DEFENDING ELECTRICITY-STEALING BASED ON DATA ANALYSIS: A NOVEL APPLICATION OF ADVANCED METERING INFRASTRUCTURE FOR DISTRIBUTION NETWORKS

Ling LUO
State Grid Shanghai EPRI – China
luol@sh.sgcc.com.cn

Kaiyan XIAO, Zhiyong CHEN
Shanghai Jiaotong University – China
xiaokaiyan@sjtu.edu.cn, zhiyongchen@sjtu.edu.cn

Jiapei JIN
State Grid Shanghai EPRI - China
jinjp@sh.sgcc.com.cn

ABSTRACT

In this paper, we propose a novel electricity anti-stealing approach based on data analysis in the advanced metering infrastructure for distribution networks. This technique detects electricity-stealing suspects by short-time load forecasting of clients via time series analysis. By processing the load data from a single client, the system is able to automatically alarm when the real load is excursive from the forecasting load. This method greatly improves the credibility and reduces the cost and complexity of anti-stealing detection. The case studies demonstrate the success in our anti-stealing application by showing accuracy of the short-term load forecasting approach.

INTRODUCTION

Electricity-stealing is a universal problem for all power utilities, causing a large amount of economic losses. Most of the common anti-stealing techniques mainly focus on the communications [1] and metering circuits [2-3], while some of the an-stealing approaches based on the power demand information acquisition systems are still at the beginning step and lack of strict mathematical models [4]. Therefore, there is an increasing need of intelligent anti-stealing technologies for the power utilities to rapidly detect the electricity-stealing suspects.

The ultimate goal of electricity-stealing is to achieve economic gains by making the demonstrated values of power consumption in the meter less than the actual values. The current defending techniques are limited, most of which are still based on manual inspection of load curves of all users followed by locally checking meters. Since these traditional anti-stealing technologies are labour-cost and inaccurate, there are still a lot of electricity-stealing phenomena are out of detection by power utilities.

This paper proposes an anti-stealing technique that can sense the possible stealing of a large set of users by data mining from the load database of consumers via a short-term statistical tool. No matter how a client steals electricity, the final feature of the load data from a meter is 'abnormal' decrease. Therefore, the key point to catch the suspects of electricity-stealing is to precisely detect the 'abnormal' decrease of load data from a meter. In such a way the technique reaches the tentative judgment of electricity-stealing, and hence significantly reduces the long list of suspects. Compared to the traditional approaches, this technique applies short-term load forecast to model the behaviour of clients via communication and storage of load data.

APPLICATION OF ADVANCED METERING INFRASTRUCTURE

The current smart grid being globally deployed changes the way energy is used. This new infrastructure offers more efficient, lower cost, and more environmentally sound energy management than its antiquated predecessor. The advanced metering infrastructure (AMI) is a crucial piece of the smart grid infrastructure [5].

AMI is a technique that automatically collects data from meters and transfers data to a central database by communication technology for remote control and analyzing, replacing the traditional analog meters with computerized systems that report usage over digital communication interfaces. AMI is composed of smart meters, regional metering devices, local and remote communication tools, two-way communication channels, power management terminals, the management software, a controlling station and other affiliated equipments. Therefore, AMI is seen as the next technique to replace manual meter reading to save human resources, meanwhile reducing meter reading costs and improving meter reading efficiency. Although an AMI system is so powerful that it can find the abnormal information to alarm in time by monitoring the data of users, there is no research work focusing on a practical application of data analysis on top of AMI to defend electricity-stealing.

One of the key applications of AMI is anti-stealing, which is more accurate and less complex, matching the development direction of the smart grid by efficiently utilizing the database of users. The main anti-stealing idea of applying AMI in this paper is to judge the abnormal load of the customer by prediction via short-term forecasting.

Among several load forecasting techniques such as regression, least square, neural network and grey model, the time series method is chosen in this paper due to its simplicity and accuracy in short-term forecast [6].

