution of power consumption data and analyze the effects of various data falsification attacks types on the parameters of the power consumption data. With the help of the observed statistical effects, *a security incident forensics criterion* is proposed that indicates the presence and the type of attack while discriminating effect of attacks from legitimate changes. Subsequently, we propose a light weight Kullback-Leibler divergence based trust model that identifies compromised meters with a high detection rate, by exploiting knowledge of the statistical impact caused by each attack type, cyclostationarity of overall power consumption patterns, and factoring in for any legitimate change in the consumption patterns.

We use a generalized linear model (GLM) and Weibull function based kernel trick to extend our trust model, for robust classification of compromised nodes based on the computed trust values with least missed detections and false alarms, even for stealthy camouflage and conflict attacks. We also perform a cost benefit and sensitivity analysis for both of our attack and defense models. Specifically, we computationally study tradeoffs, such as *breakeven time* (i.e., the time required for attacker's revenue to equal its cost of attack) and *breakdown point* (i.e., the attack strategy for which the defense mechanism is no longer able to distinguish the compromised meters from honest ones). Experimental results show that our detection technique is able to identify compromised meters with higher detection rates while incurring lower false positives, than most existing works, under rational attack strategies that may be employed by adversaries. We perform extensive sensitivity analysis to show the limits of our model.

To the best of our knowledge, our proposed work is the first effort to establish trustworthiness in AMI against multiple attacks types and organized rivals. Secondly, ours is the first work that focus on data falsification strategies other than electricity theft, which can be devised by organized adversaries rather than rogue customers. Unlike most existing works, our approach works without meter specific storage and maintenance of fine grained consumption data from each meter to obviate important privacy concerns [26]. Our proposed method is also light weight compared to the classical bad data detection mechanisms. Since our method does not require installation of additional hardware as in the state based monitoring, it is more cost effective.

## 2. RELATED WORK

Existing work on AMI data falsification can be broadly categorized into *classical bad data detection*, and *state based detection*. Both categories focus only on the electricity theft. Classical bad data detection uses techniques such as Support Vector Machine (SVM), Neural Networks and Auto Regressive Moving Average (ARMA) models. In contrast, state based detection includes sensor based monitoring, mean aggregate outlier inspections and transformer state estimation.

Classical bad data detection schemes such as Multi-class SVM and Neural Networks are used in [3, 15] for offline and retrospective identification of rogue customers stealing electricity by reporting lower usage. Such techniques contain a series of seven steps for identifying abnormal customers. The obvious disadvantage of these approaches are that they are highly computation expensive, deal with long term retrospective identification, and do not consider organized adversaries or their attack strategies, and require full and fine

grained profiling of each smart meter. A comparative analysis of classical bad data detection schemes is provided in [2] which concludes that while these schemes require full profiling of each customers' energy consumption (thus cannot protect their privacy), the detection rate of most of these schemes is approximately 60%-70%. Moreover, only two schemes provide a quantitative false positive rate.

Finally, ARMA based models [9] profile each customer's time series data separately to increase accuracy, using ARMA-GLR detector. However, in most practical cases, the consumption cannot be accurately modeled as an ARMA process [3], resulting in the detection rate of only about 62%. Additionally, several privacy threats [10, 26] are associated with such approaches since they require customer specific monitoring and maintainance of fine or coarse grained consumption data.

State based detection techniques like sensor based monitoring [5, 8], transformer state monitors [1] require additional hardware deployed at various points across the AMI and distribution network for identifying anomalies. However, most of them do not identify the compromised meters. Additional hardware requirement makes such approaches costly [2]. Some works (e.g., [8]) combines the audit logs for physical and cyber events in the meter to check for consistency in the data reported. But these approaches are nullified when the meters are compromised by external adversary who are intelligent enough to change the audit logs. Furthermore, cyber connected sensors/monitors are also similarly vulnerable to cyber attacks.

Mean aggregate based outlier approaches used in state based detection [15, 17] have the advantage that they do need not store and maintain fine grained meter specific trends on power consumption. In [17], an arithmetic mean aggregate approach is proposed; but the number of rogue meters is small compared to the population, hence the aggregate mean values are not affected enough. They also do not discriminate between legitimate and malicious changes in the mean consumption, thus incurring a high false positive rate of around 30%, and consider only electricity theft. Legitimate changes in consumption may occur due to weather and other contextual factors. Approaches focusing on only arithmetic mean aggregates and median have difficulty to discriminate malicious changes from legitimate ones when the margin of false data or the fraction of compromised nodes is higher. Thus, such approaches suffer from high false positive rates or lower detection rates. Additionally, such methods will also fail to identify camouflage and conflict attacks as discussed earlier. In [3], a false positive rate of 28% is reported although the detection rate is high.

Finally, cryptographic approaches [6, 16] may fail to provide any help as organized adversaries may be able to crack the cryptographic secrets. Moreover, given the latency critical nature of functions like demand response and management [23], advanced cryptographic defense is impractical due to additional overhead [14]. This further exacerbates the vulnerability of the AMI data falsification.

Given the above limitations, we believe there is a dire need for trustworthy computing approaches based on the anomaly detection in the data reported by each meter. This motivates us to propose a novel scheme that provides security forensics to identify various falsification attacks and a trust model based on aggregate data monitoring to identify compromised meters. Hence, our work significantly advances this field of research.

# 3. SYSTEM MODEL

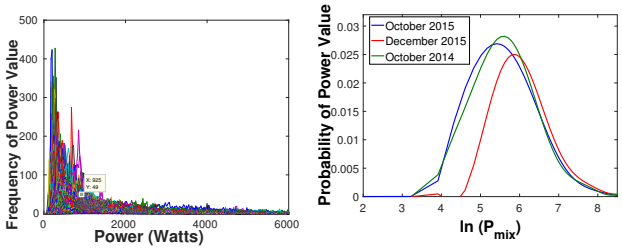
We consider a collection of $N$ smart meters reporting power consumption data to a Data Collector (DC) periodically and independently. The $i$-th smart meter, $s^i$, records an actual power consumption data $P_t^i(act)$ at the end of each time slot $t$. The reported power consumption $P_t^i(rep)$ is equal to $P_t^i(act)$ if $s^i$ is not compromised. However, $P_t^i(rep) \neq P_t^i(act)$, if $s^i$ is compromised by an adversary. We model $P_t^i(act)$ as the realizations of a random variable $P^i$. The Data Collector piggybacks data from each smart meter and sends it to the billing utility. The total power reported at a time by all $N$ meters is sent to a transformer meter.

To characterize the distribution of $P^i$ from the $i$-th smart meter, we conducted preliminary investigations on real power consumption data sets [25], of 200 houses from 16 different microgrids. Each home consists of one smart meter. We observed that for each house or meter, the power consumption can be approximated as a log normal distribution. We also observed that all such log normal distributions are *clustered close* to each other; that is, the variance between them is not arbitrarily large. Fig. 2(a) summarizes the results.

We approximate the aggregate of the individual log normals using a mixture distribution, which is also lognormal as evident from Fig. 2(a). We denote $P_{mix}$ as the random variable (r.v.) of such aggregate approximate mixture distribution.

