IET Cyber-Physical Systems: Theory & Applications

Research Article

Entropy-based electricity theft detection in AMI network

ISSN 2398-3396
Received on 12th May 2017
Revised 20th July 2017
Accepted on 23rd August 2017
doi: 10.1049/iet-cps.2017.0063
www.ietdl.org

Sandeep Kumar Singh[1] ✉, *Ranjan Bose*[1], *Anupam Joshi*[2]

[1]*Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110 016, India*
[2]*Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250, USA*
✉ *E-mail: sandeepsingh012@gmail.com*

Abstract: Advanced metering infrastructure (AMI), one of the prime components of the smart grid, has many benefits like demand response and load management. Electricity theft, a key concern in AMI security since smart meters used in AMI are vulnerable to cyber attacks, causes millions of dollar in financial losses to utilities every year. In light of this problem, the authors propose an entropy-based electricity theft detection scheme to detect electricity theft by tracking the dynamics of consumption variations of the consumers. Relative entropy is used to compute the distance between probability distributions obtained from consumption variations. When electricity theft attacks are launched against AMI, the probability distribution of consumption variations deviates from historical consumption, thus leading to a larger relative entropy. The proposed method is tested on different attack scenarios using real smart-meter data. The results show that the proposed method detects electricity theft attacks with high detection probability.

1 Introduction

The power grid has become an essential part in the present-day society. People's day-to-day life will be affected dramatically without a reliable and stable power grid [1]. Six hundred million people (about 9% of world population) have been affected in July 2012, Indian Blackout and 20 of India's 29 states were hit by power cut [2, 3]. Traditional electric grid is not suitable for today's power requirements [4]. Nowadays, nations have been modernising their existing power system into smart grid due to the evolution of information system and communication technology. Some important features of smart grid are two-way energy transmission and data communication, high reliability, improved efficiency, real-time demand response, and security. Within smart grid, one of the prime technologies being used currently is the advanced metering infrastructure (AMI).

AMI is an integration of multiple technologies such as smart metering, home area networks, integrated communications, data management applications, and standardise software interface. AMI gives consumers the information they need to make intelligent decisions, the ability to execute those decisions, and a variety of choices leading to substantial benefits that they do not currently enjoy [5]. In AMI, system operators are able to greatly improve consumer services by refining utility operating and asset management process based on AMI data. AMI offers an essential link between the grid, consumers, and their loads, and generation and storage resources. In AMI, there is no need to send a person to read electricity consumption readings on site and consumers can also monitor energy consumption remotely and do some customised control. AMI has the ability to send information to consumers about their energy use and dynamic electricity pricing. In AMI, electricity metering system has been modernised by replacing the mechanical meters by smart meters. AMI provides benefit to both consumers and utilities. Consumer benefits include more choices about price and service, and more information to manage consumption, cost, and other decisions. AMI provides benefit to utility in billing and operations.

Smart meters are low-cost computer intelligence devices that have limited resources and several years operational lifetime. The smart meters would have some minimum basic features such as measurement of electrical energy parameters, bidirectional communication, tamper event detection, recording and reporting, integrated load limiting switch/relay, power event alarms such as loss of supply, low/high voltage, remote firmware upgrade, net metering features, and on demand reading. While some security mechanism has been developed for cyber threats in smart meters, they are not sufficient to prevent attacks [6, 7]. Strong security mechanism cannot be implemented in these devices due to resource limitations. The cost of meters may increase due to the addition of extra hardware to implement strong security mechanism. Among all the cyber attacks in AMI, the study of electricity theft is important in both developed and developing countries. Commercial loss due to electricity theft in US was about $6 billion/year [8]. Electricity theft in India causes $17 billion/year revenue loss [9]. A world bank report highlights that the amount of electricity involved in non-technical losses are ∼50% in developing countries [10]. A good electricity theft detection system in modern smart grid is needed to detect electricity theft attacks efficiently.

Traditionally, electricity theft has been detected by physically checking tamper-evident seals, but adversary can easily defeat tamper-evident seal. In AMI, high-resolution meter data is collected and this data can be used in electricity theft detection. Electricity consumption data of smart meter is sent to a control centre at predefined small time intervals. One key feature of AMI is that the transferred data is highly predictable. It is easy to find out the statistical properties for each consumer's consumption pattern. Predictability of AMI data makes this system different from traditional information technology system.

Different electricity theft detection schemes have been proposed in literature. These detection schemes are broadly categorised into three classes: (i) state, (ii) game theory, and (iii) classification based. In state-based methods, Lo and Ansari [11] proposed a CONSUMER attack model to improve network observability and detection accuracy by grid sensor placement algorithm. McLaughlin *et al.* [12] propose AMI intrusion detection system that uses information fusion to combine the sensors and consumption data from a smart meter for more accurate energy theft detection. In [13], a mutual inspection strategy is proposed to discover problematic meters that report malicious consumption readings. Khoo and Cheng [14] propose a system that implements RFID technology to prevent electricity theft. In all of the above state-based detection methods, certain devices, like radio-frequency identification tags and wireless sensors are used for higher detection probability at a cost of extra investment.

