2011

A Secure Data Aggregation and Dispatch Scheme for Home Area Networks in Smart Grid

Yun Ye
University of Nebraska-Lincoln, yye@huskers.unl.edu

Yi Qian
University of Nebraska-Lincoln, yqian2@unl.edu

Hamid Sharif
University of Nebraska-Lincoln, hsharif@unlnotes.unl.edu

2011 IEEE Global Telecommunications Conference (GLOBECOM 2011)
doi: 10.1109/GLOCOM.2011.6133655

A Secure Data Aggregation and Dispatch Scheme for Home Area Networks in Smart Grid

Ye Yan, Yi Qian and Hamid Sharif
Department of Computer and Electronics Engineering
University of Nebraska-Lincoln
Emails: yy2318@ieee.org, {yqian2, hsharif}@unl.edu

***Abstract*—Cyber security for smart grid communication systems is one of the most critical requirements need to be assured before smart grid can be operationally ready for the market. Privacy is one of a very important security services. The customer information privacy in smart grid need to be protected. Smart grid data privacy encompasses confidentiality and anonymity of the information extracted from smart devices and metering transmission in a smart grid communication system. In this paper, we consider a home area network as a basic reading data aggregation and dispatch unit in smart grid systems, and we propose a secure in-network data aggregation and dispatch scheme to keep the confidentiality and anonymity for collecting power usage information of smart home devices to the household smart meter and for the reverse control message distribution procedure. More specifically, we introduce an orthogonal chip code to spread reading-data of different smart home devices into spread code, followed by a circuit shifting operation to coupling neighboring smart devices tightly. We adopt an in-network mechanism to further mask it with the spread data and its forwarding data. Finally, we analyze the cyber security protection levels using an information theoretic quantity - residual uncertainty. Simulation studies are conducted to test the performance on different metering datasets for the proposed scheme. This paper sets the ground for further research on optimizing of home power management systems with regarding to the privacy of customer power usage behaviors.**

I. Introduction

Smart grid is characterized by two-way flows of both electricity and control information to create an automated, widely distributed power network. It incorporates into the legacy electricity grid the benefits of communication networks to deliver real-time monitoring and controlling data. It enables the near-instantaneous balance of power supply and demand [1] [2] in smart grid. A smart grid differs from a legacy power grid in that it interconnects the grid components with a two-way communication system to support real-time operations such as load shedding/management, distributed energy storage (e.g. electric vehicles), and distributed energy generation (e.g. renewable energy resources). For those purposes, Advanced Metering Infrastructure (AMI) is introduced into smart grid which collects, measures and analyzes energy usage.

A smart meter in AMI can measure energy consumption in much more details than a conventional electricity meter. It can further forward the collected information to the authorized collectors (e.g. utilities), and facilitate real-time power monitoring and control in smart grid. Usually, there is a smart meter associated with each house or other resident site. A smart meter is connected to all the intelligent electrical devices or appliances through a home area network (HAN), to collect the power usage and distribute the control message from/to them.

The security of both smart metering data and electrical appliances usage data is of prime sensitive, given the scale of potential cyber threats. A classification of smart grid cyber risks and vulnerabilities has been drafted by NIST [3]. The problem of smart grid privacy has been discussed in [4]. In [5] the authors argued that AMI provides a window to the activities within a home, exposing once private activities to anyone with access to electricity usage information. Those privacy threats go beyond the exposition of private information to a hacker. AMI can facilitate the collating and analyzing of such personal data on an industrial scale.

In this paper, we investigate the security and privacy of the power usage data of electrical appliances level in a home area network. The security and privacy of home area networks is an integral part of the smart grid security. We propose an in-network data spreading method to hide the information containing in power consumption data by spreading meter reading data with an unique chip code, after masking with spreading data of others with certain circuit shift operation. From the privacy viewpoint, we can hide or obscure metering information of users so that appliance usage events cannot be detected. We use a information theoretic quantity - residual uncertainty to evaluate the offered privacy protection. Simulation results and further discussions suggest that our proposed in-network meter reading aggregation and dispatch methods can significantly improve privacy and efficiency of smart metering.