AN ANTI-STEALING TECHNIQUE BASED ON SHORT-TERM LOAD FORECASTING

Modeling the Short-Term Load Forecast Based on Time Series Analysis

In time series analysis, the data are often modeled as ARMA(p, q), where ARMA denotes autoregressive-moving-average model. In this model, p is the order of the autoregressive part and q is the order of the moving-average part [7]. The typical ARMA(p, q) model is described as follows:

$$\sum_{i=0}^{p}\varphi_i y(t-i)=\sum_{j=0}^{q}\theta_j\alpha_{t-j},\varphi_0=\theta_0=1,$$

where

- $y(t-p),y(t-p+1),\ldots,y(t-1),y(t)$ are load data in a time interval.
- $\alpha_t,\alpha_{t-1},\ldots,\alpha_0$ are white noises. In common cases their probability distribution are assumed as Gaussian, and $E[\alpha_t^2]=\sigma_\alpha^2$.
- p,q are the orders of the autoregressive part and moving-average part respectively.
- φ,θ are coefficients of autoregressive part and moving-average part respectively.

After transforming the raw data to stationary, the ARMA model can be established to analyze the data. The first procedure is to apply null transformation towards the data,

$$Y(i)=y(i)-\bar{y},t\in[1,N],$$

where N denotes the sample size. To ensure the accuracy of prediction, we normally set this size to be larger than 50. Then we can calculate the estimation of autocorrelation function ρ_k of $Y(i)$ using the following equation:

$$\rho_k=\frac{\frac{1}{N-k}\sum_{i=1}^{N-k}Y(i)Y(i+k)}{\frac{1}{N}\sum_{i=1}^{N}Y(i)^2}.$$

The estimation of partial correlation function is calculated by the recursive formula as

$$\left.\begin{aligned}a_{11}&=\rho_1\\ a_{k+1,k+1}&=(\rho_{k+1}-\sum_{j=1}^{k}a_{kj}\rho_{k+1-j})(1-\sum_{j=1}^{k}a_{kj}\rho_j)^{-1}\\ a_{k+1,j}&=a_{kj}+a_{k+1,k+1}a_{k,k-j+1}\end{aligned}\right\}.$$

According to Box Jenlkins' principle, we can determine the order of the ARMA model (p, q) by judging the tailing off property and cutting off property of ρ_k and α_{kk} with 95% confidence. In order to improve the estimation accuracy, lower order of ARMA model is preferred.

Following the identification of the model and determination of the orders, it is possible to estimate the parameters. The first step is to solve the estimation of φ by Yule-Walker equations:

$$\begin{bmatrix}\gamma_q & \gamma_{q-1} & \cdots & \gamma_{q-p+1}\\ \gamma_{q+1} & \gamma_q & \cdots & \gamma_{q-p+2}\\ \cdots & & & \cdots\\ \gamma_{q+p-1} & \gamma_{q+p-2} & \cdots & \gamma_q\end{bmatrix}\begin{bmatrix}\varphi_1\\ \varphi_2\\ \cdots\\ \varphi_p\end{bmatrix}=\begin{bmatrix}\gamma_{q+1}\\ \gamma_{q+1}\\ \cdots\\ \gamma_{q+p}\end{bmatrix}.$$

Substituting the estimation of $\varphi_1,\varphi_2,\ldots,\varphi_p$ into equation, let covariance function of $\bar{y}$ be $\bar{\gamma}$, then calculate the estimation of θ and σ_α^2 by the method of moment estimator.

$$\left.\begin{aligned}\bar{\gamma}_0&=\sigma_\alpha^2(1+\theta_1+\cdots+\theta_q)\\ \bar{\gamma}_1&=\sigma_\alpha^2(-\theta_1+\theta_2\theta_1+\cdots+\theta_q\theta_{q-1})\\ &\cdots\\ \bar{\gamma}_q&=\sigma_\alpha^2(-\theta_q)\end{aligned}\right\}$$

Our ARMA(p, q) modeling is then finished, and we have reached the short-term load forecasting equations.

