Power Line Carrier (PLC) Signal Analysis of Smart Meters for Outlier Detection

Rakesh Rao
Dept. of Computer Science
Univ. of North Carolina at Charlotte
Charlotte, North Carolina 28223
USA

Srinivas Akella
Dept. of Computer Science
Univ. of North Carolina at Charlotte
Charlotte, North Carolina 28223
USA

Gokhan Guley
Smart Grid Technology
HiSoft Technology Intl.
Charlotte, North Carolina 28210
USA

***Abstract*— A method for identifying outliers among a set of smart meters by using the power line carrier (PLC) signal strength is presented in this paper. The broad goal is to use the PLC signal as a predictor of transmission problems to proactively avert local power outages. The proposed method uses the PLC signal strength measured between the communication node (transformer) and residential smart meters. The paper presents four metrics based on the distribution of signal strengths, with each metric identifying a class of outliers. After detecting the set of outliers using these metrics, we identify the set of representative good meters.**

I. Introduction

Power Line Carrier (PLC) is a standard for carrying data on a conductor used for electric power transmission. PLC, also called Power Line Communication, uses the existing electricity infrastructure to carry communication signals by providing a standardized interface [10]. The principle behind PLC is to modulate a carrier frequency on top of existing 50/60 Hz energy carrier lines [6]. PLC has been in use since at least the 1920s [11], and recently its use is expanding into the electricity distribution area for load management, control of heating, lighting, and air conditioning as well as for broadband communication. With the recent advances in smart grid technology, many utility companies are developing technologies for system control of smart grids and to provide realtime information services for customers. Automatic meter reading (AMR) has been used to retrieve customer data using PLC [9]. PLC has also been used to measure the quality of power [8]. Cavdar [4] has described how PLC can be used in remote detection of illegal electricity usage.

This paper describes a technique for identifying outliers among a set of smart meters by using PLC signals from the smart meters. This approach categorizes outlier meters into four groups depending on the characteristic classes of signal strength behavior exhibited by the smart meters. Each group of outliers is identified by metrics described in Section III. The application of these metrics to identify the outliers and the complementary set of representative meters is described in Section IV.

There has been a lot of research done on the application of PLC to solve various transmission, distribution, and monitoring problems. However not much work has been done on using the PLC signals from smart meters to identify and analyze meter or connection problems. Gross et al. [7] describe a machine learning approach to predict feeder failures that enables preventive maintenance of the feeder cables. Hong et al. [8] describe a method for monitoring power quality using PLC. Chandola et al. [5] define three types of anomalies: point anomalies, contextual anomalies, and collective anomalies. In the first type, an individual instance of data is considered anomalous if it is seen in a state where it has not been seen before in the measurement space. We wish to identify meters considering their past behavior as well and not just the current behavior. Contextual anomalies on the other hand consider the context of the measured data instance before deciding on its anomalousness. Collective anomalies describe a set of related data instances as anomalous with respect to the entire data set. The individual data elements may not be anomalies on their own but their occurrence together as a collection is anomalous. Catterson et al. [2] describe a conditional anomaly detection technique used to model the behavior of a transformer. Their case study considers both external and internal factors while modeling the transformer behavior. Our approach aims at identifying anomalies among a set of smart meters considering the past behavior of the meters. Some work has been done in the area of anomaly detection techniques for transformers [2], [3]. However nothing has been specifically developed in the PLC arena to identify anomalies among a set of smart meters. The identification of outliers among the set of smart meters would help the utility crews to do preemptive testing and maintenance for meter or line problems.