Game theory-based methods are also discussed in literature for electricity theft detection. Amin *et al.* [15] address incentives of

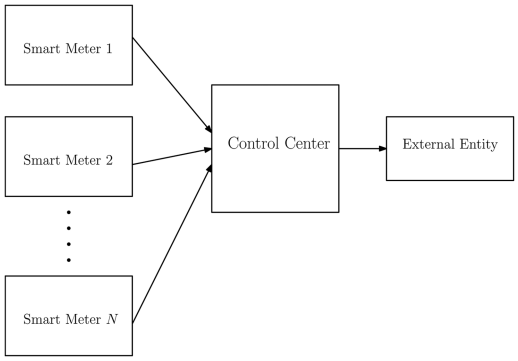

Fig. 1 *Network model*

utility company to combat non-technical losses, when company is subject to tariff regulation. They proposed that regulators should incorporate explicit targets for permissible losses to solve the problem of incentive misalignment. In [16], the problem of electricity theft detection is discussed as a game between the electricity thief and the electricity utility. Yet, in game theory-based methods, formulation of utility function for all the players is a challenging task.

Different classification-based methods have also been studied in literature. Machine learning schemes are used to train a classifier using sample database. Abnormal patterns are recognised by a trained classifier. In [17], C-means-based fuzzy clustering algorithm is proposed to detect defect measurement. Salinas *et al.* [18] propose distributed algorithms to solve linear equations for users' honesty coefficients. Authors utilise peer-to-peer computing to identify the dishonest consumers in the system. In [19], a neural network model is implemented and an encoding technique is proposed to identify malicious users. Depuru *et al.* [20] discuss the problems underlying detection of electricity theft using support vector machine (SVM). Jokar *et al.* [21] present consumption pattern-based theft detection scheme using multi-class SVM. In [22], tree-based threat model is presented to explain the energy theft in AMI. Mashima and Cárdenas [23] present threat model to detect energy theft and use real data from an AMI system to validate proposed method. There are many limitations of classification-based methods. Primary issue with these methods is data imbalance. Due to this, the benign and malicious samples are not in the same range. Benign samples are obtained from historical dataset but malicious samples (theft samples) hardly exist for a specific customer. Unavailability of dataset of malicious samples limits the detection rate (DR). Second, these schemes are vulnerable to contamination attacks. Adversary may granularly change data and pollute the dataset and it results to deceive the learning process to consider a malicious data as a normal data. There are many non-malicious reasons that can alter electricity consumption such as change of appliances, change of season, change of residents, and so on. These factors result to high false positive rate (FPR).

In this paper, we propose an entropy-based electricity theft detector (EBETD) to detect electricity theft in AMI. The main idea of the proposed scheme is to track the electricity consumption dynamics by computing the distance indices between adjoining time steps. Relative entropy (Kullback–Leibler distance) [24] is used to compute distance indices. The relative entropy is a measure of the distance between two probability distributions. In this paper, we show that if the relative entropy is more than the threshold value, there is a high probability of electricity theft attack against the AMI. The threshold value is decided on the basis of historical electricity consumption measurements. We test the performance of EBETD with real data of 5000 consumers [25]. The proposed EBETD method detects electricity theft attack with high probability of detection as compared to previously reported results.

The rest of the paper is organised as follows. In Section 2, we discuss network model and threat model. Section 3 presents proposed methodology to detect electricity theft attacks. Case study is given in Section 4. Section 5 presents results and discussion. Section 6 concludes this paper.

2 Network model and threat model

2.1 Network model

Fig. 1 shows the network model for AMI. We consider N smart meters, each meter conveying its consumption readings to a control centre. Control centre sends the final aggregate data to external entities such as a third-party service providers or grid managers. Control centre is responsible to check whether the measurements coming from the meters is benign or malicious. Therefore, all security measures are taken at the control centre.

2.2 Threat model

The goal of the adversary is to launch electricity theft attack by compromising smart meters and sending malicious readings to the control centre. The objective of electricity theft is to get financial benefit by paying less than the actual value for the consumed electricity. There are different techniques for electricity theft that an adversary may attempt against AMI. Electricity theft techniques [12] are classified into three categories: (i) physical attacks, (ii) cyber attacks, and (iii) data attacks. Attack in the third category is made possible through categories first and second. The above-mentioned attacks are discussed briefly below:

(i) *Physical attacks:* Meter tampering is one type of physical attack to hinder actual recording of electricity consumption. This may be done by a strong magnet to cause interference with the instruments. Reversing or disconnecting the meters, bypassing the meters to remove loads from measurements are some other type of physical attacks.
(ii) *Cyber attacks:* Examples of cyber attacks include compromising smart meters through remote network exploits, modifying the firmware or storage on meters, intercepting communications, and interrupting measurements. Cyber attacks can be carried out within the meter or over the communication link between the meter and the utility company.
(iii) *Data attacks:* Data attacks include targeting the electricity consumption readings and are done through cyber attacks and physical attacks.