II. Background

A. System Overview

Fig. 1 illustrates an overview of a home area network. A home area network (HAN) constructs a basic unit of a communication infrastructure for energy management in smart grid. It comprises a smart meter, several smart electrical appliances and alternative private energy sources converting and storage devices such as wind turbines, solar panels, electric vehicles, and batteries. We treat all those intelligent appliances and devices as smart devices in the context of HAN. A smart meter should collect real-time power consumption and/or generation of each smart device in HAN using a popular wireless technology (e.g., ZigBee). After collecting the reading

978-1-4244-9268-8/11/$26.00 ©2011 IEEE

of each smart device, smart meter aggregates the reading data to get the overall power consumption and report the aggregated data to local utility management office in the smart grid system.

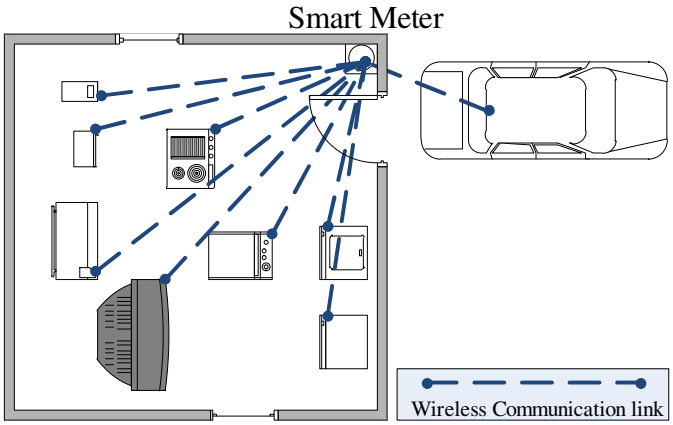

Fig. 1. A home area network

B. Privacy Problem

Smart meters are expected to provide accurate readings automatically at requested time intervals to the utility company. The expected frequency of smart meter readings is speculated that this could be as every few (1-5) minutes, which raises important privacy issues for processing of such data. The detailed energy usage information could show the daily energy usage patterns of a household and even go so far as to enable deduction of what kind of device or appliance was in use at any given time.

In [5] the authors discussed the privacy concerns with regard to expected and/or projected availability of high-frequency metering data. Another good argument for privacy is given in a recent paper on Digital Inclusion and its ramifications [6]. There are some literature in load signature algorithms which use energy measurements to extract detailed information regarding domestic appliance usage. This kind of research is typically termed NALM (Non-intrusive Appliance Load Monitoring), as originally discussed in [7]. There is also a previous work in the construction of appliance libraries and detection algorithms in [8].

III. An In-network Meter Reading Aggregation and Dispatch Scheme

In order to address the afore mentioned privacy threats, we present an in-network meter reading aggregation and dispatch scheme for smart devices in a HAN which can provide secure and efficient communications. The design criteria of the proposed in-network meter reading aggregation and dispatch scheme are discussed in the following.

1) Device authentication: The identity and legality of the smart devices should be verified before joining the HAN with a smart meter and receiving the proper control messages.
2) Data confidentiality: The smart device readings and control messages should be kept in confidential to conceal privacy of consumers and the business information of utilities from unauthorized entities.
3) Message integrity: The system should be able to verify that any meter-reading/management messages to be delivered unaltered in AMI.
4) Maintaining secrecy: It should be ensured that some secrecy of a smart device will always keep in privacy while some secrecy should be shared with particular partners to build secure communications.
5) Prevent potential cyber attacks: A smart device, by holding its own digital credential and following certain processes in the proposed in-network collaborative scheme, should be able to guarantee secure communication connections with the HAN. Even if an individual smart device is compromised, the adversary cannot use the compromised smart device to further access the information of other smart devices or take the advantage to penetrate the AMI in smart grid.
6) Facilitating communication overhead: Any security scheme will bring additional overhead especially in communication systems. The proposed in-network collaborative communication scheme should be efficient in terms of communication overhead and processing latency.