II. Signal Strength Analysis

We analyzed the Power Line Carrier (PLC) signals from about 15,000 smart meters in residential buildings covering a period of 54 weeks. The data was obtained from a leading power utility company. Each week's data is a once-a-week snapshot of signal strength for these smart meters; the smart meter data is actually available at 15-minute intervals. Initial analysis of the mean signal strength and standard deviation of signal strength (Table I) indicated that there is not much weekly variation for the aggregate set of meters. Therefore we looked into the distribution of signal strengths to help us gain better insights. We focused on a data-driven approach

978-1-4577-1703-1/11/$26.00 ©2011 IEEE

TABLE I

MEAN AND STANDARD DEVIATION OF THE PRIMARY AND SECONDARY SIGNAL STRENGTHS OF ALL METERS. DATA IS FOR A REPRESENTATIVE NINE WEEK PERIOD, WITH EACH ROW REPRESENTING A WEEK.

Week	Mean Primary	Std. Dev. Primary	Mean Secondary	Std. Dev. Secondary
30510	-9.9902993	5.6266307	-5.266214	5.7103121
31110	-10.016622	5.6492335	-5.203823	5.765126
31710	-9.9241845	5.5695217	-5.1824442	5.6638004
32410	-10.000139	5.6484687	-5.2740288	5.6857549
33110	-10.024969	5.5284079	-5.3391594	5.6307472
40610	-10.082457	5.5063082	-5.374469	5.5531313
41310	-10.046183	5.7097943	-5.2386104	5.7648316
41910	-10.018009	5.5611871	-5.2639745	5.6245254
42710	-9.9970876	5.563228	-5.2173913	5.6360458

that minimizes the use of human selected threshold values to identify anomalous signals.

A. Signal Probability Distribution

We examined the frequency distribution of signal strengths[1] to get a overview of how frequent the anomalous signals (i.e., with low signal strengths) occur across the set of smart meters. The measured signal strengths from smart meters are discretized, and measured in steps of 6 dB with values ranging from 6 to -84 dB. The signal data from a smart meter has a primary signal and a secondary signal. We use the primary signal strength for the analysis; when the primary signal is unavailable, the secondary signal can be used by the system.

Using the frequency distribution of signal strengths (Table II), we computed the probability distribution of signal strengths with add-one smoothing (Table III). Let S_i be the i^{th} signal strength level, N_i be the number of occurrences of signal strength S_i in that week, T be the total number of meter readings for that week, and D be the number of distinct signal strengths for that week. Then the probability $P(S_i)$ of signal strength S_i in a given week is

$$P(S_i) = \frac{(N_i + 1)}{(T + D)}. \tag{1}$$

Since anomalous meter readings (i.e., with low signal strengths) occur less frequently, signal strengths with low probability values are more likely to be indicative of potentially faulty connections. We therefore compute a probability score for each meter based on the signal strengths that it has over the entire observation period.

To summarize, the probability score is calculated in the following steps:

- *Step 1:* For each week, count the number of signal readings N_i of each signal strength S_i.
- *Step 2:* For each week, compute the probability $P(S_i)$ of each signal strength S_i (Equation 1).
- *Step 3:* Let S_w^j be the signal strength of meter j in week w. Then the meter probability score P_j^m for the j^{th} meter over W weeks is given by

$$P_j^m = \sum_{w=1}^{W} P(S_w^j). \tag{2}$$

  Here for each meter, we sum up the probability of its signal strength for each week over all the W weeks for which it has data to obtain its probability score.

Meters with low total probability scores are more likely to have potentially weak signals than meters with high probability scores. However probability scores alone are not enough to identify outlier meters. For example, even if a meter fluctuates between a strong signal range and weak signal range in successive weeks, the corresponding probability score could be larger than a meter with a stable strong signal over successive weeks. We describe a data driven approach to handle this and other possible scenarios in Section III.

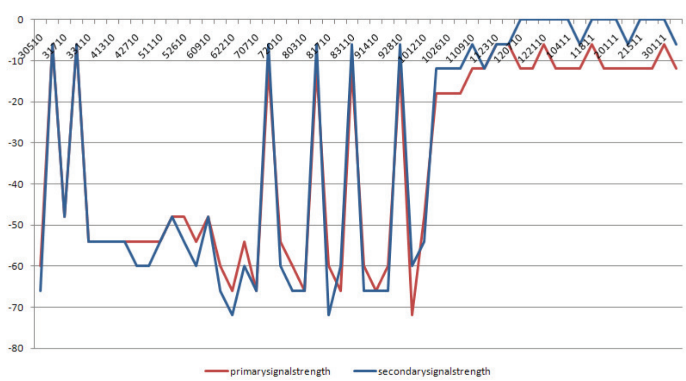

Fig. 1. Plot of primary and secondary signal strength for a meter exhibiting jumpy characteristics. The vertical axis represents signal strength and horizontal axis represents time (with dates in month-day-year format).

