How to cite this paper:

Fiza Abdul Rahim, Asmidar Abu Bakar, Salman Yussof, Roslan Ismail, & Ramona Ramli. (2017). Privacy preservation framework for advanced metering infrastructure in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference on Computing & Informatics (pp 744-749). Sintok: School of Computing.

PRIVACY PRESERVATION FRAMEWORK FOR ADVANCED METERING INFRASTRUCTURE

Fiza Abdul Rahim[1], Asmidar Abu Bakar[2], Salman Yussof[3], Roslan Ismail[4] and Ramona Ramli[5]

College of Computer Science and Information Technology, Universiti Tenaga Nasional, Malaysia,
fiza@uniten.edu.my[1], asmidar@uniten.edu.my[2], salman@uniten.edu.my[3],
roslan@uniten.edu.my[4], ramona@uniten.edu.my[5]

ABSTRACT. Integrating data from multiple sources have enabled countless sharing of data by data owners such as individuals, organizations, and governments. The sharing of data needs control and management to ensure the sensitivity of data is protected. Numerous research have studied on the management of data in the deployment of advanced metering infrastructure (AMI) to ensure the smooth process in data integration. Though, misuse of information which is considered as privacy violation may occurred in the management and sharing of sensitive data. Hence, it is essential to build a framework and design appropriate algorithms to manage integrated data efficiently and to avoid privacy violation. The algorithm will consist of data classification and ranking mechanism to ensure data is managed effectively. This paper lays out a privacy preservation framework for the proposed system. Finally, we correspondingly propose future research directions in the conclusion.

Keywords: privacy preservation, data integration, data ranking, data classification, AMI.

INTRODUCTION

Large amount of data are being collected by electric utility company through the deployment of advanced metering infrastructure (AMI). AMI utilizes two-way communication enabling utilities to collect consumer consumption data and return control signals to consumer in near real time. AMI offers consumers the chance to understand and reduce their energy usage in much greater detail as consumer able to see the usage of the energy based on the user friendly apps like "Smart Meter" develop for Android user (Mobility Minds Solutions, 2016). As for electric utility company, it helps to improve the accuracy of billing and eliminate the needs to visit premises for the purpose of meter reading.

However, the usage of AMI could present a considerable challenge because AMI generates unprecedented data volume, speed and complexity with millions of data points arriving in sub-hourly intervals. For example, for a mid-sized utility servicing a half million consumers, smart meters deployment will result in an explosive proliferation of consumer data, which may resulted to 3000 fold growth in the amount of data that must be captured, ordered, stored, and analyzed in near real-time (SunGard, 2013).

Much of these data are about consumers, their personal details and energy consumption. These data are being stored in meter data management system which later will travel across enterprise service bus to allow communication between mutually interacting software applica-

tions in a service-oriented architecture (SOA). With the large collection of data, electric utility company may also integrate data from other sources for further analysis, such as pricing details for demand response or forecasting information for renewable energy.

Nevertheless, electric utility company needs to take into account government regulations to which they must obey, guaranteeing that records are retained and archived for the time periods stated and protected according to the privacy guidelines. In Malaysia, with the enacted Personal Data Protection Act (PDPA) 2010, electric utility company needs to take specific measures related to privacy issues to ensure their compliance with the act (Laws of Malaysia, 2010).

This paper is structured as follows: The next section states the data integration in AMI, followed by a survey on related published works. The following section discusses the proposed framework and future works of this research. The last section concludes this paper.

DATA INTEGRATION IN ADVANCED METERING INFRASTRUCTURE

The implementation of smart meters in AMI leads to detailed consumer profiles to the utility company and other related parties. Scholars have revealed that fine grained power consumption data can be used to excerpt more detailed information on consumer activities (Edison Electrical Institute, 2011; McKenna, Richardson, and Thomson, 2011; UK Power, 2014). Hence, the electric utility company as a data controller must provide full protection on consumer data disclosure.

The data controller must deal with the complex business models, relationships, and new technologies, and protect the personal data at the same time. However, the complexity involves in determining responsibilities and identifing roles for each actor in the emerging business models, which involves subcontracting, outsourcing, evolving partnerships between organizations in value chains, behavioral advertising, and etc.

With the large size and velocity of information generated by smart meter, the data is often being integrated and shared among administrators with end users such as internal employees from various departments and third parties. A multitude of services are anticipated to provide additional services such as load monitoring and forecasting, dynamic billing, outage and fraud detection, and demand response. This multitude services are used for data analytics. Data analytics can be applied to gather and assess data to excerpt valuable information. The results of data analytics may be used to improve business efficiencies, validate process effectiveness, ascertain areas of key risk, fraud, errors, and influence business decisions.