Electricity theft techniques mentioned above are discussed in detail in [12]. Electricity consumers are primary attackers to carry out electricity thefts. The second type of attackers are professional hackers who use software and hardware gadgets to compromise meters. The third type of attackers is utility company insiders.

AMI systems transfer electricity consumption data to the utility at small time intervals (every 30 min or less). This time interval may change between distinct AMI deployments. This granular electricity consumption data can be used to improve theft detection mechanism of the system. The metering data is a time series $x_1, x_2, \ldots$, where x_i is the electricity consumption of the consumer from time interval between measurement x_{i-1} to x_i. A malicious time series $\hat{y}_1, \hat{y}_2, \ldots$, is generated by the adversary. The objective for the adversary is to generate a time series $\hat{y}_i$ that will reduce its electricity bill subject to the constraint that the control centre will not raise a flag with $\hat{y}_i$.

3 Proposed methodology

The proposed method detects electricity theft attacks by tracking the dynamics of the consumption data. Relative entropy is used to quantify the consumption variation.

3.1 Relative entropy

It is the expectation of the logarithmic of the likelihood ratio. Let $p(x)$ and $q(x)$ be two probability mass functions of a discrete random variable x, then relative entropy is a measure of the distance between two probability mass functions. For any x in X, $p(x) > 0$ and $q(x) > 0$ and both $p(x)$ and $q(x)$ sum up to 1. The relative entropy [24] (also called Kullback–Leibler distance) between two probability mass functions is defined as

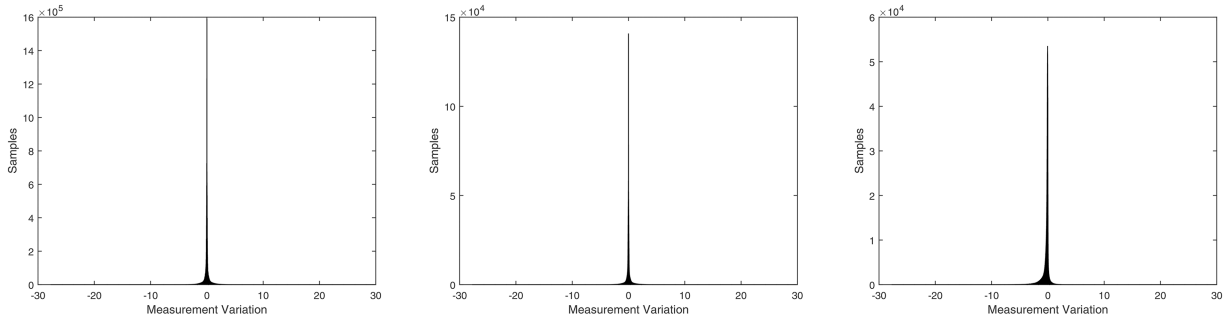

Fig. 2 *Histogram of the electricity consumption measurements in different months*
(a) July 2009 to July 2010, ***(b)*** August 2010, ***(c)*** September 2010

$$D(p \parallel q) = \sum_x p(x)\ln\frac{p(x)}{q(x)} = E_p \ln\frac{p(x)}{q(x)} \quad (1)$$

In the above definition, we use the convention that $0\log(0/0) = 0$, $0\log(0/q) = 0$, and $p\log(p/0) = \infty$. The Kullback–Leibler distance, or simply, the KL distance, of $q(x)$ from $p(x)$ is a measure of the information lost when $q(x)$ is used to approximate $p(x)$. The relative entropy is always non-negative, $D(p \parallel q) \geq 0$, and having zero value if and only if $p = q$. It does not obey triangle inequality and is not symmetric, $D(p \parallel q) \geq 0$ is not equal to $D(q \parallel p)$.

The concept of relative entropy was developed in information theory. It has been commonly used in the data mining literature. In this paper, we use the relative entropy to detect electricity theft in AMI.

The adversary sends malicious electricity consumption readings to the control centre by compromising any individual smart meter with the objective to reduce the electricity bill. The proposed method detects electricity theft attacks using dynamics of electricity consumption data. To quantify electricity consumption variation, the relative entropy $D(p \parallel q)$ is used where q represents the distribution of electricity consumption variation from the historical data and p represents the distribution of electricity consumption variation for current and previous time step.

