MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

# Electricity Theft Detection Using Smart Meter Data

Sahoo, S.; Nikovski, D.N.; Muso, T.; Tsuru, K.

TR2015-005 January 2015

**Abstract**

Electricity theft is a major concern for the utilities. With the advent of smart meters, the frequency of collecting household energy consumption data has increased, making it possible for advanced data analysis, which was not possible earlier. We have proposed a temperature dependent predictive model which uses smart meter data and data from distribution transformer to detect electricity theft in an area. The model was tested for varying amounts of power thefts and also for different types of circuit approximations. The results are encouraging and the model can be used for real world application.

*2015 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*

# Electricity Theft Detection Using Smart Meter Data

Sanujit Sahoo
Censio
Boston, MA, USA
sanujitsahoo@gmail.com

Daniel Nikovski
Mitsubishi Electric Research Labs
Cambridge, MA, USA
nikovski@merl.com

Toru Muso and Kaoru Tsuru
Mitsubishi Electric Corporation
Ofuna, Kamaruka, Japan
{Muso.Toru@aj|Tsuru.Kaoru@ah}
.MitsubishiElectric.co.jp

***Abstract*—Electricity theft is a major concern for the utilities. With the advent of smart meters, the frequency of collecting household energy consumption data has increased, making it possible for advanced data analysis, which was not possible earlier. We have proposed a temperature dependent predictive model which uses smart meter data and data from distribution transformer to detect electricity theft in an area. The model was tested for varying amounts of power thefts and also for different types of circuit approximations. The results are encouraging and the model can be used for real world application.**

***Keywords—Electricity Theft Detection; Smart Meter; Predictive Model***

## I. INTRODUCTION

Utilities incur great financial losses due to electricity thefts. It has been estimated that in the US itself, electricity worth about $ 6 billion (1% to 3% of total revenue) is stolen annually [1]. In many other countries, the losses due to power theft, as a percentage of the total power generated are much higher [2].

There are different methods used by customers to steal power from the grid [3]. A very common approach involves bypassing the meter completely [4]. This is done by directly connecting the mains of the house to the low voltage grid. This method requires lineman skills and can be very dangerous if the connection is not proper. Another common approach is to tamper with the meter installed in the house [5]. There are many ways of doing this. One of them is to short the ends of the meter installed at the house. By doing so, the user ensures that the current flowing into the house does not flow through the meter and hence the meter records very low usage. Another way is to regulate the supply voltage by disconnecting the neutral from the feeder and using a separate neutral for the return path [6]. As a result of this, the energy meter assumes that the voltage between the connected phase and the new neutral is zero, implying that the total energy consumed is zero. Other methods include tampering with the meters so that the measurements taken by the meter are inaccurate and show lower consumption than actual.

In general, loss of energy in an electricity grid is caused due to two reasons. One of them is due to the circuit itself. These losses are called technical losses (TL) and include losses due to power dissipation in resistive components, leaks due to improper isolation, etc. Since these are due to components in the circuit, the loss values depend on the current flowing in the circuit. The other category includes losses due to electricity thefts and is called non-technical losses (NTL).

Popular methods for electricity theft detection include load profile analysis of customers to detect abnormal energy consumption patterns [5]-[7]. But these methods cannot be used to detect energy thefts where there is complete bypass of meters. In such cases, losses calculated using energy balance between the energy supplied from the distribution transformer (DT) and the energy consumed by users is used for theft detection. The TL and NTL component of these losses have to be estimated accurately to detect power thefts. In [8], a model to calculate TL has been developed, but it requires topological information of the primary and secondary networks.

This work is an extension of the work mentioned in [4]. A predictive model to calculate TL was developed in [4] without using the actual topology information of the network. We have improved the predictive model to estimate technical loss in a branch of the distribution network by taking into consideration the temperature dependency of resistances in the circuit. In addition, we have tested our models on linear circuits as well. Distribution feeders mostly have linear circuits and the successful application of the NTL model on these circuits would validate their usefulness in the real world scenario. Finally, we have tested our models for different amount of power thefts to find the minimum amount of power theft that our models can detect with confidence.