For instance, the electric utility company may grant and allow third parties to perform data analytics and provide services using consumer data. However, there is a high probability that the data given to the third party may contain sensitive customer information that will violate their privacy, as well as violating government regulations. Therefore, the electric utility company must ensure that data can still be given to the third party for the purpose of doing data analytics, but at the same time the data shared must be managed to hide sensitive information.

Furthermore, much of the collected data are integrated and shared across organization. The progressively exponential increase of integrated personal data could feed data integration applications to discover real life patterns of consumer's activities. A malicious insider or untrusted third parties can abuse these datasets and extract private information about consumers.

According to Clifton *et al.* (2004) and Thanh (2015), it is important to develop techniques to enable the integration and sharing of the data consistently while preserving privacy. Together, the integrated data must be managed accordingly to classify the level of sensitivity. Therefore, it is important to develop a technique to enable the integration and sharing of data that adopts security strategies guaranteeing the non-disclosure of consumer's private information. The technique is expected to enable the integration and sharing of data, classify the level of sensitivity while preserving privacy information.

RELATED WORKS

In this section, related works on privacy preservation techniques, data integration, data ranking and data classification are discussed.

Privacy Preservation Techniques

Privacy is a vital issue in smart metering. One of the examples of the privacy breach produced when collecting readings from a household power consumption (Erkin, 2013). The consumption can be easily identified from the readings. With various powerful techniques that can use the aggregated measurements and provide approximation of the moment when each appliance is turned on and off, this reflects the importance of privacy protection.

Privacy preservation is found to be the most frequently discussed topic in the existing literatures (Abdul Rahim, Ismail, and Samy, 2014). The main goal of the privacy preservation is to ensure the ultimate security to sensitive information. For example, during an emergency situation, there is a need for the data to be shared with the rescuers (Bakar, Ghapar, and Ismail, 2014). However, some of the personal data cannot be shared to all rescuers. Thus, there is a need to ensure the privacy of personal data using privacy-preservation technique.

Many privacy-preserving techniques have been developed to be embedded in existing applications within the organization. Techniques such as randomization, anonymization, perturbation, condensation, cryptographic, privacy-preserving probabilistic record linkage, secure multi party computation, and sequential pattern hiding are among the possible techniques to be applied in AMI environment (Adam *et al.*, 2007; Schmidlin, Clough-Gorr, and Spoerri, 2015; Sukhdev and Vasava, 2015).

Data Integration

As one of the older research areas in database community (Zhang, Wang, and Zhao, 2005; Ziegler and Dittrich, 2007), data integration can be defined as the combination of technical and business processes used to integrate data from multiple sources into meaningful and valuable information in enabling better business decisions and improves business process execution (IBM Corporation, 2014).

For instance, data integration can be used for statistical analysis, querying and reporting on business activities, and data mining in business intelligence (BI) area (Ziegler and Dittrich, 2007). Furthermore, data integration is required as pre-requisite before mining data collected from various sources (Clifton *et al.*, 2004).

The continued exponential growth of distributed personal data across different databases may also be thwarted by a privacy backlash (Clifton *et al.*, 2004). For instance, healthcare data can be shared to improve scientific research, but then the necessity of obtaining permission to use identifiable information for an indvidual can be prohibitive. Another situation, fire departments may share regulatory and defense plans to improve their capability to provide community defense and fight terrorism, but fear loss of privacy could lead to liability.

Data Classification

In database research, data classification performs data analysis that extracts important data classes (Sukhdev and Vasava, 2015). Various typical classification models have been developed in previous research; decision tree, Naïve Bayesian classification and support vector machine. Several classification algorithms also have created in the data mining field for analyzing the data; Iterative Dichotomiser tree (ID3), C4.5, CART, K-Nearest Neighbors (K-NN), Naïve Bayes, and etc.

In information security, data is classified according to sensitivity level and the impact to the organization. In a specific software application, the raw data are usually classified according to a set of features but need to be reclassified when the data is later being integrated from different databases for data analysis purposes. Prior to starting any data analytics project, the

data that has been discovered and integrated must be reclassified according to its sensitivity and business criticality (Tankard, 2012).

Data classification's level of sensitivity is often classified based on varying levels of importance or confidentiality, which relates to the security measures in place to protect each classification level. In defining data classification, Lord (2016) suggested an organization may classify data as restricted, private or public. Data classification reflects the degree of impact to the organization if confidentiality, integrity or availability compromised.