When there is no electricity theft attack against the AMI, the relative entropy $D(p \parallel q)$ would be relatively small. When compromised data is sent to the AMI, the relative entropy $D(p \parallel q)$ will increase. Relative entropy $D(p \parallel q)$ for the current time step is compared with predefined threshold value which is determined by relative entropy calculated from historical electricity consumption variation, and if the relative entropy is more than predefined threshold value, control centre concludes that attack has been launched by adversary. An appropriate action is taken when electricity theft attack is detected by the control centre.

4 Case study

In our test, we have used smart-meter electricity consumption data from the Irish Social Science Data Archive [25]. The commission for energy regulation, Ireland (CER) and Sustainable Energy Authority of Ireland (SEAI) initiated the smart-metering project with the purpose of undertaking trials to assess the performance of smart meters and their impact on consumers energy consumption. The smart-metering electricity customer behaviour trials took place during 2009 and 2010 with over 5000 Irish homes and businesses participating. Residential and business customers, who participated in the trials had an electricity smart meter installed in their homes/premises and agreed to take part in research to help in establishing how smart metering can help in shaping energy usage behaviours across a different type of home sizes, lifestyles, and demographics [25].

In this database, there is a file for each consumer containing half hourly metering data over a period of 535 days. Each file for each consumer consists of 535 vectors and each vector having 48 components. For a benign dataset, sample vector $\boldsymbol{x}$ includes $[x_1, x_2, \ldots, x_{48}]$ components. For every 30 min time step, the meter sends consumption reading (in Watt) to the control centre.

In the proposed method, we consider the difference between two consecutive measurements. The measurement sent to the control centre by smart meter at time i is x_i. The measurement variation is determined as $(x_i - x_{i-1})$. In case study, we consider $N = 1000$ (number of smart meters). Histogram of measurement variation of N meters from July 2009 to July 2010 is shown in Fig. 2*a*. This historical measurement variation data includes all consecutive measurement variations of N meters. From Fig. 2*a*, it is clear that majority of measurement variations are small and close to zero.

When there are no electricity theft attacks, probabilistic distributions of electricity consumption readings are quite similar between different months. Fig. 2*b* shows the histogram of measurement variation of N meters for August 2010 with no electricity theft attacks. When there is an electricity theft attack against AMI, the histogram of consumption variation is different. The mean and standard deviation of measurement variations has been changed due to attacks. When the adversary launches electricity theft attack, malicious measurement $\hat{y}_i$ is send to control centre at time i. Histogram of measurement variation $(\hat{y}_i - x_{i-1})$ of N meters for September 2010 is shown in Fig. 2*c*.

4.1 Relative entropy-based detection

In our proposed method, we use relative entropy $D(p \parallel q)$ to quantify the dissimilarity between two distributions. Distribution q is found out from the histogram of the historical data of electricity consumption variations from July 2009 to July 2010 as shown in Fig. 2*a*. We make an assumption that the historical dataset is benign. The consumption data of the month of August 2010 is considered to be true and the threshold value is computed using this data. We have taken electricity theft attacks at each time step (every 30 min time interval) for September 2010. To calculate relative entropy, distribution p is derived from Figs. 2*b* and *c* which is the histogram of measurement variation for August and September 2010, respectively. Relative entropy $D(p \parallel q)$ is computed based on (1).

Fig. 3*a* is the histogram of relative entropy for August 2010 with no electricity theft attacks. Range of relative entropy is from 0.2112 to 0.4322. Fig. 3*b* is the histogram of relative entropy for September 2010 with electricity theft attacks. In Fig. 3*b*, relative entropies of all samples are equal to or greater than 0.4322. Therefore, we conclude that electricity theft attacks will increase relative entropy. The reason is that the relative entropy is a measure of distance between two distributions. Electricity theft attacks will affect the distribution of measurement variation, hence it will increase distance from the distribution of historical measurement variation.

To detect electricity theft attacks, we set a threshold value from the histogram of relative entropy of previous month data (Fig. 3*a*). This threshold value is compared with every sample at each time step during runtime. If the runtime relative entropy is larger than the threshold value, it is likely that electricity theft attacks have been launched into the system.

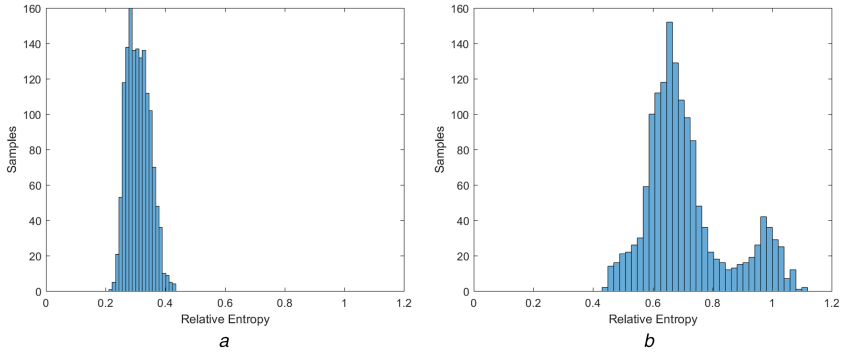

Fig. 3 *Histogram of relative entropy on*
(a) August 2010 with no attack, ***(b)*** September 2010 with attack

