Detection of False Data Injection in Smart Water Metering Infrastructure

Ayanfeoluwa Oluyomi[a], Shameek Bhattacharjee[b], and Sajal K. Das[a]
[a]Department of Computer Science, Missouri University of Science and Technology, Rolla, USA
[b]Department of Computer Science, Western Michigan University, Kalamazoo, USA
E-mail: aoonzb@mst.edu, shameek.bhattacharjee@wmich.edu, sdas@mst.edu

***Abstract*—Smart water metering (SWM) infrastructure collects real-time water usage data that is useful for automated billing, leak detection, and forecasting of peak periods. Cyber/physical attacks can lead to data falsification on water usage data. This paper proposes a learning approach that converts smart water meter data into a Pythagorean mean-based invariant that is highly stable under normal conditions but deviates under attacks. We show how adversaries can launch deductive or camouflage attacks in the SWM infrastructure to gain benefits and impact the water distribution utility. Then, we apply a two-tier approach of stateless and stateful detection, reducing false alarms without significantly sacrificing the attack detection rate. We validate our approach using real-world water usage data of 92 households in Alicante, Spain for varying attack scales and strengths and prove that our method limits the impact of undetected attacks and expected time between consecutive false alarms. Our results show that even for low-strength, low-scale deductive attacks, the model limits the impact of an undetected attack to only €0.2199375 and for high-strength, low-scale camouflage attack, the impact of an undetected attack was limited to €1.434375**

***Index Terms*—Smart Water Meter, Anomaly detection, False data injection, Smart Water Metering Infrastructure.**

I. Introduction

Smart water meters (SWMs) installed at consumers' houses are key components of a water distribution network (WDN) in a smart city. SWMs periodically capture the water usage data and send it to the utility over a wireless network. Such data allow for tasks such as determining peak periods and conservation techniques, detecting leaks, and automating billing and pricing. As a result, it is crucial that the integrity of SWM data is preserved. [12].

Motivation It has been reported that the SWM data are vulnerable to integrity attacks. In [13], a 12 billion gallon water theft has been disclosed in California since 2013. This type of attack is termed deductive [3] since the aim is to falsely report reduced water usage. Given water supply is a lifeline of a smart city, it is a target of organized adversaries, such as business competitors, utility insiders, and rival nation-states.

Related Works: Examining the existing literature, we observe that most of the works in WDN security focus on either the water treatment plants or distribution systems of waterways, instead of SWM applications. In [1], an anomaly-based attack detection is proposed in water treatment plants. The work in [5] detects SCADA cyberattacks in WDN using physics (constraints on pressure readings).

In [4], machine learning approaches were utilized in the WDN attack detection model that made use of remote sensing data (such as pipe flow sensor, nodal pressure sensor, tank water level sensor, and programmable logic controllers) In [2], the attack detection in a WDN is based on a Kalman filter estimating the system dynamics evolution.

As mentioned, data falsification on SWMs at end users' homes has not received adequate attention. The fact that it is a critical issue that can create havoc on the WDN, motivates us to explore this important topic.

From a technical point of view, the majority of existing works in water systems are physics-based models that work well for water treatment and distribution because the dynamics are mostly affected by physical principles. However, it is extremely difficult to apply such models to SWM attack detection, since the data patterns are affected by complex user behavioral relationships. Therefore, the knowledge of a stable in-control process means is typically required for the accurate implementation of a cumulative sum (CUSUM) control chart-based approach, which exhibits poor performance.

Contributions This paper focuses on attacks that target the SWM infrastructure to launch data falsification attacks on the water consumption data. First, we establish the data falsification threat landscape in SWM in terms of attack types, strengths, scales, and strategy, and how they negatively impact the operations of a SWM. Then, we identify unique behavioral properties of water usage from the houses. We further show that the problem is suitable for a time series anomaly detection approach. Specifically, using Pythagoras means we model the underlying structure of benign data from SWM and use it to derive a data-driven invariant that is highly stable in the absence of attacks. At the same time, using the properties of consumer water usage, we regulate the time granularity at which the invariant is calculated. This results in high stability in the absence of attacks, but large deviations in the invariant, as and when attacks from a subset of smart water meters are launched. The detection model consists of a two-tier approach. The first tier determines the ratio of Harmonic Mean (HM) to Arithmetic Mean (AM) of the (time series) water usage. With a goal to reduce false alarms without increasing the impact of an undetected attack, the second tier which is the sum of residual under curve (RUC) was used.