## II. POWER THEFT DETECTION BASED ON A TECHNICAL LOSS MODEL

One effective way of estimating non-technical losses in the distribution network is by correctly estimating technical losses in the network and then subtracting it from the total loss in the network. A novel way for estimating technical losses was proposed in [4]. The work assumed that the utility received data from every smart meter and DT every 30 minutes. For any particular time interval k, information about

This work was done when Sanujit Sahoo was employed by Mitsubishi Electric Research Labs.

total energy consumption for that user group ($E_{DT,k}$) was obtained from each DT and the individual smart meters provided information about the consumption of individual users ($E_k^i$). The total losses for such a user group for that time interval ($L_k$) can be calculated using (1).

$$L_k = E_{DT,k} - \sum_{i=1}^{n} E_{i,k} \quad (1)$$

where n is total number of smart meter

If the technical loss estimate ($\hat{L}_k^{TL}$) can be calculated for the user group for the same time interval, the NTL estimate can be calculated as shown in (2).

$$\hat{O}_k = L_k - \hat{L}_k^{TL} \quad (2)$$

The method in [4] approximates the distribution circuit downstream of the DT to be radial in nature as shown in Fig. 1 [9]. Since it assumed the resistances in the network to be constant, it is referred to as the Constant Resistance Technical Loss Model (CRTLM).

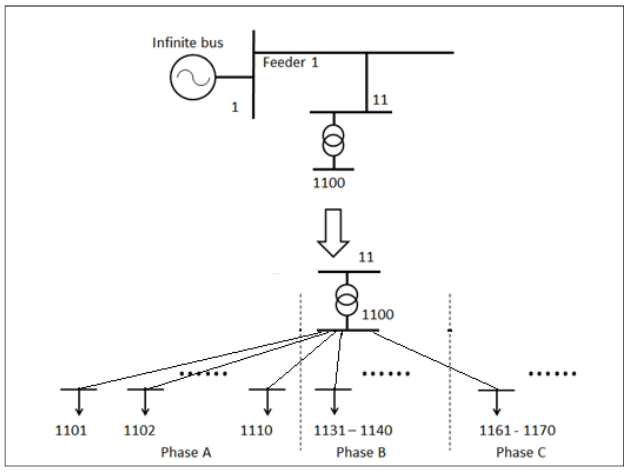


Fig. 1: Radial approximation of distribution circuit

### A. *Constant Resistance Technical Loss Model (CRTLM)*

The detailed derivation of the method can be found in [4]. A brief description is given here. The following variables were defined for the simplified circuit:

| | |
|---|---|
| $R_i$ | The actual resistance of the line to user i (assumed to be constant) |
| $\hat{R}_i$ | The estimated resistance of the line to user i |
| $I_{i,k} = I_i(t_k)$ | The measured instantaneous current of branch i at the end of time interval k |
| $L_{i,k}$ | Actual technical loss of branch i during time interval k |
| $\hat{L}_{i,k}$ | Estimated technical loss of branch i during time interval k |
| $L_k$ | Total loss during time interval k for all branches (users), obtained by means of power balance between the DT and all legal users |
| $L_k^{TL}$ | Technical loss during time interval k for all branches (users). When there is no theft, $L_k^{TL} = L_k$ |
| $l_0$ | Non-ohmic technical loss (time independent) |
| $\hat{O}_k$ | The estimated non-technical loss (NTL) during time interval k |

Total loss for all users during time interval k is given as:

$$\hat{L}_k = \sum_{i=1}^{n} \frac{I_i(t_k)^3 - I_i(t_{k-1})^3}{3s_{i,k}} R_i + l_0 \quad (3)$$

During the no theft period, the total loss is equal to the total technical loss. It is clear that in order to calculate the technical losses, the values of various resistances need to be calculated. Equation (3) can be written for all values of k in a matrix format and estimates of the resistances can be calculated using Moore–Penrose pseudoinverse as shown in (4).