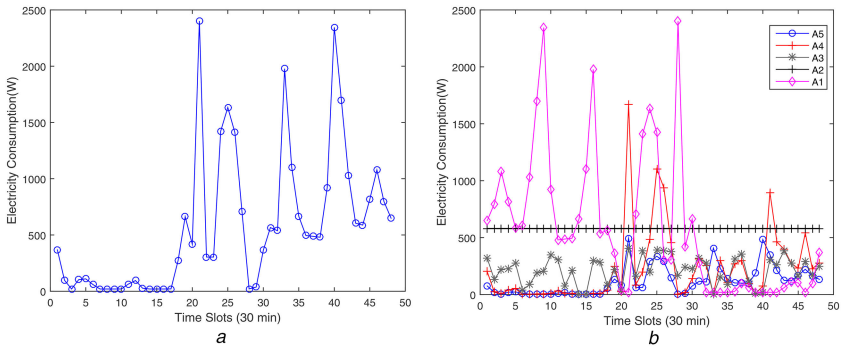

Fig. 4 *Example of the daily electricity consumption pattern of a typical consumer*
(a) True electricity consumption pattern, ***(b)*** Attacked electricity consumption patterns

Choosing a proper threshold value is a key issue in the proposed method and the detection accuracy depends on this. If we set high threshold value, some electricity theft attacks are not detectable. When the threshold value is set to low, some benign consumption readings may be categorised as malicious readings. In [26], threshold value is set using historical measurement to detect false data injection attack in AC state estimation. Similarly, in our proposed method, the threshold value is set on the basis of the histogram of historical electricity consumption. The measurement samples of one month prior to the attack (August 2010) are compared with the historical dataset to obtain the threshold value. Relative entropy at each time step is obtained from (1), and histogram of relative entropy is plotted (Fig. 3*a*). We select maximum histogram distance of relative entropy $D(p \parallel q)$ with certain trust level as threshold value where q is derived from Fig. 2*a* and p is derived from Fig. 2*b*. For example, the 99% trust level means that threshold value is larger than 99% of the historical consumption readings.

5 Results and discussion

All the simulations have been done on DELL PC with 3.20 GHz Intel Core i7 processor and 8 GB RAM on Windows 8 Enterprise. Programming has been done on MATLAB R2014b. To check the detection efficiency of our proposed scheme, we used different type of malicious samples. $X_l = [x_l, x_2, \ldots, x_{48}]$ and $\hat{Y}_l = [\hat{y}_l, \hat{y}_2, \ldots, \hat{y}_{48}]$ are true and malicious consumption data patterns of lth day for any typical smart meter. Different type of attacks [12, 18, 21, 27] for electricity theft are as follows:

i. *Attack 1 (A1):* In this case, the attacked sample is

$$\hat{Y}_l = X_{24-(l-1)} \tag{2}$$

In this attack, adversary reverses the order of consumption data of last day and sends it to control centre.

ii. *Attack 2 (A2):* In the second case, the attacked sample is

$$\hat{Y}_l = \text{mean}(X_{l-1}) \tag{3}$$

Here, adversary sends the mean value of previous day to control centre. Attacks $A1$ and $A2$ are useful in a scenario where dynamic electricity pricing is used. Utility company may charge high electricity price for peak electricity consumption hours. Total electricity consumption is same in attacks $A1$ and $A2$ but consumer can get financial profit.

iii. *Attack 3 (A3):* In the third case, the attacked sample is

$$\hat{Y}_l = \alpha_l \,\text{mean}(X_{l-1}) \tag{4}$$

where $\alpha_l = [\alpha_{l1}, \alpha_{l2}, \ldots, \alpha_{l48}]$ and $\alpha_{li} = \text{random}(\min_i, \max_i)$. In this case, mean value of the previous day's consumption is multiplied by α_l and sent to the control centre.

iv. *Attack 4 (A4):* In the fourth case, the attacked sample is

$$\hat{Y}_l = \alpha_l X_l \tag{5}$$

where $\alpha_l = [\alpha_{l1}, \alpha_{l2}, \ldots, \alpha_{l48}]$ and $\alpha_{li} = \text{random}(\min_i, \max_i)$. In this attack, all consumption readings for a particular day are multiplied by different randomly selected numbers between $(\min_i, \max_i)$.

v. *Attack 5 (A5):* In the fifth case, the attacked sample is

$$\hat{Y}_l = \alpha X_l \tag{6}$$

where $\alpha = \text{random}(\min, \max)$. In this attack, all samples for a particular day are multiplied by the same random number. $(\min, \max)$ are used to define the amount of electricity theft.