For mathematical tractability and visual intuitiveness, we transform $P_{mix}$ on a natural logarithm scale to obtain an approximate *normally distributed* r.v. denoted as $p_{mix}$. Results of this approximate normal mixture $p_{mix}$, for different months in the recent past is depicted in Fig. 2(b). Note that both $P_{mix}$ and $p_{mix}$, *do not reveal* any consumption pattern for each specific meter, but only a general trend on the consumption.

We also denote $p_t^i(act) = ln(P_t^i(act))$, as the effective power consumption report recorded at each meter $s^i$ on a log scale at any time slot $t$. Note that, for certain other data sets (like a wider area monitoring), with more than one consumption clusters, as in [15], our approach can be applied to each such cluster independently. The proposed trust model calculates and updates trust of each smart meter at the end of each month. We assume a window size of $T$ slots per month based on how $t$ is slotted. To prove the generality, we repeated the experiments for a different AMI data set [3], and reported similar observations like Figs. 2(a) and 2(b) as shown in Appendix B.




(a) (b)

**Figure 2: Power Consumption Behavior: (a) Actual (b) Normal Mixture**

## 3.1 Threat Model

We consider the following assumptions in our threat model.

### 3.1.1 Types of Adversary

We assume that the organized adversary belongs to either rival nation states, business competitors, utility insiders or cyber criminals, possessing the ability to compromise several smart meters by bypassing cryptography.

False power consumption data from a meter can be achieved in the following ways: (a) manipulation of inputs to the meter, (b) in rest at the meter, and (c) in-flight from the meter. The adversary then launches data falsification from multiple such compromised smart meters concurrently.A compromised meter in this context means either the input, content, or output coming from one specific meter is compromised.

We assume rational attackers who may have a long or short term damage objective. *Long term damage* requires evading detection for the maximum possible time, while still benefiting from attacks. The adversary may accept to face some initial loss in the hope of evading detection and accruing incremental benefits over time. Examples of long term adversarial objectives include monetary gains in terms of electricity pricing and belief manipulation of learning demand forecast models. A *short term damage*, on the other hand, requires inflicting the maximum damage in a short time, before getting detected. Examples of short term objectives include an attacker aiming to, gain quick revenue or masquerade a false high demand response. Due to the contrasting requirements on these two objectives, adversarial decisions such as fraction of compromised nodes $\rho_{mal}$ and the margin of false data are dependent on the nature of time deadlines associated with such objectives.

Unlike existing works [3], we assume organized adversaries to be intelligent enough to also tamper with transformer meters and other portions of the grid, to escape easy consistency checks on false data. We also assume that the smart meters report the consumed power to a data concentrator on every time slot (hourly). Compromised Meters spoof false data on all time slots, however, the attack margin $\Delta$ (explained below) may be more in peak periods than non-peak periods, to exploit the time dependent pricing of electricity.

### 3.1.2 Taxonomy of AMI Data Falsification

We define the manner in which the actual power consumption data $P_t^i(act)$ of each meter $s^i$ is modified as the *mode* of data falsification. We identify the following modes:

*Additive:* The adversary reports $P_t^i(rep) = P_t^i(act) + \Delta_t$, where $\delta_{min} \leq \Delta_t \leq \delta_{max}$. This mode can lead to loss of business confidence from customers due to higher bills and masquerade a critical peak leading to remote disconnect of customer appliances, thereby causing utilities to pay undue incentives.

*Deductive:* The adversary reports $P_t^i(rep) = P_t^i(act) - \Delta_t$, where $\delta_{min} \leq \Delta_t \leq \delta_{max}$. This mode can lead to loss of revenue for power utility companies.

*Camouflage:* The adversary divides the compromised meters into two teams equal in number, which simultaneously adopt an additive and deductive mode, respectively. This mode can favor smart meter of one power utility at the expense of others, and has less impact on the strategic decisions in the grid. It cannot be detected by simple mean comparison approaches, because no suspicion is raised due to negligible change in the total reported power consumption.

*Conflict:* It is a scenario where additive and deductive attacks coexist simultaneously, but are not necessarily balanced. Such a scenario represents random attacks possible if there are more than one uncoordinated adversarial teams.

### 3.1.3 Margin of Falsified Data

The value $\Delta_t$ is generated randomly within an interval $[\delta_{min}, \delta_{max}]$, for $\delta_{min}, \delta_{max} > 0$, and accordingly added to or deducted from the actual power consumption. Note that, arbitrarily high $\delta_{max}$ may facilitate intuitively easy detection, while very low $\delta_{max}$ hardly accrues any revenue. The average value of $\Delta_t$ is represented as $\Delta_{avg}$.

Apart from the type of attack, the attacker chooses a value of $\Delta_{avg}$ in the interval $[\delta_{min}, \delta_{max}]$ as part of its attack strategy. $\Delta_{avg}$ may be high or low depending on the amount of damage it wants to inflict, and the short or long time horizon of the attack. All units of $\Delta_{avg}$ values discussed throughout the paper is in Watts.

Since the distribution of power consumption is unimodal, the attacker refrains from any strategy that would make the resultant distribution, multi modal. In that sense, a uniformly distributed random noise injected into the actual smart meter data does not change the overall shape of the distribution but only effects its parameters. Such variants of uniform distribution over time is adopted from [3].

From the defender's perspective, we define the *breakdown point* $BDP = (\Delta_{avg}, \rho_{mal})$, as a combination of $\Delta_{avg}$ and $\rho_{mal}$ values for which the proposed defense model is no longer able to identify between compromised and honest nodes.

### 3.1.4 Attacker Budget

We assume that organized adversaries compromise a certain number $M_{max}$ of the $N$ smart meters based on the *attack budget*. The fraction of compromised nodes is $\rho_{mal} = \frac{M_{max}}{N}$. Hence, $\rho_{mal}$ can be high when $N$ is small.

We assume a fixed cost $C_{attack}$ is required to compromise a smart meter. We refer to the budget as the *total cost* $TC$ of the attack, such that $TC = C_{attack} \times M_{max}$. We term the attacker's revenue as $RR$ over an attack period of $D$ days in terms monetary gain in the electricity bills.

We define $T_{BE}$ as the *breakeven time*, that is the time required for the revenue accrued from attacks to match the total cost $TC$. The tradeoffs that impact $T_{BE}$ and $BDP$ is studied in Section 6.

In certain implementations, a transformer meter checks the total inflow of power versus total outflow. If $\delta_{avg}$ is high, then a smart adversary can may compromise the corresponding transformer meter, to avoid easy suspicion. Only simple attack types like additive and deductive would require the adversary compromising the transformer meter, given that in camouflage and conflict attacks, the total reported power consumption is not affected significantly.