A. In-network Aggregation and Dispatch

As illustrated in Fig. 2, our proposed in-network aggregation and dispatch scheme implement the smart device reading data aggregated using hop by hop mode while control messages dispatched using broadcast mode rather than the basic service set scheme in HANs. Using the propose in-network aggregation and dispatch scheme, the smart meter can reconstruct every reading data of each smart device since it have all the chip codes used by the smart devices. Moreover, the smart meter can use those chip codes to spread control message to each smart device and then add the spread message together into a single frame. By broadcasting this single frame to every smart devices, each smart device can only retrieve its own control message from the single frame since it only knows its own chip code.

B. Chip Codes

Walsh function based on Hadamard code is a well-known choice among many methods to generate orthogonal codes [9], [10]. In this paper, we adopt it to generate mutual orthogonal chip codes to be used in the secure in-network data aggregation and dispatch scheme we are proposing. The Hadamard matrices of different dimensions (N) are established using the following recursive expressions:

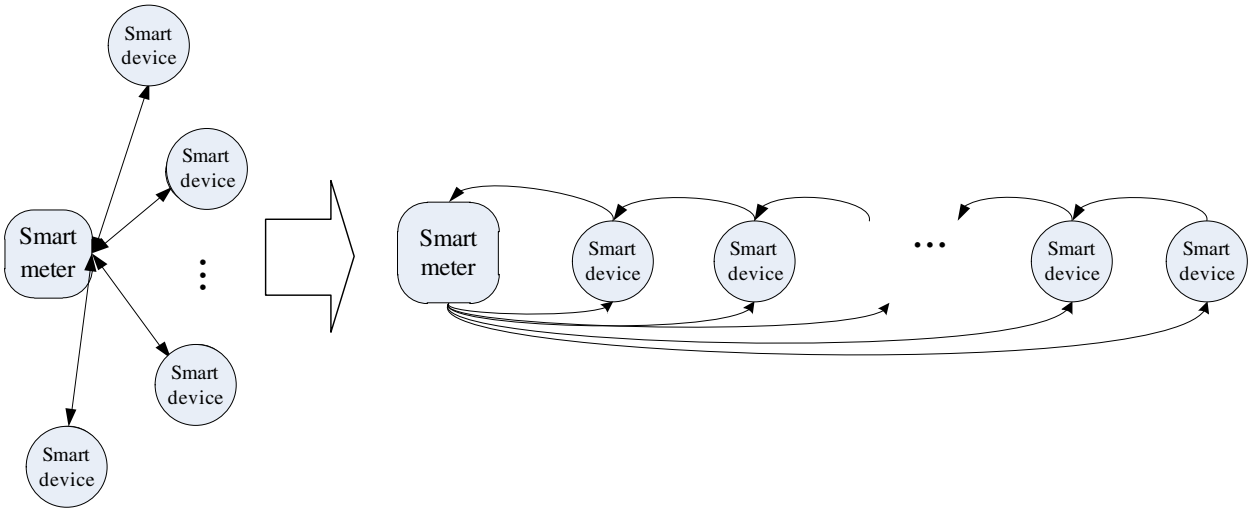

Fig. 2. Basic service set vs. in-network aggregation and dispatch system architecture for HANs

$$W^1 = [W] \qquad (W = 1 \ or \ -1)$$
$$W^2 = \begin{bmatrix} W^1 & W^1 \\ W^1 & \overline{W^1} \end{bmatrix}$$
$$\vdots \qquad (1)$$
$$W^N = \begin{bmatrix} W^{N-1} & W^{N-1} \\ W^{N-1} & \overline{W^{N-1}} \end{bmatrix}$$

and

$$N = 2^{\lceil \log_2 \mathrm{n} \rceil}, \mathrm{n} \geq 2 \qquad (2)$$

where n is the number of active smart devices communicating with smart meter within HAN. The overscore of a matrix denotes the complement operation. To convert a Hadamard matrix into orthogonal codes, we take a row of the Hadamard matrix.