The rest of the paper is organized as follows. Section II introduces the system and threat models, while Section III discusses the proposed framework. Experimental results are described in Section IV and Section V concludes the paper.

II. System and Threat Models

A. SWM data flow architecture

Fig. 1 illustrates smart water metering (SWM) infrastructure topology. The SWM installed in a house senses the water usage, which is transferred periodically to a data concentrator in the geographical area via a wireless link. The data concentrator forwards the water usage data from each smart water meter to the data center of the utility company. For more information about the SWM deployment, please see [7].

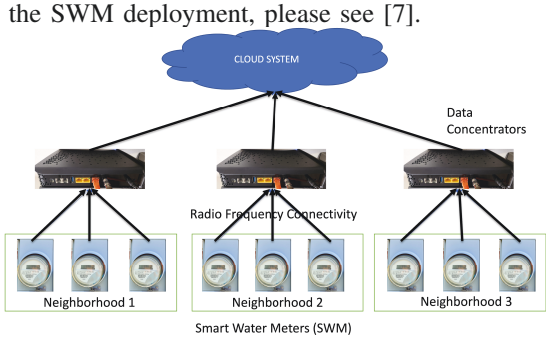


Fig. 1: Showing the flow of the water usage in WDN.

B. Threat Model.

The threat model followed in this work is similar to the one presented in [3], this will be summarized in this section False data injection (FDI) attacks on SWM imply an attacker falsifying the water usage data ultimately received by the data center. Data falsification may occur at an individual meter or a compromised data concentrator, where the attacker can intercept, view, and alter multiple smart meter data intelligently (an example of a camouflaged attack discussed later). Individual meters can be tampered with by cyber or physical attacks, e.g., transduction attack [11] that changes the accurate conversion of analog sensor signals to the digital sensor output.

In this paper, we assume an organized adversary has the resources to launch FDI attacks on multiple SWMs at a time. Since real attack data are not readily available, we generated malicious data samples having various attack parameters with the aim of an unbiased evaluation of our detection model.

Let $W_t^i(act)$ be the actual water consumption of a house i at time t, while W_t^i is the reported value. Without attacks, the $W_t^i(act) = W_t^i$, while under attacks, the reported value is perturbed by a false margin δ_t. The δ_t values are sampled from a strategic distribution whose mean depends on the intended severity of impact by the attacker and stealth level, which is discussed under *attack strength*.

Attack Types: denote the way the data is perturbed which depends on the adversary's intended nature of impact:

Deductive: Here the attacker reports a reduced amount of water usage which leads to the consumer paying less amount but will lead to a loss of revenue for the utility company. $W_t^i = W_t^i(act) - \delta_t$. This can also prevent the detection of leaks that cause water usage to be increased because the reduction prevents the utility to notice the increase in a timely manner.

Camouflage: The attacker divides the compromised meters into two equal sets and launches an additive attack (reporting an increased amount of water usage) on one set while launching a deductive attack on the second set. This favors one set of customers at the expense of the other set. This is not easily detectable as the mean consumption as seen by the utility company remains unchanged. $W_t^i = W_t^i(act) - \delta_t \quad \& \quad W_t^j = W_t^j(act) + \delta_t$.

Attack Strength is a variable that controls the extent of perturbation in the FDI. Each δ_t is randomly generated between a minimum value δ_{min} and maximum value δ_{max}, with a strategic δ_{avg} that is the average margin of false data for each compromised house. This is then subtracted and/or added to the water usage of the compromised nodes based on the attack type. The δ_{avg} is varied by an adversary depending on the amount of damage it wants to inflict.

Attack Scale An adversary may compromise a certain number of M out of N possible smart water meters, and this fraction of compromised meters is denoted as $\rho_{mal} = M/N$ and is treated as an attack variable. ρ_{mal} is a percentage of the total number of houses. The houses attacked are randomly gotten. These variables δ_t and ρ_{mal}, are taken so that one complements the other, i.e., as δ_{avg} increases (i.e., the false margin increases), the ρ_{mal} reduces and vise versa. This is to inflict an approximate amount of damage.