[1]Note that although we refer to a meter's signal strength, the signal strength in fact depends on the meter, the node (i.e., transformer), the cable, and the various connections. We use "meter" to include all these for brevity.

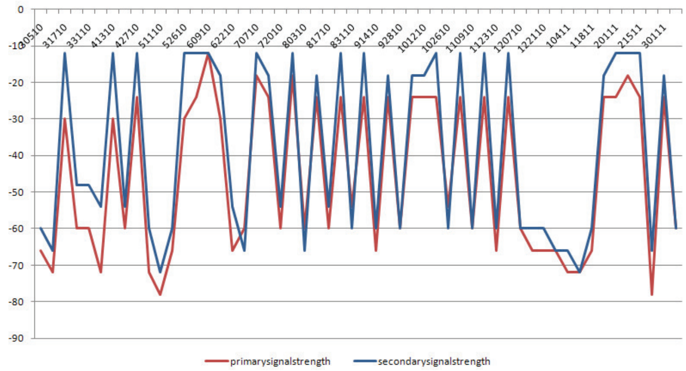

Fig. 2. Plot of primary and secondary signal strength against time for a meter exhibiting jumpy characteristics.

III. Metrics for Identifying Outliers

A. Jumpy Signal Identification

Here we present a metric that identifies meters that experience multiple frequent changes in signal strength, and have signal strengths in the low signal strength ranges. The Jumpiness score J_j of a meter j is given by

$$J_j = \sigma_j \cdot \frac{1}{P_j^m} \cdot \frac{(W_c + 1)}{(W_n + 1)}, \tag{3}$$

where σ_j is the standard deviation of the signal strength over the data period for meter j, P_j^m is the meter probability score calculated in Equation 2, W_c is the number of weeks signal strengths have changed for the meter j, and W_n is the number of weeks signal strengths have not changed for the meter j. The higher (i.e., more positive) the jumpiness score of a signal, the more likely it is to have a potentially weak signal. The intuition behind this formula is as follows. First, the standard deviation of signal strength readings for a signal is closely related to the jumpiness of the signal strength readings. Second, a jumpy signal has several signal strength readings that correspond to low probability states. Hence the overall signal strength probability score of a jumpy signal should be low, and its reciprocal will be correspondingly high. Finally, jumpy signals will exhibit more frequent changes in signal strength, and that is captured by the ratio of the number of weeks with changes in signal strengths to the number of weeks where signal strengths have not changed (with add-one smoothing). Figures 1 and 2 show plots of signal strengths for meters with high jumpiness scores.

B. Flat Signal Identification

Flat signals are signals whose strength is stable and consistently in low signal strength ranges. Flatness in meter behavior is identified by computing the Flatness score F_j of each meter j using the formula

$$F_j = \mu_j \cdot \frac{1}{P_j^m} \cdot \frac{(W_n + 1)}{(W_c + 1)} \tag{4}$$

where μ_j is the mean of the signal strength over the data period for meter j, P_j^m is the probability score for meter j calculated in Equation 2, W_c is the number of weeks signal strengths have changed for the meter j, and W_n is the number of weeks signal strengths have not changed for the meter j. The lower (i.e., more negative) the flatness score of a signal, the more likely it is to have a potentially weak signal. The intuition behind this formula is as follows. First, the mean of signal strength readings for a meter is directly related to the average signal strength readings; the lower the signal strengths, the lower the mean. Second, a meter with a large number of low signal strength readings has several signal strength readings that correspond to low probability states. Hence the overall signal strength probability score of a flat signal should be low, and its reciprocal will be correspondingly high. Finally, flat signals will exhibit less frequent changes in signal strength, and that is captured by the ratio of the number of weeks that the meter signal strength has not changed to the number of weeks with changes in the signal strength. The flatness measure seems to work well at identifying flat signals. Figures 3 and 4 show plots of signal strengths for meters with low flatness scores.