### 3.1.5 A Concrete Example

Suppose in an AMI facility of $N = 100$ smart meters, $M_{max} = 20$ implying $\rho_{mal} = 0.2$. The actual aggregate power consumption distribution has a mean and a standard deviation of $\mu_A = 2000$ units and $\sigma_A = 100$ units, respectively. If the amount of additive error to be introduced in the final mean is $\Lambda = 500$ units, the $\Delta_{avg}$ for each malicious node is given by $\Delta_{avg} = \frac{\Lambda * N}{M_{max}} = 2500$. Since false noise values are generated uniformly at random in the range $(\delta_{min}, \delta_{max})$, $\Delta_{avg} = \frac{\delta_{min} + \delta_{max}}{2}$. Therefore in this example,

if $\delta_{min} = 128$, as the minimum false value to be considered as attack, then $\delta_{max} = 4782$. Here $\delta_{max}$ and $\Delta_{avg}$ are a rather high value, which may easily be detected, given the nature of power consumption. However, if $\rho_{mal} = 0.4$, then to achieve the same $\Lambda = 500$, it is sufficient to have $\delta_{max} = 2372$ and $\Delta_{avg} = 1250$, which are more believable values.

The above proves that with higher $\rho_{mal}$, the adversary can afford to decrease the margin of false data to avoid getting intuitively and easily detected. Although the cost increases with higher $\rho_{mal}$, the adversary may reduce the chance of detection, and look to recover the initial cost in the long term. This is however not an option for adversaries with short term objectives. In this case, the attack revenue/payoff and the cost have to breakeven within a short time deadline. This implies very high $\rho_{mal}$ and low $\Delta_{avg}$ do not lead to a practical attack strategy, since low $\Delta_{avg}$ accrues slow attack revenue per unit time. In the experimental results, we study the trade-offs between $\Delta_{avg}$ and $\rho_{mal}$.

# 4. STATISTICAL EFFECTS OF VARIOUS ATTACKS ON AMI DATA

In this section, we study how different data falsification strategies affect the attacked mixture distribution from the actual (authentic) mixture distribution from real data gathered from 215 smart meters from a solar Village [25]. In particular, we show effects of various attack types, on the Arithmetic Mean (AM), Geometric Mean(GM) and Harmonic Mean (HM). The mathematical definitions of the different means are $AM_t = \frac{\sum_{i=1}^{N} x_t^i}{N}$, $GM_t = (\prod_{i=1}^{N} x_t^i)^{\frac{1}{N}}$, $HM_t = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_t^i}}$.

Based on the simultaneous changes between the various means, a security forensics criterion is provided to unravel the type of data falsification attack. This criterion that is based on the absolute difference (denoted by $AD$) between AM and HM of the observed mixture distribution, can also help to distinguish between a legitimate change and malicious change. Subsequently, a robust mean $\mu_R$ is derived exploiting the contrasting robustness of $AM, GM, HM$ measures to various types of attacks.

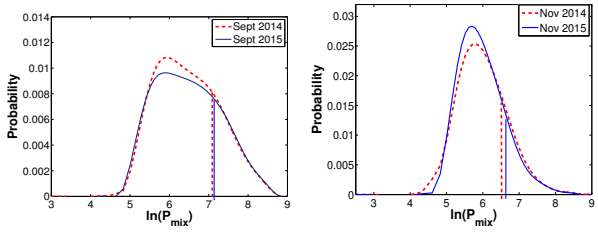
All trends on power consumption use $p^i$ values on an $ln$ scale. For comparison between legitimate and attacked data, the reference authentic distribution is called historical distribution denoted by $p_{mix}^{his}$. The attacked distribution is called the observed distribution denoted by $p_{mix}^{obs}$.

## 4.1 Investigative Comparison under Various Attacks

### 4.1.1 Authentic Data on Different Years

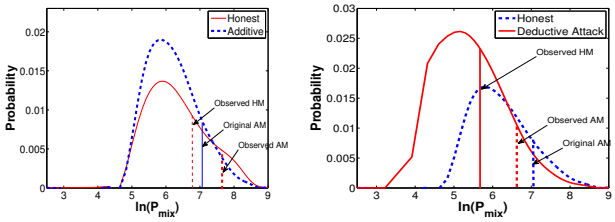
Fig. 3(a) shows the actual mixture distribution for two different years (2014 and 2015) for the month of September. We can observe, that the difference between the distributions is not large. In fact, the mean, and higher moments are very similar. This is attributed to similar coarse grained usage patterns given the weather in a particular month at the same location. Hence, power consumption at a microgrid is cyclostationary in the wide sense. The AM for 2014 and 2015 are 7.053 and 7.07 respectively. The HM for the same are 6.680 and 6.675 respectively.

However, sometimes it may happen that the same month in two different years experience varying weather conditions at certain locations. For example, winter 2015 was much warmer than winter 2014 in certain geographical locations in USA. For example in this data set, AM is 6.88 and 6.58, while the HM are 6.52 and 6.23 respectively. Such a difference is shown in Fig. 3(b). Hence, we conclude that comparison of a meter's data with the parameters of observed (current) mixture distribution ($p_{mix}^{obs}$) is equally important, as is the comparison with the historical values of power consumption.




(a) (b)

**Figure 3: Legitimate Data Comparison: (a) September (b) November**




(a) (b)

**Figure 4: Comparison: (a) Honest vs. Additive (b) Honest vs. Deductive**

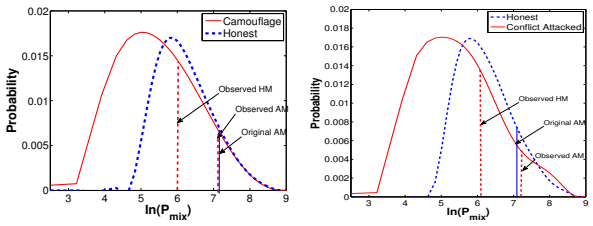
### 4.1.2 Authentic Data vs. Additive Attack

Fig. 4(a) shows the comparison between honest data set $p_{mix}^{his}$ and the same data set polluted with additive falsification, with $\Delta_{avg} = 800$ and $\rho_{mal} = 0.40$ for the month of October. Due to higher than actual power consumption reported, the observed AM is highly shifted from the original AM. Hence, when using the observed mixture distribution for anomaly detection, the observed AM is biased towards the additive false data. We observe instead that the *harmonic mean* (HM) of the observed mixture, although shifted, is closer to the original AM. Hence false readings will be located farther away from the observed HM. Hence HM is a more robust aggregate in unraveling positive outliers. Another key observation is that the absolute difference between HM and AM, given by $AD = |AM - HM|$ is higher in the attacked data set than the legitimate data set. Geometric mean (GM) is an intermediate value, but slightly closer to the AM value as compared to HM.

### 4.1.3 Authentic Data vs. Deductive Attack

Figure 4(b) shows the results for the case of deductive attacks where $\Delta_{avg} = 500$ and $\rho_{mal} = 0.40$ for October. Intuitively, the observed mixture distribution, shifts to the left, due to reporting of lower than actual consumption. As a result, the observed AM is lower than the actual AM. Nonetheless, the observed HM is even lesser than observed AM since $HM \leq AM$ is always true. Hence for deductive attacks, the observed AM is more robust than HM. However, $AD$ still increases.

Note that, the maximum possible bias introduced in the observed AM under deductive attacks is less than that of additive attacks, because the feasible margin of deductive false data is bounded by zero, because $P_t^i(rep) \geq 0$.




(a) (b)

**Figure 5: Comparison: (a) Honest vs. Camouflage (b) Honest vs. Conflict**

#### 4.1.4 *Authentic Data vs. Camouflage Attack*

Fig. 5(a) shows the effect of camouflage attacks where $\Delta_{avg} = 960$ and $\rho_{mal} = 0.40$. There is a negligible change in the AM. However, we observe that there is a shift in the HM of the observed mixture, thereby causing the resultant $AD$ to increase.