$$\hat{R} = (H^T H)^{-1} H^T L \quad (4)$$

where

$$H = \begin{bmatrix} \frac{I_1(t_2)^3 - I_1(t_1)^3}{3s_{1,2}} & \frac{I_2(t_2)^3 - I_2(t_1)^3}{3s_{2,2}} & \dots & \frac{I_n(t_2)^3 - I_n(t_1)^3}{3s_{n,2}} & 1 \\ \frac{I_1(t_3)^3 - I_1(t_2)^3}{3s_{1,3}} & \frac{I_2(t_3)^3 - I_2(t_2)^3}{3s_{2,3}} & \dots & \frac{I_n(t_3)^3 - I_n(t_2)^3}{3s_{n,3}} & 1 \\ \dots & \dots & \dots & \dots & 1 \\ \frac{I_1(t_m)^3 - I_1(t_{m-1})^3}{3s_{1,m}} & \frac{I_2(t_m)^3 - I_2(t_{m-1})^3}{3s_{2,m}} & \dots & \frac{I_n(t_m)^3 - I_n(t_{m-1})^3}{3s_{n,m}} & 1 \end{bmatrix}$$

$$L = [L_2 \quad L_3 \quad \dots \quad L_m]^T$$

$$\hat{R} = [\hat{R}_1 \quad \hat{R}_2 \quad \dots \quad \hat{R}_n \quad l_0]^T$$

Once the resistance estimates have been calculated the NTL estimate can be calculated using the following equation.

$$\hat{O}_k = L_k - \sum_{i=1}^{n} \frac{I_i(t_k)^3 - I_i(t_{k-1})^3}{3s} \hat{R}_i - l_0 \quad (5)$$

### B. *Temperature Dependent Technical Loss Model (TDTLM)*

In our work, we have improved the CRTLM by making the resistance temperature dependent. We assume that the atmospheric temperature would be available every 30 minutes and would be constant for all users in a group. It is known that the resistance of a material is linearly dependent on its temperature as represented in (6). The coefficients $a_i$ and $b_i$ are different for different materials. By making the model temperature dependent, the estimation problem now changes to estimating the coefficients $a_i$ and $b_i$ for each resistance in the circuit downstream of the DT. Substituting the value of $R_i$ in equation 3 we obtain (7) which can be expanded to get (8). The estimates of the coefficients are then obtained by using the least squares methods. The Moore–Penrose pseudoinverse (9) is used to calculate the coefficients.

$$R_i = a_i T + b_i \quad (6)$$

$$\hat{L}_k = \sum_{i=1}^{n} \frac{I_i(t_k)^3 - I_i(t_{k-1})^3}{3s_{i,k}} (a_i T_k + b_i) + l_0 \quad (7)$$

$$\hat{L}_k = \sum_{i=1}^{n} \left\{ \frac{I_i(t_k)^3 - I_i(t_{k-1})^3}{3s_{i,k}} T_k \right\} a_i + \sum_{i=1}^{n} \frac{I_i(t_k)^3 - I_i(t_{k-1})^3}{3s_{i,k}} b_i + l_0 \quad (8)$$

$$\widehat{AB} = (H_{new}^T H_{new})^{-1} H_{new}^T L \quad (9)$$

where

$$H_{new} = [H_{1new} \quad H_{2new} \quad O]$$

$$H_{1new} = \begin{bmatrix} \frac{I_1(t_2)^3 - I_1(t_1)^3}{3s_{1,2}} T_2 & \frac{I_2(t_2)^3 - I_2(t_1)^3}{3s_{2,2}} T_2 & \dots & \frac{I_n(t_2)^3 - I_n(t_1)^3}{3s_{n,2}} T_2 \\ \frac{I_1(t_3)^3 - I_1(t_2)^3}{3s_{1,3}} T_3 & \frac{I_2(t_3)^3 - I_2(t_2)^3}{3s_{2,3}} T_3 & \dots & \frac{I_n(t_3)^3 - I_n(t_2)^3}{3s_{n,3}} T_3 \\ \dots & \dots & \dots & \dots \\ \frac{I_1(t_m)^3 - I_1(t_{m-1})^3}{3s_{1,m}} T_m & \frac{I_2(t_m)^3 - I_2(t_{m-1})^3}{3s_{2,m}} T_m & \dots & \frac{I_n(t_m)^3 - I_n(t_{m-1})^3}{3s_{n,m}} T_m \end{bmatrix}$$