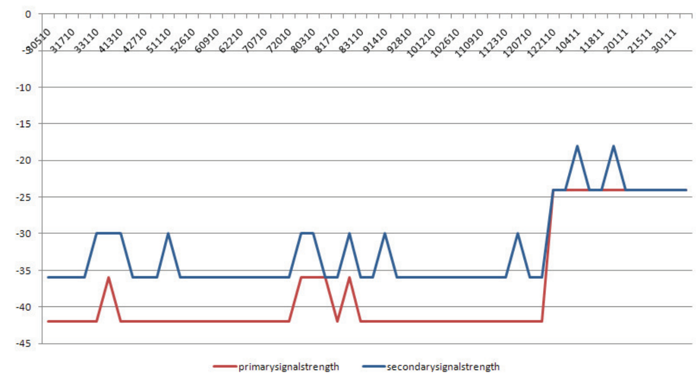

Fig. 3. Plot of primary and secondary signal strength against time for a meter exhibiting flat characteristics.

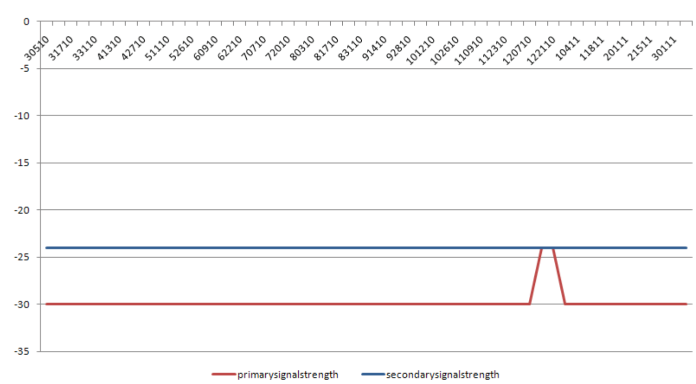

Fig. 4. Plot of primary and secondary signal strength against time for a meter exhibiting flat characteristics.

C. Large Drop Signal Identification

Since a rapid large drop in signal strength can be indicative of a connection problem, we ranked the signals using their

TABLE II

FREQUENCY DISTRIBUTION OF SIGNAL STRENGTH S_i OVER ALL METERS. EACH COLUMN REPRESENTS THE NUMBER OF OCCURRENCES OF SIGNAL S_i IN A PARTICULAR WEEK.

Week	6	0	-6	-12	-18	-24	-30	-36	-42	-48	-54	-60	-66	-72	-78	-84
30510	65	1058	5453	5362	2125	325	33	7	1	1	-	1	1	-	-	-
31110	54	1062	5498	5266	2172	344	33	4	4	1	-	-	-	1	-	-
31710	57	1036	5574	5296	2043	329	33	6	1	1	-	1	-	-	-	-
32410	65	1065	5433	5334	2080	369	34	6	2	-	-	1	-	-	-	-
33110	49	972	5528	5446	2027	357	32	4	1	-	1	1	-	-	-	-
40610	53	919	5480	5443	2091	339	26	6	-	-	1	-	-	1	-	-
41310	71	1069	5419	5276	2161	382	33	6	1	-	3	-	-	-	-	-
41910	70	981	5466	5499	2028	358	26	4	2	1	1	1	-	-	-	-
42710	64	1001	5526	5345	2114	336	26	6	1	1	1	-	-	-	-	-

TABLE III

PROBABILITY DISTRIBUTION OF SIGNALS OVER ALL METERS. $P(S_i)$ IS THE PROBABILITY OF SIGNAL STRENGTH S_i IN A PARTICULAR WEEK.