Attack Strategy An adversary can use a pulse attack strategy to launch periodic attacks on compromised SWMs. E.g., from hour t to $t+5$, there may be no attack while from hour $t+6$ to $t+10$, there is an attack. The non-continuous nature of this attack is hard to detect because it will not create a continuously large false margin (or large δ_{avg} that can easily catch the utility company's attention as water is not always expected to be low or high).

III. Proposed Framework

A. Dataset Description

We experimented with a dataset collected from Hellenic Data Service [6]. It is a deployment of water usage for 92 households over a year period (March 2016 - February 2017). We divided them into training, cross-validation, and testing sets, whose details are given later. Our approach is based on identifying the threshold limits of benign behavior from the training data using a data-driven invariant from the process data. We use the cross-validation set to understand the best hyperparameter choices that give the best performance with a small set of simulated attacks. Finally, the testing set contains unseen attacks and we verify the extent to which the thresholds of the invariant can detect the attacks.

B. Modelling the Water Consumption Behavior

During a 24-hour time frame, water is not expected to be used at all times, (e.g. during nighttime, vacation, or office hours) as a result, the water usage dataset contains a large number of correlated zero values. Similarly, we found that sharp increases in water usage are also correlated e.g. early hours of the day almost in all houses. While there is a correlation, the actual mean usage on a certain time slot of the day varies greatly over time. Hence, metrics like moving average or mean will be greatly affected by these changes as the instability can be seen in figure 2.

Examining the characteristics of the behavioral patterns, we believed that a previous theory of anomaly detection proposed in [3] but in the context of electric grids, may be transferable in this context, due to the identified parallelism between the challenges of modeling benign behavior from IoT sensing data from smart living applications. In this paper, we first perform a transfer learning of the anomaly detection criterion from the previous theory and modify the way we define and regulate time window granularity over which the anomaly detection invariant must be calculated and maintained, apart from the appropriate power transformation parameter, and finding the quantile regression weights which affect the threshold of benign behavior.

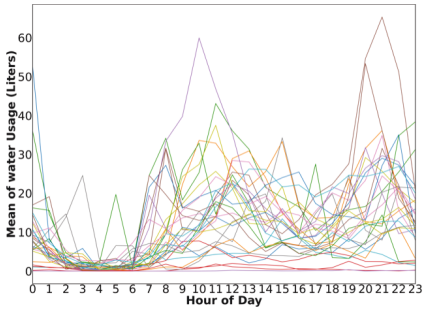


Fig. 2: Mean of water Usage for corresponding hours

C. Invariant Generation

Overview Our approach is inspired by the theory that originates from [3] which proposed a lightweight theoretical framework for the detection of anomalies in the electricity metering network. We found that the SWM behaviors can be transformed to adhere to the properties that enable us to transform. The framework consists of the design of an invariant based on Pythagoras means after which a safe margin (threshold) is determined which serves as the tier 1 stateless detection metric. Tier 2 is based on the sum of the residual distance between the observed ratio and the safe margin. In [3] the ratio of harmonic mean to arithmetic mean has been shown to be stable for smart grid and it is also shown that an attack in the data will cause the ratio to change, thus this work aims to investigate how to apply this detection model to SWM in WDN.

1) Data Distribution and Boxcox Transformation:: An investigation into water usage shows that the distribution follows an extremely left-skewed shape (a log-normal distribution). This is expected because of the properties of water usage discussed in section A above. Boxcox transformation was used to convert it from log normal to normal distribution. This step is necessary to transform the water usage to the lower portion of the real number axis (which in turn increases the sensitivity to any deviation from the normal) and also some known statistical properties of parametric estimation are easily applicable to a normal distribution.

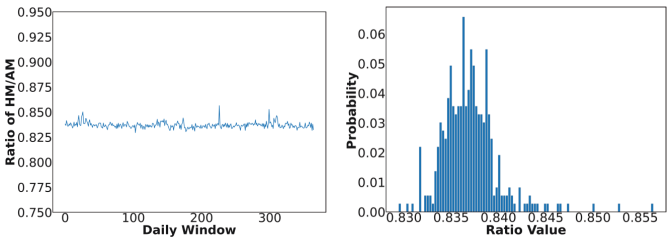
2) Time Series Invariant: Let $W_t = W_t^1, ..., W_t^N$ denote the water usage in box cox transformed scale, from N households at a time slot t, (t is slotted hourly), then the Harmonic Mean (HM) and Arithmetic Mean (AM) on a particular timeslot t is denoted by

$$HM_t = N/\left(\sum_{i=1}^{N} \frac{1}{W_t^i}\right) \qquad AM_t = \frac{\sum_{i=1}^{N} W_t^i}{N}. \quad (1)$$