#### 4.1.5 *Authentic Data vs. Conflict Attack*

Fig. 5(b) shows the effect of conflict attacks where $\Delta_{avg}$ is 700 and 500 for additive and deductive attacks, respectively, with $\rho_{mal} = 0.40$. There is a little change in the AM. However, we observe a shift in the HM of the observed mixture. Hence, $AD$ increases.

### 4.2 Security Incident Forensics

Based on the observations from comparison between authentic versus attacked data distributions, we identified a security incident forensics criteria, based on AD and the simultaneous change/bias in the observed AM, HM, and GM that indicate presence and type of data falsification (security incident). Based on this knowledge, we calculate a robust aggregate $\mu_R$, which is less biased than the otherwise observed arithmetic mean values.

#### 4.2.1 *Detecting the Anomaly from Legitimate Change*

We study how each attack type impacts various statistical parameters and identify a criterion (see Eqn. 1)that reveals the presence and the type of attack launched. From our statistical study, we found that $AD = |AM - HM|$ could be an effective indicator for anomalies.

Fig. 6 shows the comparison of instantaneous values of $AD$ between historical (2014) and current non-attacked distribution (2015) for different years. It can be verified that under no attacks, the average value of $AD$ is about 0.45 for both years, although contextual factors may have caused AM, GM and HM to readily change over time. The lowest $AD^{min}$ and highest $AD^{max}$ values of $AD$ for two years are between 0.35 and 0.55 Hence, $AD$ is almost stable for legitimate data sets, and we call this range $AD^{norm}$

In contrast, from Fig. 7, it is easy to conclude that for all attacks scenarios, the $AD^{obs}$ is larger than $AD^{norm}$ range. This figure clearly shows that when additive, deductive, camouflage, and conflict attack samples were introduced in the current data set, starting from the 250th day of 2015, $AD$ has increased for each attack type.

$$AD^{obs} : \begin{cases} \in AD^{norm} & \text{No Falsification ;} \\ > AD^{norm} & \text{Falsification Occurred;} \end{cases} \quad (1)$$

**Table 1: Effects of Different Attacks on AD**

| Parameter | Actual | Add | Deduct | Camo | Conf |
|---|---|---|---|---|---|
| AM | 7.053 | 7.68 | 6.67 | 7.04 | 7.26 |
| GM | 6.860 | 7.35 | 6.29 | 6.65 | 6.79 |
| HM | 6.680 | 6.92 | 5.88 | 6.02 | 6.11 |
| AD = \|AM-HM\| | 0.373 | 0.76 | 0.79 | 1.02 | 1.15 |

From the above, we conclude that an authentic change in the observed distribution may cause the mean consumption to increase or decrease but $AD^{obs}$ remains the same as compared to the historical range of values $AD^{norm} = [AD^{min}, AD^{max}]$. An additive attack causes the mean consumption to increase but also causes $AD^{obs}$ to increase compared to historical values. This way a legitimate versus a malicious change can be distinguished. A deductive attack causes the mean consumption to decrease and causes $AD^{obs}$ to increase from the historical range. Similarly, camouflage and conflict attacks do not have much change in the mean consumption but causes a large increase in the $AD^{obs}$ from the normal. In this way, it is possible to detect which type of data falsification has been launched.

**Table 2: Concluding the Security Incident Type**

| AD | AM | HM | GM | Conclusion |
|---|---|---|---|---|
| Increased | Increased | Increased | Increased | Additive |
| Increased | Decreased | Decreased | Decreased | Deductive |
| Increased | Same | Decreased | Decreased | Camouflage |
| Increased | Any | Any | Any | Conflict |
| Same | Don't Care | Don't Care | Don't Care | No Attack |

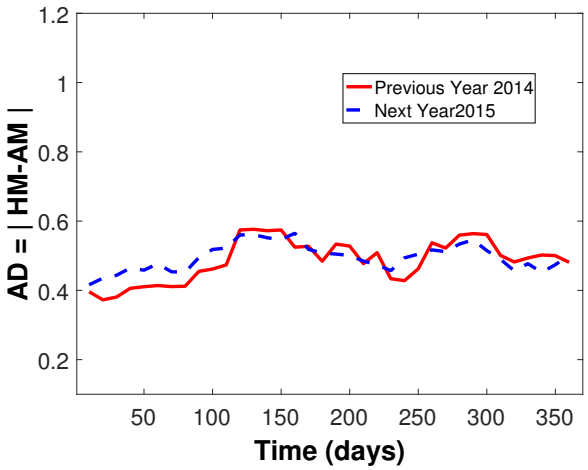



**Figure 6: AD under No Attacks**

#### 4.2.2 *A Robust Mean for Different Attacks*

The manner and extent to which different observed mean aggregates like $HM, GM$ and $AM$ get biased by different attacks is unique. We exploit this property for the calculation of robust mean. Additionally, the magnitude of the bias depends on $\Delta_{avg}$ and/or $\rho_{mal}$. Hence, an adjusted robust mean helps to get an approximate value closer to the original mean. Note that, the highest possible $\Delta_{avg}$ is lesser in deductive attacks than additive ones, because the feasible margin of deductive false data is bounded by zero. As the margins of false data or compromised fraction increases, the observed means get biased from the actual mean.

From the statistical observations, we conclude that HM is more robust than AM to the effect of additive attacks, due to slower increase in HM as opposed to AM. However, this is

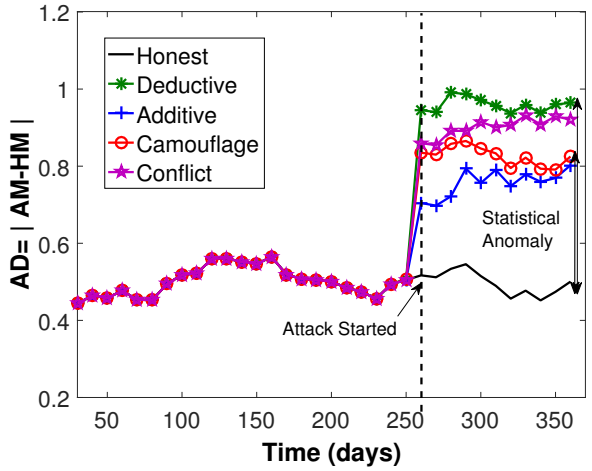



**Figure 7: AD in Observed Data: Various Attacks**

not the case for deductive attacks because of $HM \leq GM \leq AM$, causing HM to be even lesser than the already biased AM. But, $GM + AD$ is more robust than AM for deductive attacks, and results show that it is a good approximation to the actual mean. From the example in Table 2, it can be verified that for deductive attack, the robust mean $\mu_R = 6.29 + 0.79 = 7.08$ is closer to the actual mean 7.05. For camouflage attacks, AM is the most robust and hence $\mu_R$ is set as the AM. For conflict attacks, $GM$ is an intermediate robust choice as it shows a relative stability to both partially positive and negative outliers.