Week	P(6)	P(0)	P(-6)	P(-12)	P(-18)	P(-24)	P(-30)	P(-36)	P(-42)	P(-48)	P(-54)	P(-60)	P(-66)	P(-72)	P(-78)	P(-84)
30510	0.0046	0.0733	0.3775	0.3712	0.1472	0.0226	0.0024	0.0006	0.0001	0.0001	7E-05	0.0001	0.0001	7E-05	7E-05	7E-05
31110	0.0038	0.0735	0.3804	0.3644	0.1503	0.0239	0.0024	0.0003	0.0003	0.0001	7E-05	7E-05	7E-05	0.0001	7E-05	7E-05
31710	0.004	0.0721	0.3874	0.3681	0.142	0.0229	0.0024	0.0005	0.0001	0.0001	7E-05	0.0001	7E-05	7E-05	7E-05	7E-05
32410	0.0046	0.074	0.3773	0.3704	0.1445	0.0257	0.0024	0.0005	0.0002	7E-05	7E-05	0.0001	7E-05	7E-05	7E-05	7E-05
33110	0.0035	0.0674	0.3831	0.3774	0.1405	0.0248	0.0023	0.0003	0.0001	7E-05	0.0001	0.0001	7E-05	7E-05	7E-05	7E-05
40610	0.0038	0.064	0.3813	0.3787	0.1455	0.0237	0.0019	0.0005	7E-05	7E-05	0.0001	7E-05	7E-05	0.0001	7E-05	7E-05
41310	0.005	0.0741	0.3755	0.3655	0.1498	0.0265	0.0024	0.0005	0.0001	7E-05	0.0003	7E-05	7E-05	7E-05	7E-05	7E-05
41910	0.0049	0.0679	0.3783	0.3806	0.1404	0.0248	0.0019	0.0003	0.0002	0.0001	0.0001	0.0001	7E-05	7E-05	7E-05	7E-05
42710	0.0045	0.0694	0.3829	0.3703	0.1465	0.0233	0.0019	0.0005	0.0001	0.0001	0.0001	7E-05	7E-05	7E-05	7E-05	7E-05

maximum drop in signal strength over a specified set of short moving time windows (1 week, 2 weeks, 3 weeks). The signal drop values D_i^j are calculated for each meter j using the procedure below, and the meters are ranked based on their largest signal drops.

for each meter j **do**
  for each week i **do**
    $D_i^j \leftarrow \min((S_i^j - S_{i-1}^j), (S_i^j - S_{i-2}^j), (S_i^j - S_{i-3}^j), D_{i-1}^j)$
  end for
end for

Figures 5 and 6 show plots of signal strengths for meters with large drops in signal strength. Note that there is overlap between the set of meters exhibiting jumpiness characteristics and the set of meters with large drop characteristics.

D. Z-Score Signal Analysis

Z-scores are normalized scores to compute how many standard deviations each signal strength is away from the mean so that we can track how much of an outlier the meter's signal strength is from week to week. Since we are primarily interested in low signal strength readings, we compute the z-score based on how many standard deviations each signal strength is below the mean signal strength. The z-score Z_w^j for each meter j during week w is calculated using the formula

$$Z_w^j = -\left(\frac{S_w^j - \mu_w}{\sigma_w}\right) \quad (5)$$

where S_w^j is the signal strength of meter j in week w, μ_w is the mean of signal strengths of all meters in week w and σ_w is the standard deviation of signal strengths of all meters in week w. For each meter, we count the number of weeks the z-score calculated in the above step is greater than or equal to 3, and divide by the total number of weeks (W) to