Detection of Electric-Stealing Suspects

Based on the historical data of customers, we apply time-series analysis to model the short-term load forecast of a single user. This section is aimed at discussing how to determine the probability of the possible electricity-stealing behavior of a customer.

For a single user, the behavior of electricity-stealing is composed of two parts-- the amount of power it steals and the duration it steals power. Thus these two parameters should be set up in anti-stealing analysis: the alarm threshold factor *r* and the alarm window *W*. Since the load of a single user is a stochastic process, the defending system needs an alarm window to reduce the false alarm probability.

Letting the forecast load be $\underline{y}$, the alarm threshold is

$$y_{th}=r\underline{y}.$$

As forecasting electricity-stealing behavior, we are going to compare the recorded real load curve with alarm threshold level. In an alarm window starting from time *n*, if all of the values of observed load are lower than alarm threshold:

$$y(t)<y_{th},\forall t\in[n,n+W],$$

then the system will provides an "electricity-stealing" warning. When there is any real load which is higher than the alarm threshold, the system will automatically reset the starting point of the alarm window. The procedure of our anti-stealing suspects detection is shown in figure 1.

Based on the characteristics of electricity-stealing behavior as well as local circumstances, the defender can set up reasonable alarming threshold index *r*, such as 1/2 or 2/3.

Generally speaking, the electricity-stealing behavior is sustainable during a considerable time interval, since the operation is relatively complicated in practice. Therefore, the defender can set up reasonable alarm window length *W* based on historical record in specific cases. If *W* is too small, the probability of forecasting electricity-stealing behavior will decrease. On the other hand, if the *W* is too large, the probability of missing alarm will increase.

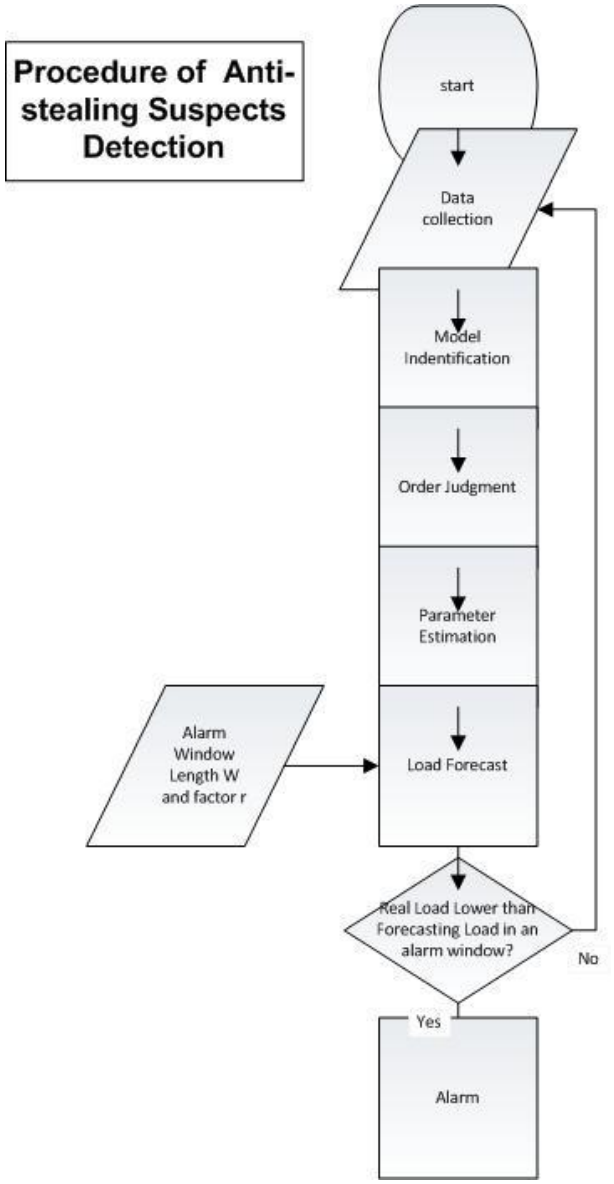

Figure 1. Procedure of anti-stealing suspects detection

Once the system gives an alarming signal, the daily data is then regarded as unconvincing, which means that the data with electricity-stealing behavior will not be recorded in the database. Meanwhile, only the convincing data will be recorded to forecast future load.