**Table 3: Robust Aggregate guided by Incident Type**

| Security Incident | Choice of Aggregate $\mu_R$ |
|---|---|
| Additive | HM |
| Deductive | GM+AD |
| Camouflage | AM |
| Conflict | GM |
| No Attack | AM |

# 5. AN ENTROPY BASED TRUST MODEL

We pursue a light weight supervised learning approach for defending against data falsification from compromised smart meters. A prior historical data set is considered as the authentic distribution of power consumption. From the historical data set, a *true proximity distribution* denoted as $\boldsymbol{X}_i$ for each smart meter is generated based on its reported consumption's proximity to the arithmetic mean of the authentic data set. Since the authentic historical data set is attack-free, the measure of mean is arithmetic mean (AM), denoted by $\mu$.

Then an observed data set is considered with data from spurious meters. We define $\mu_R$ as the robust mean of the observed distribution calculated as discussed in Section 4 based on the occurred security incident. The *current proximity distribution* $\boldsymbol{Y}_i$ of each smart meter $s^i$ is calculated based on the proximity of its reported consumptions to $\mu_R$. In contrast, to the historical distribution, when an attack is present, we set $\mu_R$ according to Table 3.

If the true distribution is very different from the current distribution, it is an indication that this meter is *unusually* far from the aggregate. This difference is measured as *Kullback-Leibler divergence* (also called KL Distance) which measures the *relative entropy* between the two distributions. The higher the divergence between the two distributions, the more the indication of anomalous behavior. The trust of a meter is calculated at the end of the window (in days). The total number of observations ($T$) over the window depends on how time is slotted.

## 5.1 True and Current Proximity Distributions

We introduce a binary random variable $X_i = \{0, 1\}$ for each meter $s^i$, for $i = 1, \ldots, N$, which acts as a historical reference distribution. If the historical data reported $p_t^i(rep)$ at time $t$ from meter $s^i$ falls within one standard deviation of $\mu_t$, then $X_i = 1$, else 0. Formally,

$$X_i(t) = \begin{cases} 1 & \text{if } p_t^i(rep) \in \{\mu_t \pm \sigma_t\}; \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $X_i(t)$ follows a Bernoulli distribution with parameter $r$, that is the probability of $X_i = 1$ is $r$, and the probability of $X_i = 0$ is $1 - r$.

Suppose, $S(X)$ be the variable that denotes the number of successes, that is $S(X_i) = \sum_{t=1}^{T} X_i(t)$. Let $S(X) = k$ be the observed value of the variable.

Similarly, we have a binary random variable $Y_i$ for the current distribution of each smart meter, such that the probability of $Y = 1$ is $q$ and the probability of $Y = 0$ is $1 - q$. In this case, the number of successes is denoted by a variable $R(Y_i) = \sum_{t=1}^{T} Y_i(t)$. Let $l$ be the observed value of $R(Y)$. If an anomaly has been detected through monitoring the HM, AM and AD, then $\mu_{R_t}$ is assigned accordingly, and the corresponding standard deviation $\sigma_{R_t}$ is calculated. In absence of attacks, $\mu_{R_t} = \mu_t$. Thus,

$$Y_i(t) = \begin{cases} 1 & \text{if } p_t^i(rep) \in \{\mu_{R_t} \pm \sigma_{R_t}\}; \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Intuitively, in absence of attacks the distribution of $Y$ should be very close to $X$. On the contrary, the two distributions should show a difference when an attack is present.

## 5.2 Estimating Parameters of True and Current Proximity Distributions

Next we need to estimate the parameters $r$ and $q$ for corresponding distributions $X_i$ and $Y_i$. An obvious estimate is the minimum variance unbiased estimate (frequentist), which is the sum of all successes divided by the total number of observations $T$. However, this approach may cause $r = 0, q = 0$, or $r = 1, q = 1$, for which the relative entropy (see Eqn. 9) is undefined. Moreover, frequentist probability unbiased estimator makes sense only if there is a large set of observations [13]. However, since our trust model works on a shorter horizon of time (typically on a few days or monthly basis), such approaches are improper. Hence, we need to accommodate a Bayesian approach for estimation of $r$ and $q$, so it is theoretically sound and mathematically tractable. Since the following is true for all meter's $s^i$, we drop the suffix $i$ from the notational simplicity.

First, we estimate the parameter $r$. We prove that the estimated probability $r = \frac{k+1}{T+2}$, where $k$ is the realization of the total number of successes observed. Thus $S(X) = k$ follows a binomial distribution with parameter $r$.

Hence, the probability of observing exactly $k$ successes out $T$ times, given the probability of success of each trial was $r$, is given by,

$$P(S(X) = k|r) = \binom{T}{k} r^k (1-r)^{T-k} \tag{4}$$

The Bayesian posterior estimate of $r$, based on prior $T$ observations by Bayes theorem, is given as:

$$P(X(T+1) = 1|S(X) = k) = \frac{P(X(T+1)), S(X) = k)}{P(S(X) = k)} \tag{5}$$

The denominator is the marginal probability of $P(S(X) = k)$ marginalized over all possible outcomes of $r$. Hence,

$$P(S(X)) = \int_0^1 \binom{T}{k} r^k (1-r)^{T-k} f(r) dr \tag{6}$$

Assuming conditional independence between $S(X)$, $r$ and $X_i(t+1)$ of the prior and likelihood can be solved as:

$$P(X_i(T+1)), S(X) = k) \Rightarrow$$
$$= \int_0^1 P(X(T+1) = 1|r) P(S(X) = k|r) dr \tag{7}$$

Since there is no prior information on $r$, we assume a non-informative prior such that $f(r) = 1$, for the above Eqn (6) and Eqn. (7). Plugging in Eqn. (6) and Eqn. (7) to Eqn. (5), it can shown that:

$$P(X_i(T+1) = 1|S(X) = k) = \frac{k+1}{T+2} = r \tag{8}$$

Hence, $r = \frac{k+1}{T+2}$. Similarly, $q = \frac{l+1}{T+2}$.
It can be verified that $r, q \neq 0, 1$. Hence, the logarithms of distributions $X$ and $Y$ in terms of $r$ and $q$ are always defined and exist as evident from Eqn (9).

## 5.3 Kullback-Leibler Divergence based Trust Model

We adopt the *Kullback Leibler divergence* to measure the difference between the historical distribution $X_i$ and the observed distribution $Y_i$ for a smart meter. It may be noted that $X_i$ and $Y_i$ are not consumption patterns but a trend on proximity to the aggregate. Subsequently, the KL distance is transformed into a trust value between zero and one. The trust values are fed to a generalized linear model based logit link function for linearly separable trust values that facilitate classification between compromised and honest meters through a single threshold.

The KL distance between two distributions X and Y for a smart meter $s^i$, is given by:

$$D_i(X_i||Y_i) = (1-r) \times ln\Big(\frac{1-r}{1-q}\Big) + p \times ln\Big(\frac{r}{q}\Big) \tag{9}$$

The $D_i(X||Y)$ is a positive real value. The final trust value of a smart meter $s^i$, is given by:

$$Q_i = \frac{1}{1 + \sqrt{D_i(X||Y)}} \qquad 0 \leq Q_i \leq 1 \tag{10}$$

Any classification problem such as identifying compromised meters from honest ones, require a threshold for separation. In order to ensure the efficiency of our method, our goal is to ensure that the compromised and honest meters form two clearly linearly separable clusters in terms of their trust values. However, for certain attacks, especially camouflage and conflict attacks, the distributions may not be sufficiently far from each other to ensure linear separation through a threshold.