Daily electricity consumption of a typical consumer is shown in Fig. 4*a*. Fig. 4*b* shows the malicious electricity consumption patterns of 24 h under different attacks. Outcome of the three categories of electricity theft techniques discussed in Section 2 is manipulation of the meter readings. All these five type of attacks can be launched using previously discussed electricity theft techniques.

Performance of the proposed method is checked by plotting receiver operating characteristic (ROC) curve [28, 29]. ROC curve

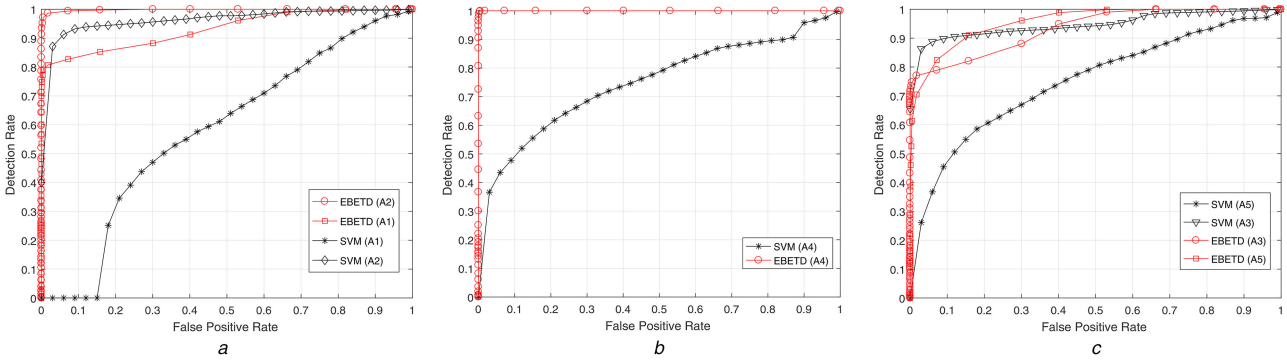

Fig. 5 *ROC curves for EBETD and SVM under different attacks*
(a) For $A1$ and $A2$, ***(b)*** For $A4$, ***(c)*** For $A3$ and $A5$

shows true positive rate (DR) versus FPR (equivalently sensitivity versus 1−specificity) for different threshold. True positive rate or DR is the proportion of true positives that are correctly identified by the test and specificity is the proportion of true negatives that are correctly identified by the test. 1−specificity is FPR.

5.1 Test results

To check the performance of EBETD, samples from July 2009 to July 2010 are used for historical distribution. Samples of August 2010 are used for threshold selection. The data for September 2010 is malicious. We choose min and max equal to 0.2 and 0.8, respectively. In test setup, we have taken number of smart meters N equal to 1000.

The key feature in AMI network is that the consumption of all meters follow a certain statistical pattern at any time. In our test, we consider that consumers launch attack in a coordinative manner so that control centre may not detect the attack. Control centre stores measurement data from all consumers and run EBETD algorithm at each time step to check that data coming from all consumers is benign or malicious.

Table 1 summarises the test results. From the test results, we can see that EBETD successfully detects attacks with high DR. Attacks $A1$ and $A2$ are launched especially in dynamic electricity pricing scenario. The proposed scheme detects attacks $A1$ and $A2$ with 81.24 and 98.92% DR, respectively. Attacks $A3$, $A4$, and $A5$ are launched to reduce the electricity bill by sending the malicious consumption readings having value less than the actual consumption readings. The proposed scheme detects attacks $A3$, $A4$, and $A5$ with 77.03, 100, and 94.22% DR, respectively.

We compare the performance of the EBETD algorithm with classification based method like SVM [21]. In SVM, we train the classifier using malicious and benign data samples and employ k-means clustering algorithm on benign data. For most of the consumers, $k = 1$ or 2 is obtained from the silhouette plot. In SVM test, training set includes 420 day's benign data samples and 420 day's malicious data samples. Thirty day's data samples of September 2010 are used in testing phase. Fig. 5 shows the performance of EBETD and SVM for different attack scenario. We observe that EBETD provides better performance as compared to the multi-class SVM.

Table 1 Summery of test results for electricity theft attacks

Type of attack	Detected samples
attack 1	81.24
attack 2	98.92
attack 3	77.03
attack 4	100
attack 5	94.22

Table 2 Effect of threshold on detection performance

Trust level, %	FPR, %	DR, %
90	23.13	100
91	21.25	100
92	20.31	100
93	18.09	100
94	16.14	100
95	14.59	100
96	11.97	100
97	10.15	100
98	7.95	100
99	2.82	100
100	0.47	99.96

Table 3 Effect of α on detection performance

α		Detected samples, %		
Min	Max	Attack 3	Attack 4	Attack 5
0.1	0.9	74.29	98.14	90.03
0.2	0.8	77.03	100	94.22
0.3	0.7	99.99	100	98.89
0.4	0.6	100	100	99.95
0.5	0.5	100	100	100