CASE STUDY

In this section, we take the data of a tire manufactory company as an example. The AMI collects the load data from this company every 15 minutes. Let $X_{i,j}$ represent the j-th collection of the data in the i-th day, and thus we will get 96 points each day. The data from a single day are represented as $\{X_{i,1}, X_{i,2}, \ldots, X_{i,96}\}$. Then we assemble the convincing data for 50 days in a row in the data set X:

$$X = \{X_{i,j}\}, i \in [1,50],\ j \in [1,96].$$

Figure 2 shows the sample of load in the last 5 days.

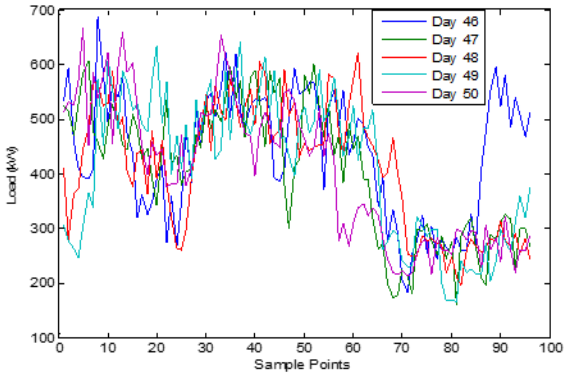

Figure 2. Load data during the last 5 days

Scrutinizing the periodicity of the data, we are clear that the data is periodic with T=96, which means that the present load is interrelated to the load of the previous day. In fact, if we present the data set as

$$Z = \{Z_{i*96+j}\}, i \in [1,50],\ j \in [1,96],$$

where Z normally has periodicity with 96 or 672, because the load feature of a regular user varies daily or weekly.

Letting $Y_j = \{X_{i,j}\}$, $i \in [1,50]$, $j \in [1,96]$, we have

$$Y_j(1) = X_{1,j}, Y_j(2) = X_{2,j}, \cdots Y_j(50) = X_{50,j}.$$

Before we model the data set Y_j by time-series method, identifying the model and judging the orders are still essential. Estimating the autocorrelation function and partial correlation function of Y_j, we find that both of them tail off, and thus Y_j suits the AR(p) model. Aiming at lowering the order of the model, we determine 3 as a reasonable value of p. So the equation of the model of $Y_j(t)$ is

$$Y_j(t) - \varphi_{j,1}Y_j(t-1) - \varphi_{j,2}Y_j(t-2) - \varphi_{j,3}Y_j(t-3) = \alpha_{j,t}.$$

Then we can solve the estimation of the parameters $\{\varphi_{j,1}, \varphi_{j,2}, \varphi_{j,3}, \sigma_\alpha^2\}$ by solving Yule-Walker equation of AR(3).

Based on the load values during last 3 days $\{Y_j(48), Y_j(49), Y_j(50)\}$, we can predict the load in the 51-th day $Y_j(51)$. In this case, we determine r as 2/3 and W as 2 hours (8 sampling points in a row). Then we aggregates $Y_j(51), \forall j \in [1,96]$ in the forecasting data set, and compare with the real load curve of day 51 in figure 3. The black line represents the predicted load curve; the blue line represents the collected load curve from AMI; the red line represents the alarm threshold due to load forecast.