To address this problem, we introduce a *kernelized trust metric* that maps the trust values into a higher dimension. We use a light weight two step kernel mapping function. The first step is the use of a generalized linear model (GLM) predictor (logit link function) where $Q^i$ is mapped into $W^i \in \mathbb{R}$ as follows.

$$W^i = log_2\left(\frac{Q^i}{1-Q^i}\right) \tag{11}$$

The second step is a Weibull scaling function converting $W^i$ into the final kernelized trust metric $KT^i \in [-1, +1]$:

$$KT^i = \begin{cases} 1 - e^{-|W^i|} & \text{if } W^i > 0; \\ -(1 - e^{-|W^i|}) & \text{if } W^i < 0; \\ 0 & \text{if } W^i = 0 \end{cases} \tag{12}$$

Eqn. (11), is a logit link function used in logistic regression for binary classification problems. Since, our problem is to classify malicious from honest, the corresponding link function for such response variables is a logit function.

**Remark on privacy concerns**: Note that, our defense model does not require the storage and maintainance of actual power consumption trends of each individual house. Rather, we only store the information of $X_i$, thus less privacy intrusive. So, by policy the individual $P_t^i(rep)$ need not be used and may be discarded. In particular, the data collector only knows the historical mixture distribution $p_{mix}$ and the historical private parameter $r$ of each house. The $q$ parameter is the current private parameter of each house. The challenge is resolved by depending on $p_{mix}^{his}$ rather than the individual $p_i$ distribution. Since $r$ and $q$ are compared for KL divergence, both comparison to history and comparison with current mixture distribution is achieved in a privacy preserved manner.

# 6. PERFORMANCE EVALUATION

The data set from three residential micro-grids of $N = 215$ houses, was obtained from PeCan Street Project [25], containing hourly power consumption data from a solar village near Austin, Texas. We studied some results of anomaly detection and trust model for various attacks. To display the performance of the defense models, a 30 day data from 2014 was used as a training data set. The training data set is used to derive a threshold (through K-means clustering) that linearly separates between honest and compromised labels based on their trust score. A data set for a 30 day period in 2015 is used a testing data set.

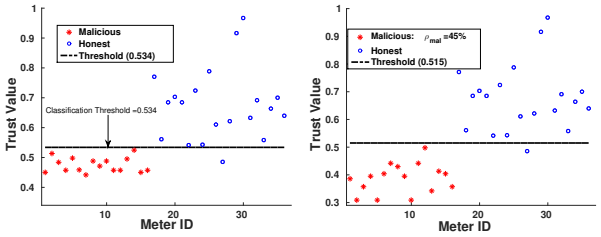
The malicious data sets were generated from the real data samples that were fed into a simulated AMI micro-grid, and $\rho_{mal}$ and $\Delta_{avg}$ were carefully chosen to avoid overfitting or underfitting. In the compromised testing data set, the resulting distributions of each smart meter is modeled as the distribution $Y$. The KL distance is calculated for each smart meter, and subsequently their trust values are plotted. Then, the threshold obtained from training is applied

## 6.1 Training Set

We use a training data set from 36 houses and use power consumption reported in 2014 for a month. We label some

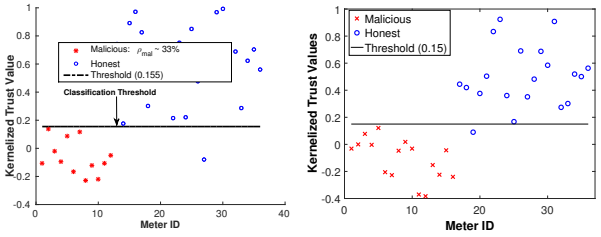
meters as compromised and alter their reported values and plot the corresponding trust values. We use these experiments to calculate a threshold that can linearly separate between compromised and non-compromised nodes. We use a logistic regression classifier to find the optimal linearly separable classifier. We choose $\rho_{mal} = 0.4$, and $\Delta_{avg} = 1024W$ which are intermediate values to prevent overfitting or underfitting. The results of training for various attack strategies are shown in Figs. 8(a) and 8(b).

Figure 9(a) shows the training sets for camouflage attacks with kernelized trust metrics bounded between $[-1, +1]$, for $\delta_{avg} = 960W$ and $\rho_{mal} = 0.45$. The kernel mapping function yield a clear separation between honest and compromised labels even for stealthy camouflage attacks. We derive the threshold as 0.155. Fig. 9(b) shows the training results for conflict attacks where $\Delta_{avg}$ is 900 and 600 for additive and deductive attacking meters respectively and $\rho_{mal} = 0.45$.




(a) (b)

**Figure 8: Training Data: (a) Additive; (b) Deductive**




(a) (b)

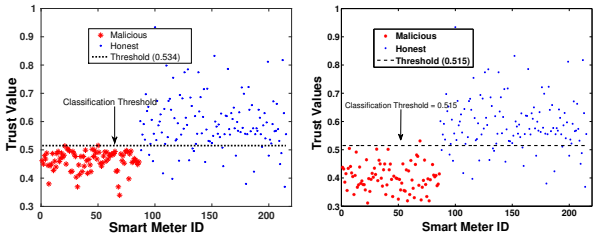
**Figure 9: Training Data: (a) Camouflage; (b) Conflict**

## 6.2 Performance with Testing Set

We use the data set from 2015 as testing set. We set $\rho_{mal} = 0.4$ and $\Delta_{avg} = 768W$. More results with different $\rho_{mal} = \{0.2, 0.3\}$ are presented in Appendix A to prove the scalability. The results for additive and deductive attacks are shown in Figs. 10(a) and 10(b). They exhibit a clear separation between honest and compromised nodes with a false alarm rate of 8.3% in the additive case, and 9.3% in the deductive case. The missed detection rate is 0% for all attack modes, and no compromised meter remains undetected.

Figure 11(a) shows the results for the testing set for $\delta_{avg} = 880W$, demonstrating a clear difference between honest and compromised nodes. The false alarm rate in this case is 11.6%.

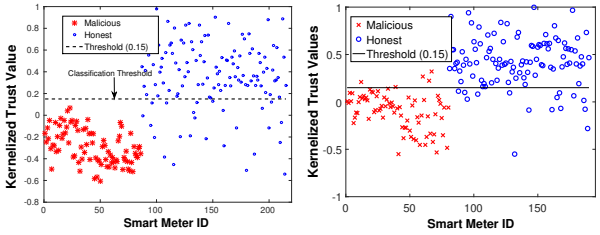
For conflict attacks, about 48% of the total meters are compromised, with additive attack of $\delta_{avg} = 1300W$, while the deductive attacks with $\delta_{avg} = 900W$. This is a case which considers random attacks from unorganized rivals.




(a) (b)

**Figure 10: Testing Data: (a) Additive; (b) Deductive**

From Fig. 11(b), we show that our approach works with a high detection rate of 98% and the false alarm rate is about 7%. All the testing results, show improvement from most existing works in [2].