Data Ranking

Despite the vast research on the areas of data classification and its importance, researches on data ranking are intermittent. Several research works have been done in data ranking such as in entertainment industry where data ranking is used to retrieve music based on its relevance and importance (Ruxanda *et al.*, 2008) and real estate field where spatial data is ranked to determine the potential flats in real state agency database with respect to the appropriateness of the location, quality of other facilities within the neighborhood (Yiu *et al.*, 2011).

Data ranking should be done whenever the newly integrated data is created. The possibility of the existing security mechanism might not be appropriate when data is integrated from various sources. Therefore, it is important to rank the data along with an analysis of appropriate security measures to identify whether existing security mechanisms are still sufficient with the new classification.

PROPOSED FRAMEWORK

The aim of this research is to develop privacy preservation framework for AMI data that comprises of several privacy preservation techniques. The proposed framework comprises of five phases as shown in Figure 1.

In the first phase, the raw data is collected from various sources. In this example, head end equipment, web portal, and various systems in the infrastructure such as System A, and System B would be the data donors. After the data collection, the raw data are integrated and kept in meter data management system.

The newly integrated data would be requested by various systems in the infrastructure in the second phase. For instance, requests from third parties that would require certain data to perform data analytics. Based on the request, the integrated data will be classified according to its level of privacy protection in the third phase. This is to guarantee which privacy preservation techniques to be applied in the requested integrated data.

In the fourth phase, data ranking technique will be applied for further evaluation. In this research, data ranking is defined as how data is quantified based on level of sensitivity. It is a process which involves reevaluation of the data sensitivity contained in the requested integrated data. This reevaluation also will align with the objectives of privacy protection and the purpose of request. Several privacy preservation techniques will be applied in the following phase. Finally, the output which is the privacy-preserved integrated data will be distributed to the requested system resources or other parties.

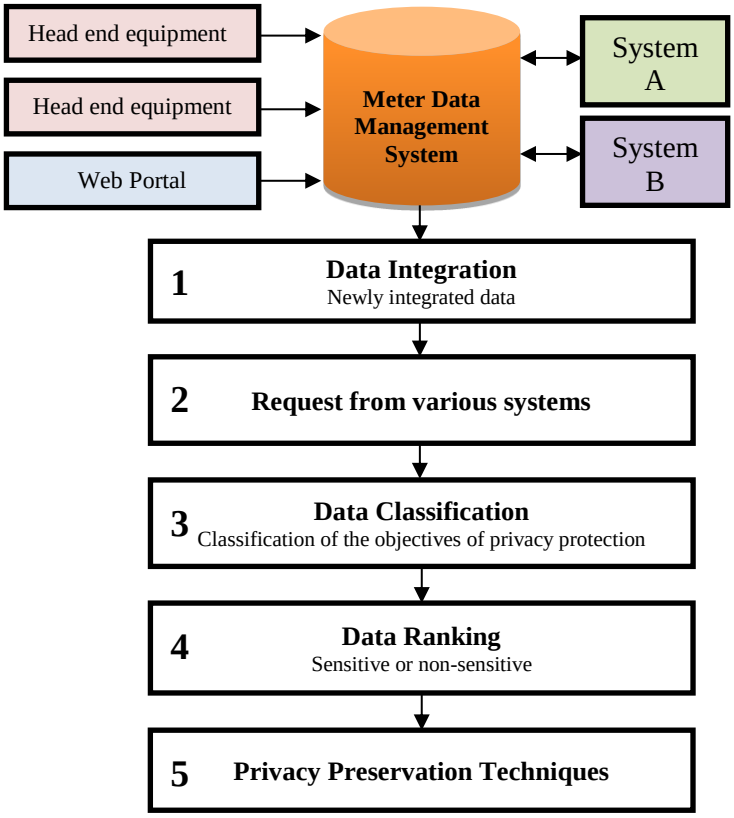

Figure 1. Proposed Framework

CONCLUSION

The goal of this research is to propose a practical privacy preservation framework to ensure the privacy of consumer data while keeping the sensitive data useful for further analytics processes. In this paper, the combination of data classification and data ranking are proposed to be performed before identifying any appropriate privacy preservation techniques to be applied on the newly integrated data. Furthermore, this research is also emphasize on the importance of data classification and data ranking for the newly integrated data. This in-progress work will proceed to the next stage to identify the most suitable mechanism for data classification, data ranking and privacy preservation techniques to be used.