Notice that the dot product for any pair of orthogonal codes is zero.

$$W_i^N \odot W_{i'}^N = 0 \quad i, i' \in \{1, ..., N\} \ and \ i \neq i' \qquad (3)$$

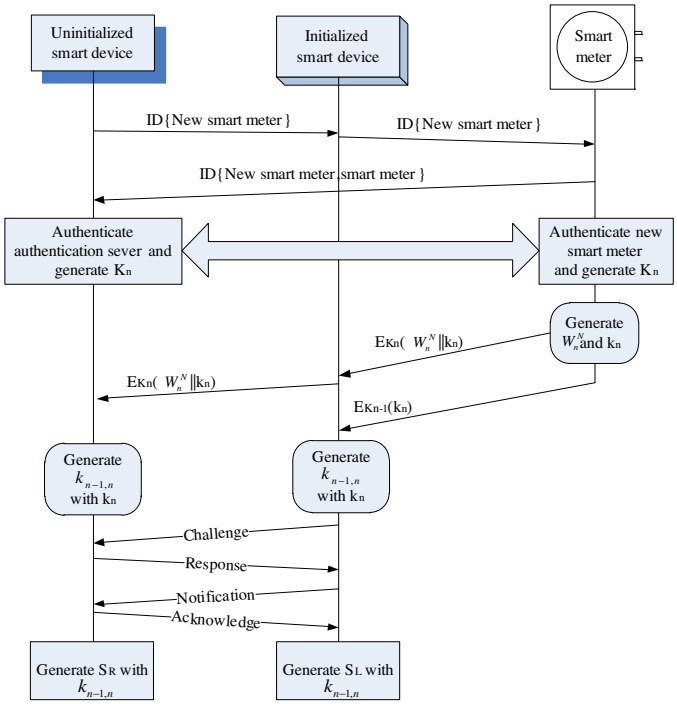

Fig. 3. Initialization Process for A Newly Boot-up Smart Device

C. Initialization Procedure

Fig. 3 illustrates the initialization procedure for each newly boot-up smart device in the mutual authentication with both smart meter and neighboring initialized smart device. Before joining an authenticated HAN, the newly boot-up smart device must be authenticated by the smart meter as an authentication server as a legal device to terminal customers. The neighboring authenticated smart device plays as a authenticator in the initialization procedure to convey the authentication transaction messages between new smart device and smart meter. The smart meter acts as a portal in both home electricity system and HAN.

The newly boot-up smart device send smart meter its digital identification as an authentication request, which can be authenticated by the smart meter. If the authentication for the new smart device identity is successful, the smart meter sends back the verified digital identification of both the new smart device and the smart meter. It will generate a key K for the new smart device. The newly boot-up device can also authenticate the smart meter by verifying the smart meter digital identification and generate the K itself. After the successful mutual authentications, the smart meter updates its active device number n and Hadamard matrix W^N if

necessary.

The K generated after mutual authentication between newly boot-up smart device and smart meter is unconcealed to any devices else including its neighboring smart device. Because both the consequent data masking chip code W_n^N and neighboring device mutual authentication key k between newly boot-up smart device and neighboring initiated smart device are encrypted by this K. If the identity of a newly boot-up smart device is authenticated as a valid smart device, the corresponding credential is established between the smart meter and the new smart device. Then, the smart meter will pick a row W_n^N from unoccupied Hadamard matrix rows as the chip code for the authenticated smart device concatenating with k. The W_n^N and k will be encrypted by K, so the new smart device can decrypt and get its W_n^N and k. Meanwhile, the smart meter sends the k to the neighboring smart device encrypted with their own K' since the initiated neighboring smart device authenticated with K'. So the neighboring smart device knows the k, too.