The HM and AM were first calculated for all the houses for each time slot t over a time window T. Each T is composed of 24-time slots that represent a particular day in the dataset, i.e. the T-th day of the year. Then the Ratio of HM and AM is given as

$$Q^r(T) = \frac{\sum_{t=1}^{24} HM_t(T)}{\sum_{t=1}^{24} AM_t(T)} \quad (2)$$

Where $0 \leq Q_r(T) \leq 1$ due to the known Pythagorean mean inequality, $HM \leq AM$ The stability of the ratio can be seen in Fig. 3a making it suitable for anomaly detection. Figure 3b shows the probability distribution for the ratio and a right tail can be observed which informs that outliers exist in the ratio data which can lead to lots of false alarms to prevent this, a novel design of the system identification phase for tier one was proposed as discussed in section III-D.



(a) HM/AM Ratio (b) Ratio Distribution

Fig. 3: Stability and Distribution of the Ratio

D. Safe Margin Generation

Stateless residuals track the difference between the range of expected value(s) of the invariant at a time window T (called safe margin), versus the actual invariant value at a time window T. In [3], it was shown that it could be used to reduce false alarms while not increasing missed detection in the context of smart meters.

However, the behavior of a smart water system is different compared to smart meters. Hence, we need to find an intelligent way of finding the range of expected value(s) of the invariant, which we do by temporal reasoning. The range of expected value requires us to find some measure of location and scale parameters of the ratios at a given time context.

1) Expected Value of Invariant: We define T_ϕ as a unique day in a week, i.e. T_ϕ can be Sunday through to Saturday. Therefore, ϕ varies from 1 to 7; 1 corresponds to Sunday and 7 corresponds to Saturday. We define the $Q^r(T_\phi^h)$ to represent the average of ratio values at each corresponding T_ϕ (i.e. of the same ϕ) across the training data. For this work, the 7 months of the training set is approximately 30 weeks, so h can take values from 1 to 30 for each ϕ).

The average of all T_ϕ^h through the training data was taken as $Q^r(T_\phi^h)$. For example, if there are exactly 30 weeks (that is 30 Sundays, $\cdots$ 30 Saturdays), then the summation is taken for each ϕ, i.e., T_ϕ^h, where $h = [1, 30]$). This summation was then

divided by the total number of occurrences of the unique T_ϕ which is represented with μ_ϕ^h (for the example above, $\mu_\phi^h = 30$ for each ϕ). This mean is represented by $Q^r(T_\phi^h)$.

As a result, there will be a total of 7 distinct values of $Q^r(T_\phi^h)$ that will form the margin of each day of the week. For example for a Monday, the HM/AM ratio for all the days of the historical data was first obtained. The sum of the ratio for all Mondays in the historical data was obtained, this was divided by the number of times Mondays appear in the historical data. This was done considering the observed premise that daily water usage patterns are correlated to the day of the week.

2) Safe Range of the Invariant: Now we need to find a range around the expected value of $Q^r(T_\phi^h)$ (location parameter) that indicates a safe margin of operation. We use a scalar factor ϵ of the median absolute deviation of the entire ratio set $MAD(Q^r)$ as a measure of spread in the benign data.

Therefore, the $\epsilon.MAD(Q^r)$ was used to build the safe margin around the observed instantaneous ratio as shown in Equations (3) and (4). This is gotten by calculating the median absolute deviation (MAD) of the different ratios obtained from the whole training set, i.e., MAD across the ratios of the training set. ϵ is parameterized as $\epsilon \in (0, 3)$ with a step size of 0.02. The MAD was used because it is known to be more robust in handling outliers than standard deviation (SD) [8]. It was also seen experimentally to minimize the upper and lower bounds than SD. The upper limit $\Gamma_{high}(T)$ and lower limit $\Gamma_{low}(T)$ of the safe margins are defined as:

$$\Gamma_{high}(T) = Q^r(T_\phi^h) + \epsilon.MAD(Q^r) \tag{3}$$

$$\Gamma_{low}(T) = Q^r(T_\phi^h) - \epsilon.MAD(Q^r) \tag{4}$$

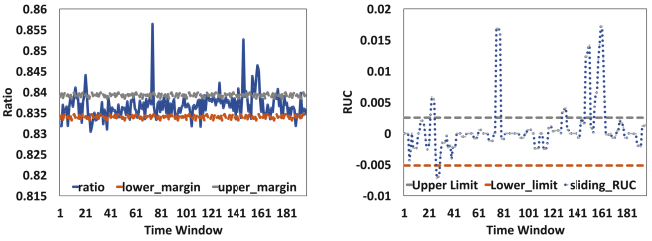
$$Where \quad Q^r(T_\phi^h) = \frac{\sum_{i=1}^{h} Q^r(T_\phi)}{\mu_\phi^h}. \tag{5}$$

The ratio was obtained for all days throughout the historical data (7 consecutive months). Given time window T denotes a day of the week, then the arithmetic means for the ratio for each unique day T of the week through the historical data was used as the safe margin. As a result, there are only seven distinct values (number of days of a week) for the safe margin. This is represented with $Q^r(T_\phi^h)$. $\sum_{i=1}^{h} Q^r(T_\phi^h)$ = summation of the ratio for each unique T across the historical data, $\mu^h\phi$ = the number of occurrences of each unique T (with the same ϕ) across the historical data.

Larger ϵ values signify that there will be larger safe margins which can result in more missed detection and smaller ϵ values can result in more false alarms, as a result, a trade-off is necessary for selecting ϵ such that there will be a lowered false alarms while ensuring that the missed detection is not on the high side. Section III-I discusses hyperparameter learning.

E. Stateless and Stateful Residuals

For the second tier of detection, the sum of Residual distance RUC(T) was used. This calculates the difference (also known as residual) between the observed ratio and the safe margins. The sum of this difference is taken over a sliding time frame. This is done by first determining the signed residual



(a) Ratio (b) RUC

Fig. 4: Detection model and learning phase for the training set

distance $\nabla(j)$ between the observed ratio and the safe margin for that time window. This is given by:

$$\nabla(T) = \begin{cases} Q^r(T) - \Gamma_{high}(T), & \text{if } Q^r(T) > \Gamma_{high}(T); \\ Q^r(T) - \Gamma_{low}(T), & \text{if } Q^r(T) < \Gamma_{low}(T); \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The sum of the residual is calculated over a sliding frame length of past FL days. The RUC(T) is given by:

$$RUC(T) = \sum_{j=T-FL}^{T} \nabla(j). \tag{7}$$

F. Thresholds of Detection (Standard Limits):

The threshold for the RUC(T) was determined using Algorithm 1 where τ_{max} and τ_{min} are the upper and lower limits. Cost c and Penalty p were used to prevent overfitting and underfitting respectively. Weight ω was hyper-parameterized along with the FL with the cross-validation set to get the optimal value that minimizes the false alarm and simultaneously minimizes the impact of the attack. For τ_{max}, the algorithm searches among all non-zero positive RUC(T), this is what is shown in Algorithm 1 below. τ_{min} searches among the non-zero negative values in RUC(T).

Algorithm 1: Determining the Upper Limit τ_{max}

```
Data: List of τ: [τ]
Result: τ_max
for RUC(T), τ do
    if RUC(T) > 0 then
        if (RUC(T) < τ then
            c_max = |τ − RUC(T)| / ω;
            ℂ ← c_max
        else
            p_max = ω|RUC(T) − τ|;
            ℙ ← p_max
        end
    end
end
τ_max = argmin_τ |Σ_ℂ(c_max) − Σ_ℙ(p_max)|
```

G. Detection Process during Testing

With the safe margin defined in section III-D, there will only be 7 distinct values of $Q^r(T_\phi^h)$ and with the optimal $\epsilon.MAD(Q^r)$, 7 distinct values of $\Gamma_{high}(T)$ and $\Gamma_{low}(T)$ were obtained and these represent the seven days of the week. The instantaneous ratio gotten from currently observed water usage is represented by $Q^r(T_\phi^c)$. This is defined as the currently observed ratio for a particular day in a week. Each instantaneously observed ratio for a day of the week is checked with the safe margins of the same day (i.e., both the current

ratio and the historical data will have the same ϕ) For example, a current ratio observed on a Monday will be checked with the safe margins of Monday (gotten from the historical data). If $Q^r(T^c_\phi)$ is within the safe margin, then it is counted as normal water usage for that day, else, an anomaly is flagged. The equation is given below.