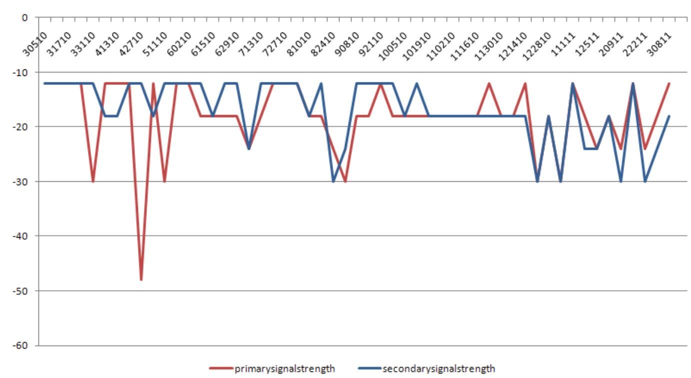

Fig. 5. Plot of primary and secondary signal strength against time for a meter exhibiting large drop characteristics.

TABLE IV
MEAN AND STANDARD DEVIATION OF THE PRIMARY AND SECONDARY SIGNAL STRENGTHS OF THE REPRESENTATIVE METERS AFTER REMOVING THE OUTLIERS.

Week	Mean Primary	Std. Dev. Primary	Min. Primary	Max. Primary	Mean Second.	Std. Dev. Second.	Min. Second.	Max. Second.
30510	-9.7861816	5.3054534	-24	6	-5.0406212	5.3365727	-30	6
31110	-9.8335605	5.3485612	-24	6	-4.9987804	5.4253103	-84	6
31710	-9.7348703	5.2711226	-24	6	-4.9824207	5.3475755	-30	6
32410	-9.8052257	5.3649757	-24	6	-5.0619737	5.3367431	-30	6
33110	-9.8443088	5.2433352	-24	6	-5.1412567	5.3153847	-24	6
40610	-9.8899387	5.209434	-24	6	-5.1699964	5.2346575	-24	6
41310	-9.84305	5.4053001	-24	6	-5.0221137	5.4170056	-24	6
41910	-9.8178689	5.2575503	-24	6	-5.0466083	5.2826573	-24	6
42710	-9.820568	5.2843349	-24	6	-5.0275387	5.3350311	-24	6

TABLE V
FREQUENCY DISTRIBUTION OF PRIMARY SIGNAL STRENGTHS FOR THE REPRESENTATIVE METERS IDENTIFIED IN SECTION IV. EACH COLUMN REPRESENTS THE NUMBER OF OCCURRENCE OF SIGNAL STRENGTH S_i IN A PARTICULAR WEEK.

Week	6	0	-6	-12	-18	-24	-30	-36	-42	-48	-54	-60	-66	-72	-78	-84
30510	57	1015	5353	5272	2002	210	-	-	-	-	-	-	-	-	-	-
31110	49	1008	5411	5164	2073	234	-	-	-	-	-	-	-	-	-	-
31710	52	988	5479	5200	1951	210	-	-	-	-	-	-	-	-	-	-
32410	60	1021	5346	5235	1976	255	-	-	-	-	-	-	-	-	-	-
33110	44	926	5428	5365	1915	247	-	-	-	-	-	-	-	-	-	-
40610	48	880	5382	5356	1988	211	-	-	-	-	-	-	-	-	-	-
41310	62	1026	5341	5191	2044	264	-	-	-	-	-	-	-	-	-	-
41910	65	934	5388	5412	1915	232	-	-	-	-	-	-	-	-	-	-
42710	57	954	5441	5252	2025	215	-	-	-	-	-	-	-	-	-	-

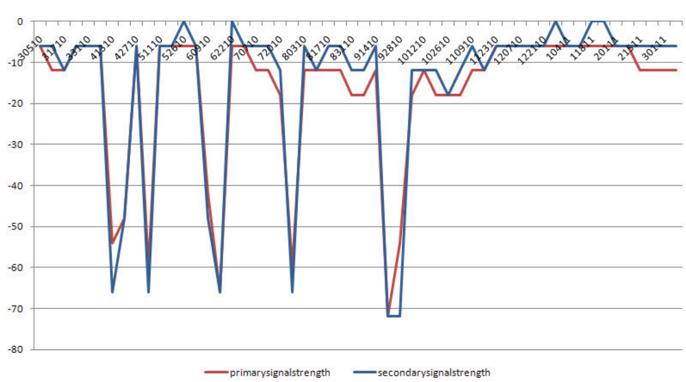

Fig. 6. Plot of primary and secondary signal strength against time for a meter exhibiting large drop characteristics.