With its k, the new smart device is initialized and it can generate the $k_{n-1,n}$ individually and then conduct a four-way handshake procedure to fulfill another mutual authentication between them. After the successful mutual authentication, $k_{n-1,n}$ at both the new smart device and the initialized smart device sides are validated and ready to the consequent reading data message authentication code (MAC) generation and validation. S_R and S_L are negotiated during the four-way handshake procedure, too. Where S_R and S_L are the number of bit positions to be right shifted and left shifted, as described in Section III.E.

D. Data Spread

Before data spread operation, we must represent the original data bit 1 or 0 into bipolar form as well as consider the null state when there is no data to present or in error state, as shown in the following:

$$\begin{aligned} &data: 1 \rightarrow +1 \\ &data: 0 \rightarrow -1 \\ &null: \rightarrow 0 \end{aligned} \tag{4}$$

Then, each original data bit can be spread into $N \times (N+1)$ bits spread code using its assigned chip code. Fig. 4 shows an example that one bit original reading data have been spread for 4 active smart devices with their chip codes accordingly.

After all bits of original reading data have been spread, the MAC and padding bits if necessary will be appended to the spread data as illustrated in Fig. 5. The MAC is generated with key $k_{n,n-1}$. Therefore, its neighboring smart device can validate the integrity of the data it received.

E. Circuit Shift

Instead transmit the spread data and MAC plus padding bits directly, we further introduce a circuit shift operation for the frame. As illustrate in Fig. 6, the spread reading data as payload will shift to right hand side with S_R bits.

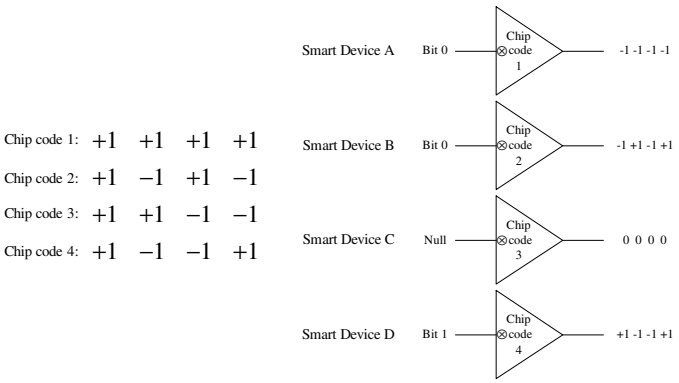

Fig. 4. Example of N = 4 chip codes for 4 active smart device data spread

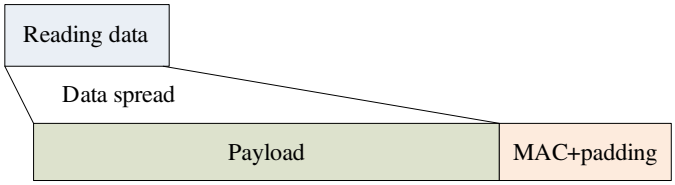

Fig. 5. Data spread and MAC plus padding operation for reading data

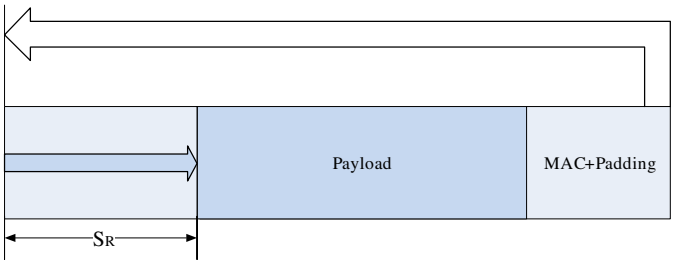

Fig. 6. Circuit shift of payload to right-hand side with S_R

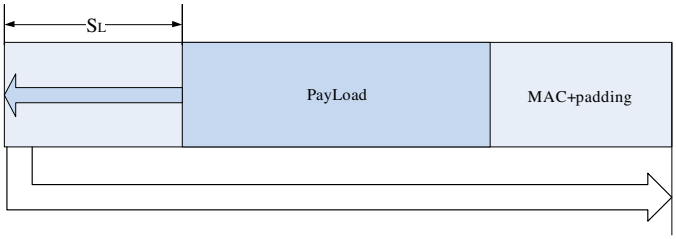

Fig. 7. Circuit shift of payload to left-hand side with S_L

After received the right shifted frame, only the authenticated neighboring smart device knowing S_L can do the exactly left hand side circuit shift with S_L bits back to reverse the frame. Then, the payload, MAC and padding bits are back to its right position for the MAC verification, data aggregation or data reconstruction.