5.2 Effect of threshold on performance

Threshold selection is a key issue in the proposed scheme. In [26], threshold value is set using historical measurement to detect false data injection attack in AC state estimation. Similarly, in our proposed method, the threshold value is set from the histogram of relative entropy (as shown in Fig. 3*a*) with certain trust level. We check the performance of our proposed scheme for different trust levels. Table 2 shows FPR and DR for different trust levels when the adversary launches $A4$ attack. From Table 2, it is clear that if we set a threshold value with low trust level, high DR is achieved with high FPR. The opposite is true for high trust level. For example, DR is 100% and FPR is 23.13% when trust level is 90% but if we set high trust level like 99%, FPR is 2.82% and DR is 100%. In proposed scheme, we select 99% trust level because low FPR is obtained at this value.

5.3 Effect of α on performance

In our proposed scheme, parameter α, used in attack formation, depends on min and max values. We check the performance of the proposed scheme for different (min , max) pairs under different attack scenario. We considered five different (min , max) pairs as (0.1, 0.9), (0.2, 0.8), (0.3, 0.7), (0.4, 0.6), and (0.5, 0.5). For these (min , max) pairs, mean value of compromised meter readings is same but having different variances for each pair. Table 3 shows DR for different α when adversary launched attacks $A3$, $A4$, and $A5$. From Table 3, we observe that the proposed scheme is capable to detect all attacks successfully with high DR for different (min , max) pairs.

Table 4 Effect of sampling rate on detection performance

Sampling rate, samples/day	FPR, %	DR, %
6	18.92	33.51
12	1.08	43.92
24	2.56	92.69
48	2.82	100

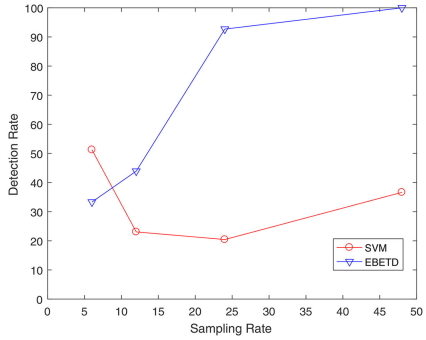

Fig. 6 *DR for EBETD and SVM*

5.4 Effect of sampling rate on performance

We analysed the effect of sampling rate on performance of the proposed scheme. Sampling rate indicates the amount of information that can be extracted from the users' data. Table 4 shows the result when adversary launches *A*4 attack. We observe that when the sampling rate is 6 samples/day, DR is 33.51% and for 48 samples/day, DR is 100%. Table 4 shows that as we increase sampling rate, DR increases. In the proposed scheme, we have chosen sampling rate equal to 48 samples/day.

We compare the performance of EBETD with SVM for different sampling rates. Fig. 6 shows DR for EBETD and SVM when sampling rate is 6, 12, 24, and 48 samples/day (for *A*4 attack). From Fig. 6, we observe that our proposed scheme achieves high DR than SVM.

6 Conclusion

Electricity theft is a key threat in AMI security. In this paper, we have proposed EBETD, a novel scheme, based on relative entropy to detect energy theft in AMI. EBETD relies on the dynamics of consumption variation of consumers. Under normal scenario with no electricity theft attack, the relative entropy is small. After an electricity theft attack, the relative entropy tends to be larger than a preselected threshold.

We have analysed the performance of proposed scheme under different attack scenarios using real dataset and compared test results with SVM. Test results show that EBETD can detect attacks with high detection probability. It also outperforms SVM-based theft detectors. We have also analysed the effect of the threshold, α, and the sampling rate on detection performance.

7 References

[1] Jiang, R., Lu, R., Lai, C., *et al.*: 'Robust group key management with revocation and collusion resistance for scada in smart grid'. Proc. IEEE Global Communications Conf. (GLOBECOM), December 2013, pp. 802–807

[2] 'India blackouts leave 700 million without power'. Available at https://www.theguardian.com/world/2012/jul/31/india-blackout-electricity-power-cuts

[3] 'Second day of India's electricity outage hits 620 million'. Available at http://usatoday30.usatoday.com/news/world/story/2012-07-31/india-power-outage/56600520/1

[4] Lu, R., Liang, X., Li, X., *et al.*: 'Eppa: an efficient and privacy-preserving aggregation scheme for secure smart grid communications', *IEEE Trans. Parallel Distrib. Syst.*, 2012, **23**, (9), pp. 1621–1631

[5] 'The NETL Modern Grid Strategy Powering our 21st-Century Economy: Advanced Metering Infrastructure', February 2008. Available at https://www.smartgrid.gov/files/NISTSGInteropReportPostcommentperiodversion200808.pdf