$$Q^r(T^c_\phi) : \begin{cases} \in [Q^r(T^h_\phi)) \pm \epsilon.MAD(Q^r)] & \text{Normal ;} \\ \notin [Q^r(T^h_\phi)) \pm \epsilon.MAD(Q^r)] & \text{Suspected;} \end{cases} \tag{6}$$

Where T^c is the current time window and in this case, the current day, $Q^r(T^h) \pm \epsilon.MAD(Q^r)$ is the safe margin as $Q^r(T^h) + \epsilon.MAD(Q^r)$ represents $\Gamma_{high}(T^h)$ and $Q^r(T^h) - \epsilon.MAD(Q^r)$ represents $\Gamma_{low}(T^h)$.

$$RUC(T^c) : \begin{cases} \in [\tau_{\min},\ \tau_{\max}],\ \text{No Anomaly;} \\ \notin [\tau_{\min},\ \tau_{\max}],\ \text{Anomaly.} \end{cases} \tag{9}$$

The second tier of detection, the $RUC(T_c)$ is the residual under curve for the currently observed time window, while τ_{min} and τ_{max} is the lower and upper threshold for the RUC. This is needed to reduce the false alarms seen with tier 1.

H. Performance Evaluation Metrics

Expected time Between false alarm E(T_{fa}): False alarm rate is not the best for security evaluation since it largely depends on the time duration of the study and the granularity of the time series detector. As proposed by [9] expected time between false alarms $E(T_{fa})$ should be used instead to account for the base rate fallacy. The higher the $E(T_{fa})$, the better the model is. Mathematically, it is defined as $E(T_{fa}) = \frac{\sum_1^{nFA} T_E}{nFA}$ Where nFA is the number of false alarms and T_E is the duration between two consecutive false alarms.

Impact of an Undetected attack (I_u): We model this as the revenue damage per hour for an undetected attack. For the purpose of a security evaluation, this is a better metric given the attacker who knows our approach may bypass detection. Mathematically, it is quantified as $I_u = RR/24$ where RR is the total revenue loss, such that $RR = \delta_{avg} \times M \times \eta \times E$, where η is the number of days of attack before the first detection, M = number of houses compromised and E = cost of water per liter. The average cost of water in Alicante, Spain (the city of data collection) is €2.55/m^3 [10]. Hence, E = €0.00255/liters (measurements are in liters in Spain).

The $E(T_{fa})$ versus I_u, for varying thresholds of $\epsilon.MAD(Q^r)$ will be obtained for different attack parameters instead of the typical ROC curves. If the impact of attack of undetected attack does not drastically increase for higher $E(T_{fa})$ then it indicates high performing time series anomaly based attack detector.

I. Hyper-parameter Learning

To obtain the optimal value for ϵ, ω, and FL, the cross-validation set with low attack strength and scale ($\delta_{avg} = 10$ and $\rho_{mal} = 10\%$) was used. The safe margins were gotten from the historical data and the residual distance was calculated. Using this, different sets of standard limits for different ω were calculated using algorithm 1. The ratio of the cross-validation set was obtained. Using this ratio, the residual distance values for the cross-validation set was calculated, then different sliding frame length (from 2 - 14) was used on the RUC. This set of different values gotten from the different frame lengths was used with the different set of standard limits initially gotten. This process was repeated for all possible values of $\epsilon.MAD(Q^r)$ taken from 0 to 3 with a step size of 0.02. The optimal values obtained from this process are $FL = 5$, $\omega = 3$, and ϵ = 1.96.

IV. Experimental Results

We divided the 92-household data into three (3), 7 months (August 2016 - February 2017) training sets which also served as the historical data, 2 months (March 2016 - April 2016) cross-validation set with some attack ($\delta_{avg} = 10, \rho_{mal} = 10\%$) to obtain the optimal hyperparameter for FL and ϵ and ω. 3 months (May 2016 - July 2016) test sets were used to validate our model.