F. Data Aggregation and Reconstruction

As shown in the left part of Fig. 8, if the received data is verified through MAC as validated, the received spread reading data will be added to its spread reading data by each spread code. Then each spread code will update the value hop by hop

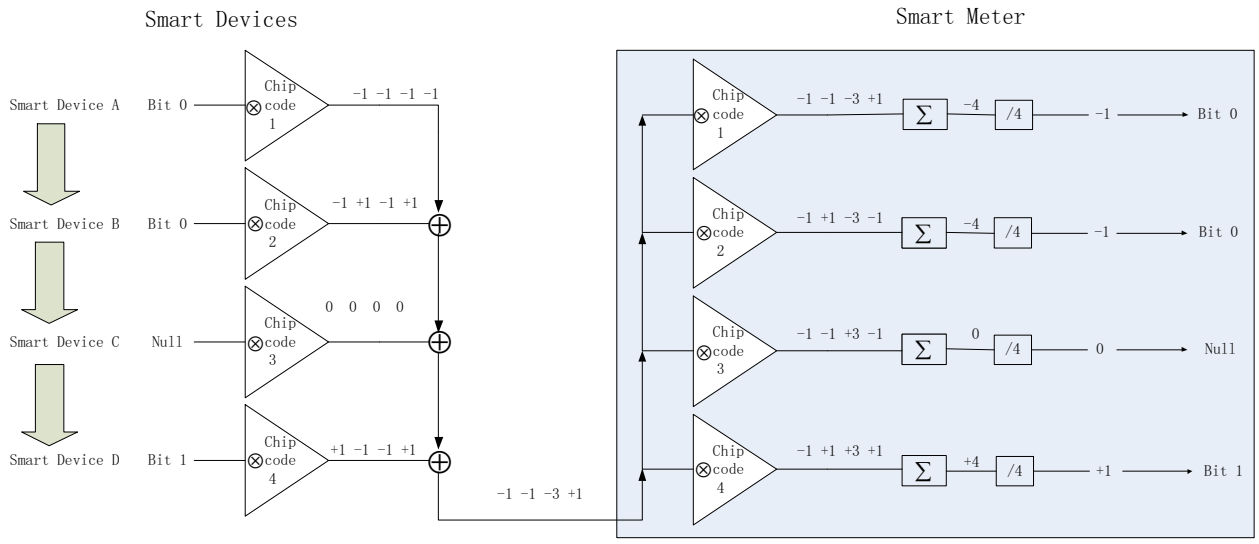

Fig. 8. An example for data aggregation and reconstruction in 4 active smart devices and smart meter

until it reaches the smart meter. In case the received data did not pass the MAC verification, the smart device can simply set it null and continue the procedure without any interruption to the data aggregation procedure.

After verified the received data from the immediate smart device, the smart meter can reconstruct all the reading data from the aggregated data since the smart meter has all the chip codes of the participated data aggregation smart devices. The smart meter can even point out the exact smart device that from whom the reading data become null in the data aggregation procedure. The smart meter can take the proper actions using its control message dispatch.

IV. Performance Evaluation

A. Security Metric

In cryptography, privacy is typically discussed in the context of anonymity, the property of hiding the identity of a user associated with a message [11]. Typically, privacy levels are analyzed stochastically. However, it is useful to quantify the privacy level using information theoretic metrics [12], [13].