$$H_{2new} = \begin{bmatrix} \frac{I_1(t_2)^3 - I_1(t_1)^3}{3s_{1,2}} & \frac{I_2(t_2)^3 - I_2(t_1)^3}{3s_{2,2}} & \dots & \frac{I_n(t_2)^3 - I_n(t_1)^3}{3s_{n,2}} \\ \frac{I_1(t_3)^3 - I_1(t_2)^3}{3s_{1,3}} & \frac{I_2(t_3)^3 - I_2(t_2)^3}{3s_{2,3}} & \dots & \frac{I_n(t_3)^3 - I_n(t_2)^3}{3s_{n,3}} \\ \dots & \dots & \dots & \dots \\ \frac{I_1(t_m)^3 - I_1(t_{m-1})^3}{3s_{1,m}} & \frac{I_2(t_m)^3 - I_2(t_{m-1})^3}{3s_{2,m}} & \dots & \frac{I_n(t_m)^3 - I_n(t_{m-1})^3}{3s_{n,m}} \end{bmatrix}$$

$$O = [1 \quad 1 \quad \dots \quad 1]^T$$

$$L = [L_2 \quad L_3 \quad \dots \quad L_m]^T$$

$$\widehat{AB} = [\hat{a}_1 \quad \hat{a}_2 \quad \dots \quad \hat{a}_n \quad \hat{b}_1 \quad \hat{b}_2 \quad \dots \quad \hat{b}_n \quad l_0]^T$$

Once the coefficients are obtained, the TL (8) and NTL estimates (5) are calculated and a decision regarding power theft can be taken.

## III. EXPERIMENTAL SET-UP

We have verified the performance of TDTLM on a system similar to the one mentioned in [4]. The test branch of the distribution system consists of 30 users with 10 users being connected to each phase. The data from each user was generated using a smart meter simulator. Each user's smart meter sends measurement data back to the utility every 30 minutes. The theft analysis has been done on 10 users connected to phase A. One of the users was connected to the grid illegally. The rogue user stole electricity by bypassing the smart meter and the amount varied roughly between 1% and 10% of the total energy consumed by the 10 users. Data was collected from these users for a period of six days (144 hours). It was assumed that there was no theft for the first four days. Data collected on the first two days was used for calculating parameters of the predictive model used for NTL estimation. This model was then tested using the data collected over the next four days.

The smart meter simulator calculated the power quality measurements like power consumed, instantaneous current and voltage readings every 30 minutes for every user. The load profile for each user (Fig. 2) was generated from [10] in the same manner as mentioned in [4].

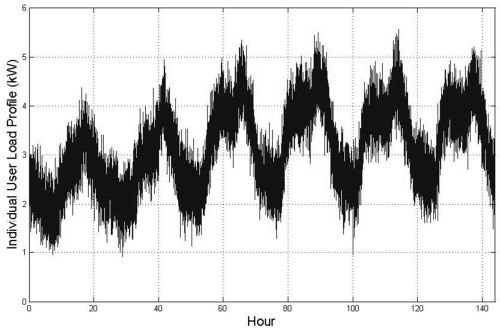


Fig. 2: Load profile of one of the users

The simulator took the load profiles and the temperature profile for the area as inputs and calculated the state of the nodes connected to each user (instantaneous current, voltage measurements, etc.) by executing power flow calculations every 10 seconds. It was assumed that copper wires were used in cables and the resistance values were calculated by using the temperature coefficients of resistance for copper. The power consumption values were aggregated every 30 minutes and communicated back to the utility along with other instantaneous measurements.