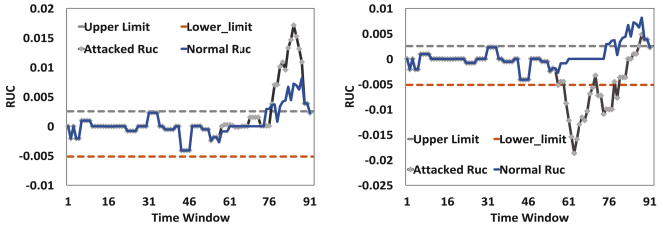
A. Attack Parameters

Using the DAIAD dataset, the experiments for different falsification margins (from 10 - 50, δ_{avg}) along with the different percentages of houses attacked (from 10 - 50, ρ_{mal}) was carried out with the test set. Both additive and camouflage attacks were launched with the pulse strategy by taking various combinations of the attack parameters δ_{avg} and ρ_{mal}.

B. Deductive Attack

The tier 2 with low strength, low scale deductive attack ($\delta_{avg} = 15, \rho_{mal} = 10\%$) is shown in Figure 5a.

It can also be observed that for deductive attacks, the ratio deviated upward instead of an expected downward deviation as observed in previous works. A further investigation into the AM and HM shows that there is a downward deviation for both (for a deductive attack). This phenomenon happened because the standard deviation of the data after an attack was launched decreased instead of an expected increase (which is the observation of other cyber-physical systems (CPS) data like electricity grid and transportation as observed in previous work). Also, The log-normal shape of the data distribution and the outliers greatly affected the attack datapoint to fall on the right of the mean and since $\Delta HM < \Delta AM$, the ratio will increase with reduced values and the SD will decrease.



(a) Deductive ($\delta = 15$, $\rho = 10\%$) (b) Camouflage ($\delta = 10$, $\rho = 50\%$)

Fig. 5: Attack detection

C. Camouflage Attack:

The attack was performed by taking the extreme values of the attack scale and attack strength (that is low scale high strength and high scale high strength). Only the results for tier 2 detection for $\delta_{avg} = 10$ and $\rho = 50\%$ is shown in Fig. 5b. The model detected the camouflage attack because of the difference between the rate of decay of the AM and HM. From figure 5b, the attack was confirmed with tier 2 within 2 days. The low attack strength that the adversary employs in order to remain in the system for a long time also ensures that impact I_u is not high. The ratio is seen to be decreasing in value because the standard deviation for the attack data is greater than the standard deviation of the actual data.

TABLE I: Experimental Results

$\rho_{mal}(\%)$	δ_{avg}	$E(T_{FA})$	I_u(€)
Deductive Attack			
10	50	23	0.669375
20	40	23	0.21675
30	30	23	0.0796875
40	20	23	0.07225
50	10	23	0.0669375
Worst case deductive attack i.e. low attack strength			
10	15	23	0.3155625
10	10	23	0.2199375
Camouflage attack			
10	50	92	1.434375
50	10	92	0.08925

D. Performance Evaluation

In this section, we report the performance over varying attack parameters to give an average sense of performance. To evaluate the performance of the model, the expected time between false alarms $E(T_{FA})$ and the impact of undetected attack I_u was plotted for each threshold $\epsilon.MAD(Q^r)$. With each $\epsilon.MAD(Q^r)$, different sets of $E(T_{FA})$, and I_u was gotten for deductive attack. The result in Figure 6 shows that our model performs well mostly by reducing the impact of the attack while increasing the expected time between false alarms. The threshold chosen for $\epsilon.MAD(Q^r) = 1.96$ was observed to be the point where the expected time between false alarms is high without a considerable increase in the impact of the attack. ***Table 1*** shows the results for the complete attack spectrum the attacker can employ and it can be seen that the expected time between false alarms and the impact of undetected attack shows a good result.

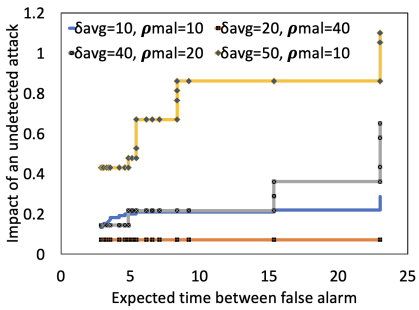


Fig. 6: Impact I_u Vs Expected time between false alarms

V. Conclusion

In this work, we presented a real-time, lightweight, privacy-preserving detection model for the FDI attacks in SWM in WDN. We show that the HM/AM ratio can be used as an invariant for the detection of the attacks as it is stable under normal conditions but deviates under an attack. The second tier of detection was used which is based on the first which aims to reduce false alarms, and with this second tier, the expected time between false alarms was increased. Different attack margins were tested with different types of attacks, and the results were presented.