The undetectability property is to hide the information, specifically as the source reading data in this paper. The nature of information security evaluation problem allows us to use a stochastic quantity - information entropy [14]. It is a well known information theoretic quantity which can be used to evaluate the sources of information. To employ this metric in our security context, we assume that the processed data can be modeled as a stochastic process with probability measure P. If $p(x)$ is the probability density function (pdf) of P, the entropy H is defined:

$$H = p(x) \log \frac{1}{p(x)} \tag{5}$$

For data spreading, there are $L = log_2 N$ possible trials to determine the right data spreading dimension N if using brutal force search,

$$H_{spread}(x) = 2 \sum_{x \in L} x \log x + 1 \tag{6}$$

The entropy of coupling operation of right and left circuit shifts is

$$H_{shift}(x) = \log S \tag{7}$$

where $S = S_L = S_R$ is the bits for the circuit shift between coupling neighboring devices

The overall topology entropy for the in-network operation of data spreading and circuit shift can be calculated as

$$H_{topology}(x) = 2 \log S_{max} \sum_{x \in N_{max}} x \log x \tag{8}$$

Now, we can define the equivocation of the residual uncertainty R for an eavesdropper. With respect to the spread dimension N as

$$\begin{aligned} R &= \frac{H(x|N)}{H(x)} \\ &= \frac{H_{spread}(x) \cdot H_{shift}(x)}{H_{topology}(x)} \end{aligned} \tag{9}$$

When R = 0, the eavesdropper completely knows the information source. When R = 1, the processed data does not provide any information for the eavesdropper.

Fig. 9 shows residual uncertainty R with various reading data length in term of the number of active smart devices. We can observe that the residual uncertainty increases with the number of active smart device, as the spread dimension N increase while the various reading data length will not change the residual uncertainty R severely when the number of active smart device is small.

B. Delay Performance

We implemented a ns-2 simulation to evaluate the delay performance of the proposed in-network data aggregation and dispatch scheme. We assume that there are n active smart

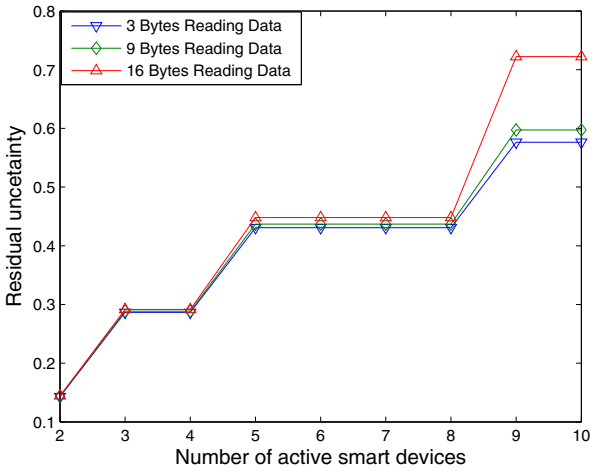

Fig. 9. Residual uncertainty vs. number of active smart devices

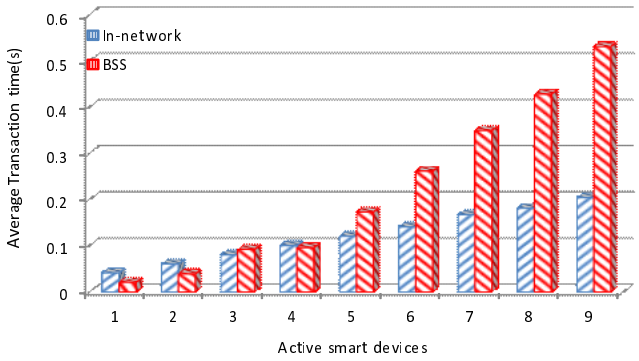

Fig. 10. Residual uncertainty vs. number of active smart devices

devices associated with a smart meter in a home area network. Each smart device generates a data packet of reading every second (1 packet/sencond) independently and need to send its reading data packet to smart meter in 1 second. For simplicity, we assume the reading data lifetime is equal to that of the arrival of data packet. We assume that each smart device reading data consists of 16 bytes. We run 10 realizations with active smart device number from 1 to 10 and the simulation time lasted for 100 seconds for each realization to obtain the statistics.