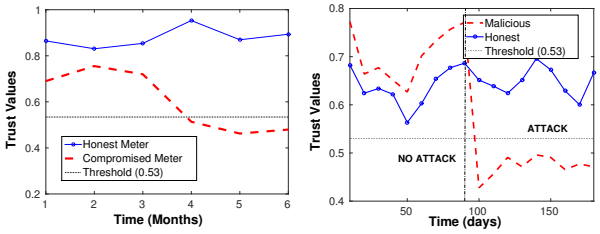



(a) (b)

**Figure 11: Testing Sets: (a) Camouflage; (b) Conflict**

## 6.3 Trust Values over Time

Figs. 12(a) and 12(b) show the trust value comparison between an honest meter and a compromised meter over time. Fig. 12(a), shows trust values calculated every 30 days while Fig. 12(b) shows them for every 10 days. The first 90 days are attack free, and hence trust values are above the threshold. After the 90th day, the attack starts, and a decrease in the trust of compromised meter is clear while the honest meter's trust is unaffected.




(a) (b)

**Figure 12: Trust Propagation Over Time: (a) Every 30 days; (b) Every 10 days**

## 6.4 Breakdown and Breakeven Point Analysis

We model the attacker revenue $RR$ over an attack period of $D$ days in terms monetary gain in the electricity bills as:

$$RR = \frac{\Delta_{avg} \times M_{max} \times \eta \times D \times E}{1000} \quad (13)$$

where $\Delta_{avg}$ is the average attack margin, $\eta$ is the number of reports a day, and $E$ is the per unit (KW-Hour) cost of electricity in dollars.

Recall that the *breakdown point* occurs for some $\Delta_{avg}$ and $\rho_{mal}$ such that the proposed model is no longer able to distinguish between compromised and honest nodes because the average trust values of compromised meters are higher than honest ones. We study the existence of such breakdown points, the feasibility of the adversary achieving the breakdown point and associated cost benefit analysis. This analysis can be used by the insider attackers to accordingly design their attack strategy to evade detection.

The breakdown point could be achieved with very high or low margin $\Delta_{avg}$ given $\rho_{mal}$. Very low margins make short term attacks impossible and very large margins make attacks very obvious. Therefore, low margins make sense for long term attacks, but require either a large $\rho_{mal}$ or a long time duration to be effective.

Figure 13 shows different breakdown points for $\rho_{mal}$ values 0.10, 0.25, 0.4 and 0.6 for additive attacks over various $\Delta_{avg}$ values. Two break down points exist only for $\rho_{mal} = 0.6$, which suggest that to evade detection with lower margin of false data, the adversary has to compromise a large number of nodes, thus increasing its cost. For lower $\rho_{mal}$, the attacker cannot evade detection, unless $\Delta$ is very small, which in turn rules out short term attacks and increases breakeven time ($T_{BE}$) significantly.

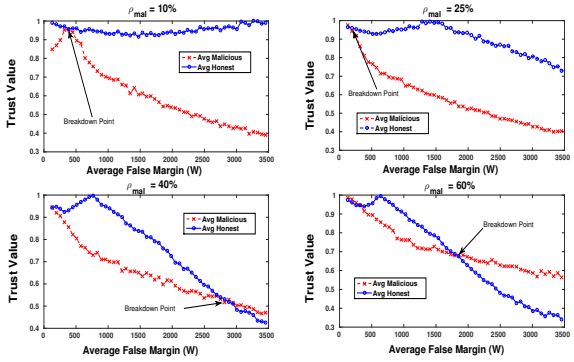



**Figure 13: Sensitivity Analysis over $\Delta_{avg}$: Additive**

The breakdown points for deductive and camouflage attacks for different $\rho_{mal}$ values are shown in Figs. 14 and 15.

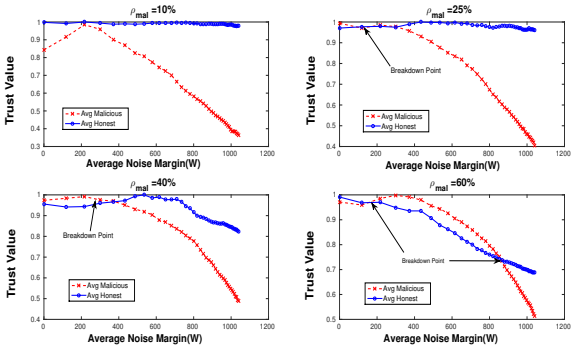



**Figure 14: Sensitivity Analysis over $\Delta_{avg}$: Deductive**

Table 4 numerically shows that for a very low margin $\Delta_{avg} = 256W$ and $\rho_{mal} = 0.5$, the adversary needs 22 months to recover its initial investment and start to gain profit. This acts as deterrent to implement such a strategy although it may evade detection.

Another aspect shown in the plots and Table 4 is that the breakdown point and attack evasion could be achieved with higher $\rho_{mal} \geq 0.42$ and a simultaneous higher margin $\Delta_{avg}$ of false data. However, our model ensures the breakdown happens only at very high levels of $\Delta_{avg}$. This is intuitively detectable, since the power consumption has a thinner tail, and most of the power consumptions are below $2500W$.

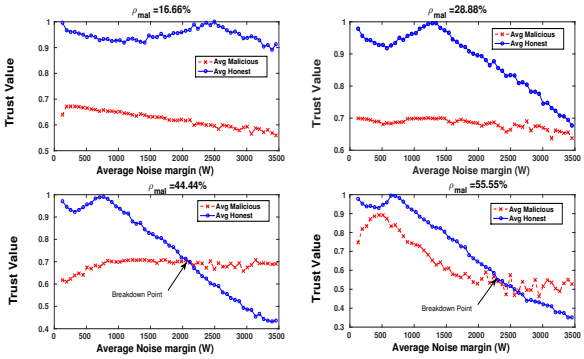



**Figure 15: Sensitivity Analysis over $\Delta_{avg}$: Camouflage**

In Table 4 we also observe that when $\Delta_{avg} = 3328W$ and $\Delta_{avg} = 3072W$, the breakdown point is achieved at lower cost of $\rho_{mal} = 0.42$, and breakeven duration is also significantly smaller. The problem for the adversary is the very high margin and fraction of compromised meters, which makes it easily detectable and cost inefficient. This observation shows that although our defense model has breakdown points, most of these correspond to strategies which are hardly convenient for the adversary.