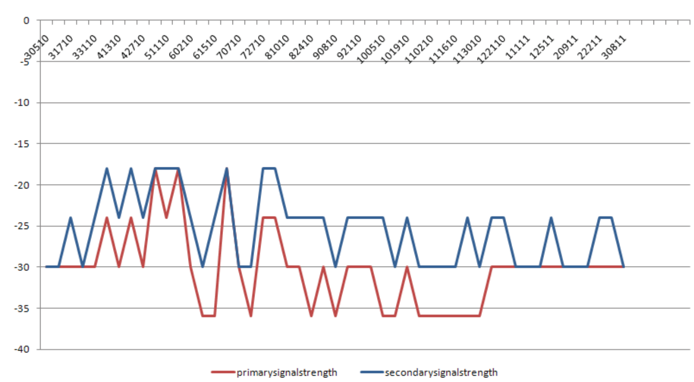

Fig. 7. Plot of primary and secondary signal strength against time for a meter with a large normalized z-score outlier score.

obtain the normalized z-score outlier score. The more often a meter exhibits low signal strengths, the greater its normalized z-score outlier score will be. Figures 7 and 8 show plots of signal strengths for meters with large normalized z-score outlier scores. The normalized z-score outlier score for each meter is calculated as shown below.

for each meter j **do**
  for all Z_w^j calculated for the meter j **do**
    if $Z_w^j \geq 3$ **then**
      $Z_{outlier}^j \leftarrow Z_{outlier}^j + 1$
    end if
  end for
  $Z_{normalized}^j \leftarrow \frac{Z_{outlier}^j}{W}$
end for

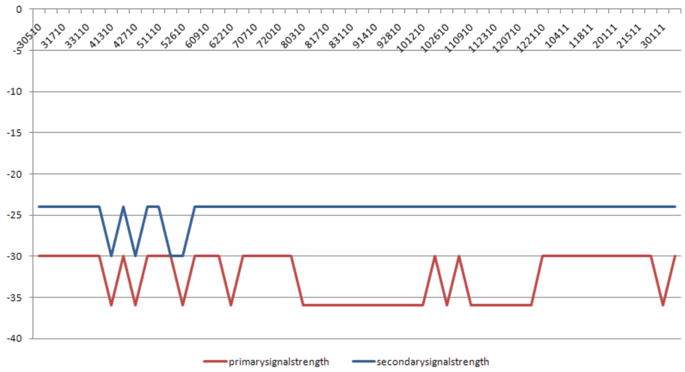

Fig. 8. Plot of primary and secondary signal strength against time for a meter with a large normalized z-score outlier score.

IV. Identifying Outliers and Representative Meters Among the Set of Smart Meters

We used the above metrics to identify potentially weak signals with jumpy (top 1%), flat (top 1%), normalized z-score outlier score (> 0), and large drop characteristics (drop ≤ -18). Based on these, we identified 677 meters (4.66% of 14524 meters) as exhibiting potential connection anomalies. For the remaining representative meters (meters that did not exhibit jumpy, flat, z-score and large drop characteristics, i.e., properly functioning meters), we computed their weekly mean and standard deviation (Table IV).

We selected the top 1% threshold for the meters identified by jumpy and flat metrics by testing the metrics with other percentage values (3%, 2%, and 0.5% of meters identified by jumpy and flat metrics). In each case we identified the representative set of meters and compared its frequency distribution of signals with the frequency distribution of signals for the entire set of meters. We found that the frequency distribution of representative meters identified with 2% and 3% threshold for jumpy and flat metrics eliminated some of the normal meters. On the other hand the frequency distribution of representative meters with 0.5% threshold for jumpy and flat metrics included some meters with large signal drops.