The average transaction delays are shown in Fig. 10 for various numbers of active smart devices. A transaction contains reading data collection from all active smart devices to the smart meter and the reverse operation of control message dispatch from the smart meter to smart devices. We can observe that as the number of active smart devices increases, the gap of average delay between the legacy reading data collection and dispatch scheme using basic service set (BSS) and our proposed in-network scheme becomes larger, although BBS has a lower average transaction delay when the number of active smart devices is small. The increasing translation delay of the BSS with respect to the increasing number of smart devices is much larger than that of the proposed in-network scheme. Therefore, it is more practical to adopt the proposed in-network scheme when there are many smart devices in a home area network.

V. Conclusion

In this paper, we proposed a secure and efficient in-network data aggregation and dispatch scheme for advance metering infrastructure in home area networks for smart grid. In the proposed scheme, the original reading data are spread and then mixed up with the spreading code of other smart devices. Only the smart meter can reconstruct the original reading data from the mixed data using the chip code established with smart devices in their initialization procedure through mutual authentications. The performance of the reading data aggregation and dispatch has been analyzed subject to the HAN setting. The levels of security has been discussed qualitatively, which lies in the secrecy of pseudo-random spreading codes and circuit shift. Simulation results have demonstrated the advantage of the proposed scheme over the traditional BSS approach.

References

[1] A. Ipakchi and F. Albuyeh, “Grid of the future,” IEEE Power and Energy Magazine, vol. 7, no. 2, pp. 52-62, 2009.
[2] Z. Fan, G. Kalogridis, C. Efthymiou, M. Sooriyabandara, M. Serizawa, and J. McGeehan, “The New Frontier of Communications Research: Smart Grid,” in Proceedings of the 2010 ACM e-Energy Conference, Passau, Germany, Apr. 2010.
[3] A. Lee and T. Brewer, “Smart grid Cyber Security Strategy and Requirements,” NIST IR 7628, 2nd Draft, Feb. 2010
[4] http://voices.washingtonpost.com/securityfix/2009/11/experts_smart_ _rid_poses_priva.html.
[5] E. L. Quinn, “Privacy and the New Energy Infrastructure”, Social Science Research Network (SSRN), February 2009
[6] R. Stallman, “Is digital inclusion a good thing? How can we make sure it is?,” IEEE Communications Magazine, vol. 48, pp. 112-118, February 2010
[7] G. W. Hart, “Nonintrusive appliance load monitoring”, Proceedings of the IEEE, vol.80, no.12, pp.1870-1891, December 1992
[8] H. Y. Lam, G. S. K. Fung, and W. K. Lee, “A Novel Method to Construct Taxonomy of Electrical Appliances Based on Load Signatures,” IEEE Transactions on Consumer Electronics, vol.53, no.2, pp.653-660, May 2007
[9] V. M. DaSilva and E. S. Sousa, “Multicarrier orthogonal CDMA signals for quasi-synchronous communications systems,” IEEE J. Select. Areas Commun., vol. 12, pp. 842-852, June 1994.
[10] X. D. Lin and K. H. Chang, “Optimal sequence design for quasi-synchronous CDMA communication system,” IEEE Trans. Commun., vol. 45, pp. 221-226, Feb. 1997.
[11] D. Chaum, “The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability,” Journal of Cryptology, vol. 1, no. 1, pp. 65-75, 1988.
[12] Y. Deng, J. Pang, and W. P., “Measuring Anonymity with Relative Entropy,” in Proceedings of the 4th International Workshop on Formal Aspects in Security and Trust. Springer, LNCS 4691, 2006, pp. 65-79.
[13] A. Serjantov and G. Danezis, “Towards an information theoretic metric for anonymity,” in Proceedings of the workshop on Privacy Enhancing Technologies, R. Dingledine and P. F. Syverson, Eds. Springer, LNCS 2482, 2002, pp. 41-53.
[14] T. M. Cover and J. A. Thomas, Elements of information theory. John Wiley & Sons, Inc. New York, NY, USA, 2006.