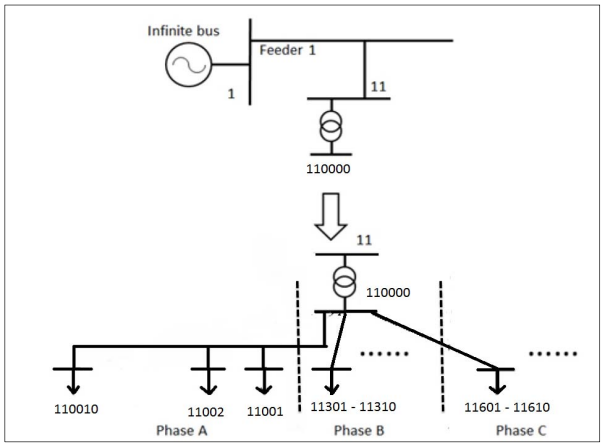


Fig. 3: Linear approximation of distribution circuit

In this work, the circuit for the distribution system has been approximated in two ways. In the first case, a radial circuit is considered. Downstream of the DT, all users were directly connected to the bus as shown in Fig. 1. The technical loss calculation for CRTLM and TDTLM is done assuming this type of circuit approximation. The second way

of approximating the circuit is to assume a linear circuit as shown in Fig. 3. In Fig. 3, each phase of the bus downstream of the DT is connected to 10 users (similar to the previous case) but in a linear way. This is a more realistic approximation of the actual distribution feeder circuit. Successful application of the technical loss models on this type of linear circuit would indicate the robustness of the proposed models, hence implying that they could be deployed to detect power theft using real data from actual smart meters.

For each loss model and each type of circuit approximation, the data from the first two days (no theft) is used to train the predictive model. A well-trained model for this period should be able to estimate the technical losses very well and the root mean square error (RMSE) for the training set should be low. The non-technical losses for this period should be low and be distributed around zero. The data from the next two days (no theft) is used to validate the predictive model. Ideally, a good predictive model would have a low RMSE for this validation data set too. The error would be slightly higher, since this data is completely new to the model but a good performance (low RMSE) would imply that the model did not over fit the training data. The maximum non-technical loss in the validation set is used as a threshold to separate the theft cases from the non-theft cases in the third data set. The third data set contained data collected on the last two days. During this period, one of the users was stealing power by bypassing the meters. If the calculated non-technical loss exceeded the threshold calculated during the validation step, then it was assumed there was theft in the user group. The power theft detection algorithms were tested for different amounts of power thefts in order to evaluate the algorithms and check the minimum amount of theft that they could catch with confidence. The power theft percentage was varied from roughly 1% of total consumption to 10% and the performance of the algorithm was seen.

## IV. Experimental Results

Data was collected every 30 minutes from the smart meter simulator for six days. During each of the three phases – training, validation and testing, 96 data points were obtained. Non-technical losses were estimated for each time instant and have been discussed.

### A. Radial Circuit

In case of CRTLM, 11 parameters (10 resistance and 1 non ohmic non technical loss parameter ($l_0$)) were first estimated from these 96 data points while for TDTLM 21 parameters (10 $a_i$s, 10 $b_i$s and $l_0$). The TL estimates for both models were very close to the actual TL in the system and hence the RMSE was very low. The RMSE for TDTLM was found to be lower than that of CRTLM as shown in Fig. 4. Similarly, the RMSE for both models was calculated using data from the validation set. The RMSE for both models was low although it increased slightly in comparison with the value obtained with training data. This proved that the models were robust and didn't over fit the training data. The RMSE of the TDTLM was again lower than the RMSE of the CRTLM.