The signal strength quality of the representative meters selected as described above can be inferred by examining their signal strength distributions (Table V). We see that all signal strengths are -24 or greater. A comparison of signal strength distributions of the entire set of meters in Table II with signal strength distributions for the representative meters in Table V shows that meters exhibiting potential connection anomalies (i.e., meters exhibiting either jumpy, large drop, flat, or higher z-score characteristics) have been eliminated. The representative meters identified could be used to build baseline data (e.g., mean, standard deviation) over a period of time. The baseline data can help in trend analysis and modeling of the meter behavior.

V. Conclusion

This paper presents metrics for detecting the set of outliers among a set of smart meters, and shows they can be used to identify different types of connection anomalies. The result of applying the techniques on a fifty-four week set of data from 14524 meters showed that roughly 677 smart meters exhibited potential connection anomalies. Further analysis of the anomalous meters showed that some of them were assigned to incorrect parent nodes which caused the signals to hop across neighboring nodes resulting in large signal drops. Note the improvement in signal strength in Figure 1 after this was corrected.

This paper shows how to detect anomalous meters among a large set of smart meters. The baseline data from the representative meters could be used to perform trend analysis and examine the behavior of meters in response to external factors like severe weather. The results of this study could be used to visualize the entire hierarchy of smart meters and nodes with visual elements depicting the quality of the connections among the meters and nodes. Future work will include exploring the use of time series analysis techniques [1] to identify anomalous smart meters.

Acknowledgments

We thank David Masters, Vicky McLean-Young, and Randy Brown for their guidance and support in making this project possible. This research was supported in part by Duke Energy.

References

[1] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2nd edition, Mar. 2002.

[2] V. M. Catterson, S. D. J. McArthur, and G. Moss. Online conditional anomaly detection in multivariate data for transformer monitoring. *IEEE Transactions on Power Delivery*, 25(4):2556–2564, Oct. 2010.

[3] V. M. Catterson, S. E. Rudd, S. D. J. McArthur, and G. Moss. On-line transformer condition monitoring through diagnostics and anomaly detection. In *15th International Conference on Intelligent System Applications to Power Systems (ISAP '09)*, pages 1–6, Nov. 2009.

[4] I. H. Cavdar. A solution to remote detection of illegal electricity usage via power line communications. *IEEE Transactions on Power Delivery*, 19(4):1663–1667, Oct. 2004.

[5] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41:15:1–15:58, July 2009.

[6] D. Fink and R. J. Jeung. Feasible connectivity solutions of PLC for rural and remote areas. In *2008 IEEE International Symposium on Power Line Communications and Its Applications*, pages 158–163, Apr. 2008.

[7] P. Gross, A. Boulanger, M. Arias, D. Waltz, P. M. Long, C. Lawson, R. Anderson, M. Koenig, M. Mastrocinque, W. Fairechio, J. A. Johnson, S. Lee, F. Doherty, and A. Kressner. Predicting electricity distribution feeder failures using machine learning susceptibility analysis. In *Proceedings of the 18th Conference on Innovative Applications of Artificial Intelligence, Volume 2*, pages 1705–1711. AAAI Press, 2006.

[8] D. Hong, J. Lee, and J. Choi. Power quality monitoring system using power line communication. In *Fifth International Conference on Information, Communications and Signal Processing*, pages 931–935, Jan. 2005.

[9] T. T. Ku, C. S. Chen, C. H. Lin, M. S. Kang, and H. J. Chuang. Identification of customers served by distribution transformer using power line carrier technology. In *4th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 3476–3481, May 2009.

[10] P. Topfer. Technology review of Power Line Communications (PLC) technologies and their uses in Australia. Technical report, PB Associates: Prepared for The Department of Communications, Information Technology and Arts, 2003.

[11] M. S. Yousuf and M. El-Shafei. Power Line Communications: An Overview - Part I. In *4th International Conference on Innovations in Information Technology (IIT '07)*, pages 218–222, Nov. 2007.