ACKNOWLEDGEMENT

We would like to express profound gratitude to Smart Billing Project, Distribution Network Department, Distribution Division, Tenaga Nasional Berhad for the invaluable support, encouragement, supervision and useful suggestions throughout this research. The project is funded under TNB Seeding Fund.

REFERENCES

Abdul Rahim, F., Ismail, Z., and Samy, G. N. (2014). Privacy Challenges in Electronic Medical Records : A Systematic Review. In *Knowledge Management International Conference (KMICe)* (pp. 12–15).

Adam, N., White, T., Shafiq, B., Vaidya, J., He, X., Vaidya, J., … White, T. (2007). Privacy preserving integration of health care data. *International Journal of Computational Models and Algorithms in Medicine (IJCMAM)*, *2007*(2), 22–36.

Bakar, A. A., Ghapar, A. A., and Ismail, R. (2014). Access control and privacy in MANET emergency environment. In *2014 International Conference on Computer and Information Sciences (ICCOINS)* (pp. 1–6). http://doi.org/10.1109/ICCOINS.2014.6868389

Clifton, C., Kantarcı, M., Doan, A., Schadow, G., Vaidya, J., Elmagarmid, A., and Suciu, D. (2004). Privacy-Preserving Data Integration and Sharing. *Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 19–26. http://doi.org/10.1145/1008694.1008698

Edison Electrical Institute. (2011). *Smart Meters and Smart Meter Systems : A Metering Industry Perspective*. Retrieved from www.eei.org

Erkin, Z. (2013). Privacy-Preserving Data Aggregation in Smart Metering Systems: An Overview. *Signal Processing Magazine*, (march), 75–86. http://doi.org/10.1109/MSP.2012.2228343

IBM Corporation. (2014). Data Integration. Retrieved October 29, 2016, from https://www-01.ibm.com/software/data/integration/

Laws of Malaysia. Act 709: Personal Data Protection Act 2010 (2010).

Lord, N. (2016). What is Data Classification ? A Data Classification Definition. Retrieved December 14, 2016, from https://digitalguardian.com/blog/what-data-classification-data-classification-definition

McKenna, E., Richardson, I., and Thomson, M. (2011). Smart meter data: balancing consumer privacy concerns with legitimate applications. *Energy Policy*, (December).

Mobility Minds Solutions. (2016). Smart Meter. Retrieved from https://play.google.com/store/apps/details?id=com.jems.smartmeter&hl=en

Ruxanda, M. M., Nanopoulos, A., Jensen, C. S., and Manolopoulos, Y. (2008). Ranking Music Data by Relevance and Importance. In *Proceedings of IEEE International Conference on Multimedia & Expo* (pp. 549–552).

Schmidlin, K., Clough-Gorr, K. M., and Spoerri, A. (2015). Privacy preserving probabilistic record linkage (P3RL): a novel method for linking existing health-related data and maintaining participant confidentiality. *BMC Med Res Methodol*, *15*, 46. http://doi.org/10.1186/s12874-015-0038-6

Sukhdev, S., and Vasava, H. (2015). Privacy Preserving Data Mining With Classification And Encryption Methods, *2*(5), 19–23.

SunGard. (2013). *Big Data Challenges and Opportunities for the energy industry*.

Tankard, C. (2012). Big data security. *Network Security*, *2012*(7), 5–8. JOUR. http://doi.org/http://dx.doi.org/10.1016/S1353-4858(12)70063-6

Thanh, D. S. (2015). *A survey of privacy on data integration*.

UK Power. (2014). Smart Energy Meters - A boon or a bane to consumers? Retrieved from https://www.ukpower.co.uk/gas_electricity_news/7636-smart-energy-meters---a-boon-or-a-bane-to-consumers

Yiu, M. L., Lu, H., Mamoulis, N., and Vaitis, M. (2011). Ranking spatial data by quality preferences. In *IEEE Transactions on Knowledge and Data Engineering* (Vol. 23, pp. 433–446). http://doi.org/10.1109/TKDE.2010.119

Zhang, N., Wang, S., and Zhao, W. (2005). A new scheme on privacy-preserving data classification. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 374–383). http://doi.org/10.1145/1081870.1081913

Ziegler, P., and Dittrich, K. R. (2007). Data Integration — Problems, Approaches, and Perspectives. *Conceptual Modelling in Information Systems Engineering*, 39–58. http://doi.org/10.1007/978-3-540-72677-7