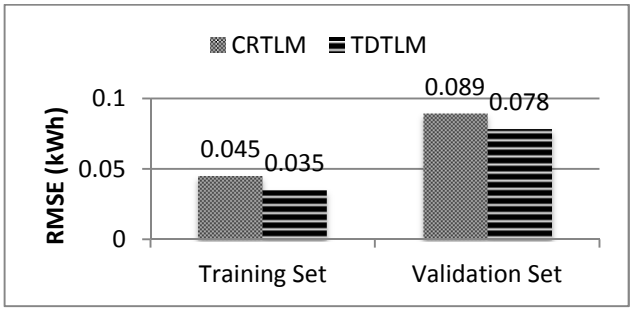


Fig. 4: RMSE comparison for radial circuit approximation

The non-technical loss estimates for the constant resistance model during the training phase was obtained as in [4]. On similar lines, once the parameters of the temperature dependent predictive model were obtained, the non-technical loss (NTL) estimate for every 30-minute interval was calculated for the three phases. The distribution of the NTL estimates for the different periods has been shown in Fig. 5.

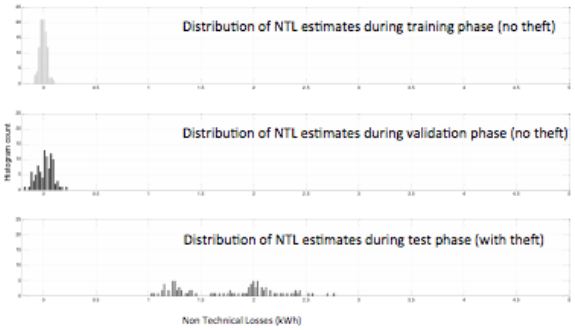


Fig. 5: Distribution of NTL estimates during different phases for temperature dependent model

It can be seen that during the no theft period, the NTL estimates are low and distributed around zero. But when power theft occurs (Fig. 5 shows the distribution for case when the power theft is 10% of the total power consumption by all users), the estimates are no longer small values. In addition, the NTL estimates are no longer distributed around zero. The maximum NTL estimate calculated during the validation phase was taken as the threshold and whenever an NTL estimate was more than this threshold, it was assumed that there was power theft in the user group. During the testing phase, all the NTL estimates were greater than the threshold and this can be seen in Fig. 5. Hence the power theft detection rate was 100% when the power theft was 10% of the total power consumption.

The detection rate of both the models was calculated for different amount of power thefts and is shown in Fig. 6. For both the models power theft can be detected 100% of times

when the power theft percentage is greater than 4%. For lower percentage of power theft, the detection rate falls, but in all cases, the performance of TDTLM is either better than or similar as that of the CRTLM.

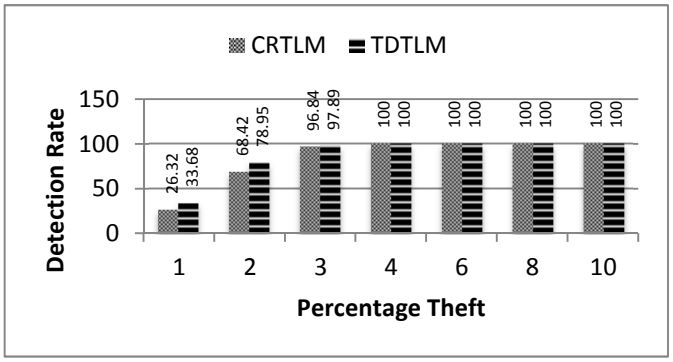


Fig. 6: Detection rates for different amounts of power theft in radial circuits

### B. *Linear Circuit*

The process described for radial circuits was repeated on the new linear circuit type to obtain the results. In this case also, the technical loss estimates were very close to actual technical loss and the RMSE was very low with the training data.