[6] Wright, J.: 'Smart meters have security holes', 2010

[7] Ward, M.: 'Smart meters can be hacked to cut power bills', October 2014. Available at http://www.bbc.com/news/technology-29643276

[8] McDaniel, P., McLaughlin, S.: 'Security and privacy challenges in the smart grid', *IEEE Secur. Privacy*, 2009, **7**, (3), pp. 75–77

[9] Katakey, R.: 'India fights to keep the lights on', June 2014. Available at http://www.bloomberg.com/news/articles/2014-06-05/india-fights-electricity-theft-as-modi-pledges-energy-upgrade

[10] 'Reducing technical and non-technical losses in the power sector. Washington, DC: World bank group'. Available at http://documents.worldbank.org/curated/en/829751468326689826/Reducing-technical-and-non-technical-losses-in-the-power-sector

[11] Lo, C.-H., Ansari, N.: 'Consumer: a novel hybrid intrusion detection system for distribution networks in smart grid', *IEEE Trans. Emerg. Top. Comput.*, 2013, **1**, (1), pp. 33–44

[12] McLaughlin, S., Holbert, B., Fawaz, A., *et al.*: 'A multi-sensor energy theft detection framework for advanced metering infrastructures', *IEEE J. Sel. Areas Commun.*, 2013, **31**, (7), pp. 1319–1330

[13] Xiao, Z., Xiao, Y., Du, D.H.c.: 'Non-repudiation in neighborhood area networks for smart grid', *IEEE Commun. Mag.*, 2013, **51**, (1), pp. 18–26

[14] Khoo, B., Cheng, Y.: 'Using RFID for anti-theft in a Chinese electrical supply company: a cost-benefit analysis'. Wireless Telecommunications Symp. (WTS), April 2011, pp. 1–6

[15] Amin, S., Schwartz, G.A., Tembine, H.: 'Incentives and security in electricity distribution networks'. Int. Conf. on Decision and Game Theory for Security, GameSec, 2012, pp. 264–280

[16] Crdenas, A.A., Amin, S., Schwartz, G., *et al.*: 'A game theory model for electricity theft detection and privacy-aware control in AMI systems'. 50th Annual Allerton Conf. on Communication, Control, and Computing (Allerton), 2012, October 2012, pp. 1830–1837

[17] Angelos, E.W.S., Saavedra, O.R., Corts, O.A.C., *et al.*: 'Detection and identification of abnormalities in customer consumptions in power distribution systems', *IEEE Trans. Power Deliv.*, 2011, **26**, (4), pp. 2436–2442

[18] Salinas, S., Li, M., Li, P.: 'Privacy-preserving energy theft detection in smart grids: a p2p computing approach', *IEEE J. Sel. Areas Commun.*, 2013, **31**, (9), pp. 257–267

[19] Depuru, S.S.S.R., Wang, L., Devabhaktuni, V., *et al.*: 'A hybrid neural network model and encoding technique for enhanced classification of energy consumption data'. IEEE Power and Energy Society General Meeting, July 2011, pp. 1–8

[20] Depuru, S.S.S.R., Wang, L., Devabhaktuni, V.: 'Support vector machine based data classification for detection of electricity theft'. Power Systems Conf. and Exposition (PSCE), 2011, March 2011, pp. 1–8

[21] Jokar, P., Arianpoo, N., Leung, V.C.M.: 'Electricity theft detection in AMI using customers consumption patterns', *IEEE Trans. Smart Grid*, 2016, **7**, (1), pp. 216–226

[22] Jiang, R., Lu, R., Wang, Y., *et al.*: 'Energy theft detection issues for advanced metering infrastructure in smart grid', *Tsinghua Sci. Technol.*, 2014, **19**, (2), pp. 105–120

[23] Mashima, D., Cárdenas, A.A.: '*Evaluating electricity theft detectors in smart grid networks*' (Springer Berlin Heidelberg, Berlin, 2012), pp. 210–229

[24] Cover, T.M., Thomas, J.A.: '*Elements of information theory*' (Wiley-Interscience, 2006)

[25] 'Irish Social Science Data Archive'. Available at http://www.ucd.ie/issda/data/commissionforenergyregulationcer/

[26] Chaojun, G., Jirutitijaroen, P., Motani, M.: 'Detecting false data injection attacks in ac state estimation', *IEEE Trans. Smart Grid*, 2015, **6**, (5), pp. 2476–2483

[27] Krishna, V.B., Weaver, G.A., Sanders, W.H., *et al.*: 'Pca-based method for detecting integrity attacks on advanced metering infrastructure'. Proc. of the 12th Int. Conf. on Quantitative Evaluation of Systems, New York, NY, USA, 2015, pp. 70–85

[28] Fawcett, T.: 'Roc graphs: notes and practical considerations for researchers'. Technical Report, 2004

[29] Zweig, M.H., Campbell, G.: 'Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine', *Clin. Chem.*, 1993, **39**, (4), pp. 561–577