The main contribution of this work is as follows; A base understanding of the behavioral pattern of water usage in households was done. This enables us to be able to have a structure of benign behavior of water usage. With this behavior known, any deviations from the norm were categorized as an attack. Based on this behavior, specifically the extreme values, a new way of determining the safe margin was done and this was able to reduce the false alarms by handling the outliers in the historical data. Further studies will consider more years of data with more households.

Acknowledgements: This research is supported by the National Science Foundation grants under award numbers: SATC-2030611, SATC-2030624, DGE-1914771, and OAC-2017289.

References

[1] S. Adepu and A. Mathur, "Distributed attack detection in a water treatment plant: Method and case study," IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 1, pp. 86–99, 2018.

[2] C. M. Ahmed, C. Murguia, and J. Ruths, "Model-based attack detection scheme for smart water distribution networks," in Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, 2017, pp. 101–113.

[3] S. Bhattacharjee and S. K. Das, "Detection and Forensics against Stealthy Data Falsification in Smart Metering Infrastructure," IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 1, pp. 356–371, Jan. 2021.

[4] Y. H. Choi, A. Sadollah, and J. H. Kim, "Improvement of Cyber-Attack Detection Accuracy from Urban Water Systems Using Extreme Learning Machine," Applied Sciences, vol. 10, no. 22, p. 8179, 2020.

[5] F. Moazeni and J. Khazaei, "MINLP modeling for detection of SCADA cyberattacks in water distribution systems," in World Environmental and Water Resources Congress 2020: Hydraulics, Waterways, and Water Distribution Systems Analysis, 2020, pp. 340–350.

[6] Helix, "Smart Water Meter Consumption Time Series - Datasets - HELIX," Hellenic Data Service, 2020. https://data.hellenicdataservice.gr/dataset/78776f38-a58b-4a2a-a8f9-85b964fe5c95 (accessed Apr. 03, 2022).

[7] Daiad, "Trials Evaluation and Social Experiment Results," European Commission's 7th Framework Programme., 2017. daiad.eu/wp-contentuploads/201711D7.3_Trials_Evaluation_v1.0.pdf (accessed Apr. 03, 2022).

[8] P. Roy, S. Bhattacharjee, and S. K. Das, "Real Time Stream Mining based Attack Detection in Distribution Level PMUs for Smart Grids," 2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings, Dec. 2020.

[9] D. I. Urbina et al., "Limiting the impact of stealthy attacks on industrial control systems," dl.acm.org, vol. 24-28-October-2016, pp. 1092–1105, Oct. 2016.

[10] Á. F. Morote, M. Hernández, J. Olcina, and A. M. Rico, "Water Consumption and Management in Schools in the City of Alicante (Southern Spain) (2000–2017): Free Water Helps Promote Saving Water?," Water 2020, Vol. 12, Page 1052, vol. 12, no. 4, p. 1052, Apr. 2020.

[11] C. Yan, H. Shin, C. Bolton, W. Xu, Y. Kim, and K. Fu, "SoK: A minimalist approach to formalizing analog sensor security," Proc IEEE Symp Secur Priv, vol. 2020-May, pp. 233–248, May 2020, doi: 10.1109/SP40000.2020.00026.

[12] [Online] M2M, "Challenges of Smart Water Metering - b31, Cat-M, LwM2M, MBus, MQTT, nb-iot, S0," m2mserver, Available at: https://m2mserver.com/en/challenges-of-smart-water-metering/ (accessed Apr. 03, 2022).

[13] [Online] CTVnews, "Thieves in California are stealing scarce water amid extreme drought, 'devastating' some communities — CTV News," 2021. Available at: https://www.ctvnews.ca/climate-and-environment/thieves-in-california-are-stealing-scarce-water-amid-extreme-drought-devastating-some-communities-1.5524683 (accessed May 05, 2022).