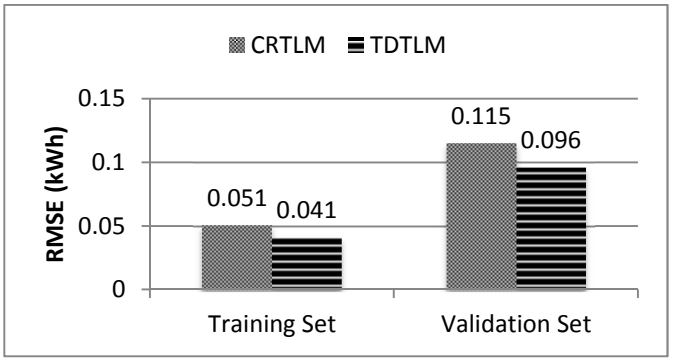


Fig. 7: RMSE comparison for linear circuit approximation

As expected, the RMSE for each linear circuit model was greater than the RMSE obtained for the radial circuit. With the validation data also, the RMSE followed an expected pattern. The RMSE value did go up but was not high. This confirmed the robustness of the models. In terms of theft detection rate, both the models performed very accurately when the power theft percentage was more 4% of the total consumption. It can be seen again (Fig. 7 and Fig. 8) that TDTLM performs better than the CRTLM. It can be inferred from Fig. 6 and Fig. 8 that the models performed slightly better on radial circuits but overall their performances is very good. This proves that these predictive models work very well even with the linear approximation of distribution networks and hence can be applied to real data.

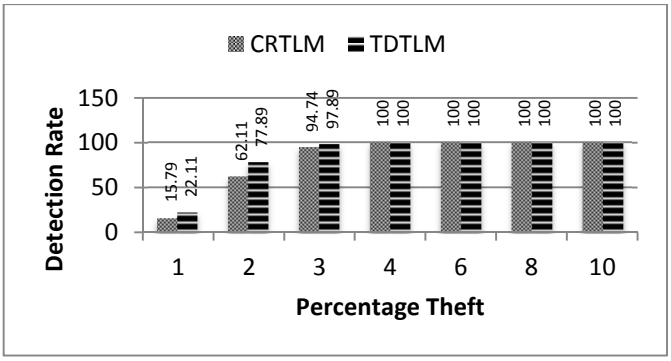


Fig. 8: Detection rates for different amounts of power theft in linear circuits

## V. CONCLUSIONS

In this work, we have fine-tuned the predictive model for calculating technical loss for a branch in the distribution network by incorporating the temperature dependency of resistances in a distribution network. The new model performed better than the constant resistance model and gave better power theft detection rates. In addition, we tested the proposed predictive models on distribution circuits, which were approximating them as linear circuits. The performance of our models on these circuits was also very good which implies that they can used to detect electricity thefts using data from actual smart meters.

## REFERENCES


[1] "Electricity Thefts Surge in BadTimes," March 16, 2009, USA Today,via Factiva, © 2009 USA Today. Available: http://usatoday30.usatoday.com/money/industries/energy/2009-03-16-electricity-thefts_n.htm

[2] "Fighting Electricity Theft with Advanced Metering Infrastructure", March 2011 ECI Telecom Ltd.

[3] "Achieving high performance with theft analytics", August 29, 2011, Accenture

[4] D. Nikovski, Z. Wang, A. Esenther, H. Sun, K. Sugiura, T. Muso, and K. Tsuru, "Smart Meter Data Analysis for Power Theft Detection", Machine Learning and Data Mining in Pattern Recognition, Lecture Notes in Computer Science Volume 7988, 2013, pp 379-389

[5] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines", IEEE Trans. Power Del., vol. 25, no. 2, pp. 1162-1171, Apr. 2010.

[6] S.S.S.R. Depuru, "Modeling, Detection, and Prevention of Electricity Theft for Enhanced Performance and Security of Power Grid," The University of Toledo, Aug. 2012.

[7] J. Nagi, K.S. Yap, S.K. Tiong, S.K. Ahmed, and A.M. Mohammad, "Detection of abnormalities and electricity theft using genetic support vector machines" Proc. IEEE Region 10 Conference TENCON, Hyderabad, India, Jan. 2009, pp. 1–6.

[8] C. C. B. de Oliveira; N. Kagan; A. Meffe; S. L. Caparroz; J. L. Cavaretti, 2001, "A New Method for the Computation of Technical Losses in Electrical Power Distribution Systems", Proceedings ClRED 2001.

[9] T. McAvinew, R. Mulley, "Control System Documentation", ISA, p. 165 (2004)

[10] Electricity Transmission Operational Data, National Grid UK, Available: www.nationalgrid.com/uk/Electricity/Data/Demand+Data