**Table 4: Breakdown and Breakeven analysis**

| $\Delta_{avg}$ | $\rho_{mal}$ | $TC$ | $RR$ | $T_{BE}$ |
|---|---|---|---|---|
| 256 | 0.5 | 9000 | 398.13 | 22 |
| 512 | 0.69 | 12500 | 1105.9 | 11 |
| 768 | 0.80 | 14500 | 1924.3 | 7.53 |
| 1024 | 0.86 | 15500 | 2742.6 | 5.6 |
| 1280 | 0.72 | 13000 | 2875.3 | 4.52 |
| 1536 | 0.61 | 11000 | 2919.6 | 3.76 |
| 1792 | 0.58 | 10500 | 3251.4 | 3.22 |
| 2048 | 0.53 | 9500 | 3361.9 | 2.82 |
| 2304 | 0.50 | 9000 | 3583.1 | 2.5 |
| 2560 | 0.47 | 8500 | 3760.1 | 2.26 |
| 2816 | 0.44 | 8000 | 3892.8 | 2.05 |
| 3072 | 0.42 | 7500 | 3981.6 | 1.88 |
| 3328 | 0.42 | 7500 | 4313.0 | 1.73 |

## 7. CONCLUSION

In this paper we presented a taxonomy of various data falsification strategies in AMI micro-grids, as may be devised by powerful and organized adversaries such as rival nation states, business competitors, etc. rather than individual selfish customers only. We proposed statistical anomaly detection and forensics technique to identify presence of various attacks and a trust model based on Kullback-Leibler divergence to identify the compromised smart meters. Our analysis on real data sets shows that both the margin of false data and the fraction of compromised nodes play a key role in understanding the limits of a distributed detection scheme. We also studied some strategies that could be employed by attackers to escape detection and the cost benefit analysis of such strategies.

**Acknowledgment**: The work is partially supported by the NSF grants under award numbers CNS-1545037, CNS-1545050 and DGE-1433659.

# 8. REFERENCES


[1] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, "Non-technical loss detection using state estimation and analysis of variance", *IEEE Trans. on Power Systems*, 28(3):2959-2966, Aug. 2013.

[2] R. Jiang , R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-Theft detection issues for advanced metering infrastructure in smart grids", *Tsinghua Science and Technology*, 19(2):105-120, April 2014.

[3] P. Jokar, N. Arianpoo, and V. Leung, "Electricity theft detection in AMI using customers' consumption patterns", *IEEE Trans. on Smart Grid*, 7(1):216-226, Jan. 2016.

[4] T. Koppel, "Lights Out: A Cyberattack, A Nation Unprepared, Surviving the Aftermath", *Crown Publishers, New York*, 2015.

[5] C.-H. Lo and N. Ansari, "CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid", *IEEE Trans. on Emerging Topics in Computing*, 1(1):33-44, 2013.

[6] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Trans. on Parallel and Distributed Systems* 23(9):1621-1631, Sept. 2012.

[7] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure", *Proc. of Critical Information Infrastructures Security*, Springer-Verlag, pp. 176-187, Sept. 2009.

[8] S. McLaughlin, B. Holbert, S. Zonouz, and R. Berthier, "AMIDS: A multi-sensor energy theft detection framework for advanced metering infrastructures", *IEEE SmartGridComm*, pp. 354-359, Nov. 2012.

[9] D. Mashima and A. Alvaro, "Evaluating electricity theft detectors in smart grid networks", *Springer Intl. Workshop on Recent Advances in Intrusion Detection*, pp. 210-229, Sept. 2012.

[10] R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, "A survey on advanced metering infrastructure", *Elsevier Journal of Electrical Power & Energy Systems*, 63:473-484, Dec. 2014.

[11] A. Rad and A.L. Garcia, "Distributed internet-based load altering attacks against smart power grids", *IEEE Trans. on Smart Grids*, 2(4):667-674, Dec. 2011.

[12] R. Sevlian and R. Rajagopal, "Value of aggregation in smart grids", *IEEE SmartGridComm*, pp. 714-719, Oct. 2013.

[13] Y.L. Sun, W. Yu, Z. Han, K.J. Ray Liu, "Information Theoretic Framework of Trust Model and Evaluation for Ad Hoc Networks",*IEEE Journal on Sel. Areas in Communications*, 24(2):305-317, Feb. 2006.

[14] W. Wang and Z. Lu, "Cyber security in smart grid: Survey and challenges", *Computer Networks*, 57(5):1344-1371, Apr. 2013.

[15] E. Werley, S. Angelos, O. Saavedra, O. Cortes, and A. Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems", *IEEE Trans. on Power Delivery*, 26(4):2436-2442, Oct. 2011.

[16] J. Xia and Y. Wang, "Secure key distribution for the smart grid", *IEEE Trans. on Smart Grid*, 3(3):1437-1443, Sept. 2012.

[17] W. Yu, D. Griffith, L. Ge, S. Bhattarai and N. Golmie, "An integrated detection system against false data injection attacks in the Smart Grid, *Security and Commun. Networks*, 8(2):91-109, Jan. 2015.

[18] https://skyvisionsolutions.files.wordpress.com/2014/08/utility-smart-meters-invade-privacy-22-aug-2014.pdf

[19] http://www.nytimes.com/2009/12/14/us/14meters.html?ref=energy-environment&_r=0

[20] https://www.maximintegrated.com/content/dam/files/design/technical-documents/white-papers/smart-grid-security-recent-history-demonstrates.pdf

[21] https://energy-solution.com/2015/01/29/enabling-automated-demand-response-pge-dras/

[22] https://www.smartgrid.gov/files/The_Smart_Grid_Promise_DemandSide_Management_201003.pdf

[23] http://blog.comverge.com/intelligent-energy-management/does-ami-have-what-it-takes-for-demand-response/

[24] https://www.whitehouse.gov/sites/default/files/microsites/ostp/nstc-smart-grid-june2011.pdf

[25] https://www.smartgrid.gov/project/pecan_street_project_inc_energy_internet _demonstration.html

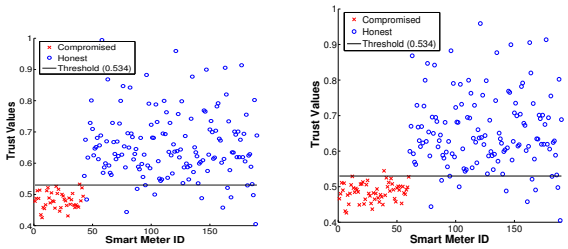
[26] https://skyvisionsolutions.files.wordpress.com/2014/08/utility-smart-meters-invade-privacy-22-aug-2014.pdf

[27] http://energy.gov/sites/prod/files/oeprod/Documentsand Media/14-AMI_System_Security_Requirements_updated.pdf


# APPENDIX

## A. TESTING SETS: LOWER FRACTION OF COMPROMISED NODES

Figs. 16(a) and 16(b) prove that our approach works for testing sets with $\rho_{mal}$ values that very different from the training sets.

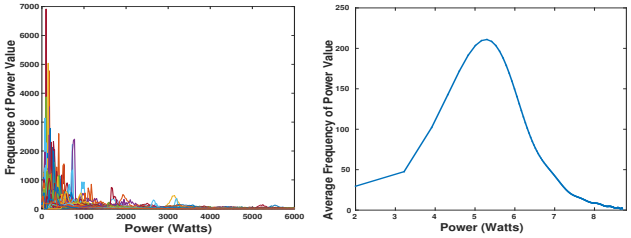



(a) (b)

**Figure 16: Testing Set for Additive: (a) $\rho_{mal} = 20\%$ (b) $\rho_{mal} = 30\%$**

## B. POWER CONSUMPTION DATA: DIFFERENT REGION

To prove that the nature of power consumption studied is generic, we show in Figs. 17(a) and 17(b) that the nature of power consumption is also similar for a different AMI data set as used in [3].




(a) (b)

**Figure 17: Power Consumption Behavior: Different Data Set**