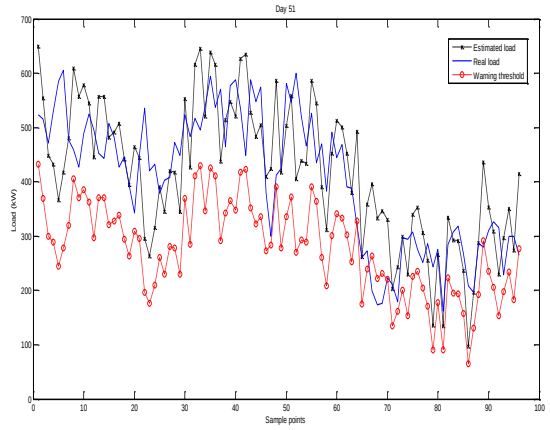

Figure 3 Estimated load and real load in day 51

In figure 3, we can see that the load value from AMI is close to the estimated load value. No collected data is able to cause electricity-stealing alarm; thereby the data is regarded convincing and is recorded in the database.

According to the updated set *X*, we apply the same method to predict the load curve in day 52, which is presented in figure 4.

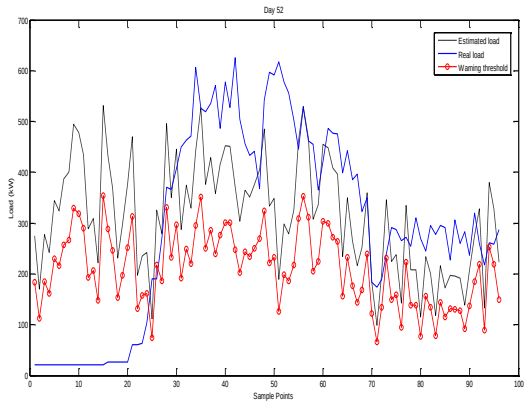

Figure 4 Estimated load and real load in day 52

It is obvious to see that the first 24 points (6 hours) in the collected data of AMI are lower than the alarm threshold, and the system should produce an alarm signal. Meanwhile, the following 72 points are close to the forecasting data. From our method, it is reasonable to judge that this customer has a large probability to steal electricity from 0 a.m. to 6 a.m. in day 52. The data is regarded as unconvincing and should be precluded from the data set *X*. We then examine the suspect through further local investigation, and the fact proves that the client indeed steals the electricity during this time, according with the judgment via our anti-stealing technique.

CONCLUSIONS

This paper presents a novel application of advanced metering infrastructure for distribution networks, which can defend electricity-stealing from a large set of users via short-term load forecast instead of traditional on-site checking. Proved to be applicable by statistical data analysis from the practical load database of clients, this scalable technique, tracking the electricity-stealing suspects at distribution networks, brings the emerging smart grid a labour-saving defence of electricity-stealing.

REFERENCES

[1] W. Zhou, R. Zhu and J. Wang, 2004, “GSM-based monitoring and control system against electricity stealing”, *Electric Power Auto. Equip.* vol. 2, 64-66.

[2] M. Shen, 2006, "Theories for Anti-stealing Electricity", *Power System Tech.* vol. 30, 236-238.

[3] B. Zhao, Y. Lv and H. Zou, 2009, “Design of a new anti-stealing electricity device”, *Power System Protection and Control*, vol. 37 (23), 116-118.

[4] J. Wang et al., 2008, “The present situation and development trend of anti electric stolen function of power demand information acquisition system”, *Power System Tech.* vol. 32, 177-178.

[5] S. McLaughlin, D. Podkuiko, and P. McDaniel, 2010, “energy theft in the advanced metering infrastructure”, *Critical Information Infrastructures Security*, vol. 6027, 176-187.

[6] L. Feng and J. Qiu, “electrical load forecasting based on load patterns”, *Power System Tech.* vol. 29 (4), 23-26.

[6] J. Yang, 2008, “ARIMA time series modeling and forecasting of electricity consumption”, *Chinese J. of Eng. Math.* vol. 25 (4), 611-615.