# Preserving Prosumer Privacy in a District Level Smart Grid

B. Yuce, M. Mourshed, Y Rezgui
BRE Trust Centre for Sustainable Engineering
Cardiff University, CF24 3AA
Cardiff, United Kingdom
YuceB@Cardiff.ac.uk, MourshedM@Cardiff.ac.uk,
RezguiY@Cardiff.ac.uk

O. F. Rana
School of Computer Science and Informatics,
Cardiff University, CF24 3AA
Cardiff, United Kingdom
RanaOF@cardiff.ac.uk

***Abstract*—This study presents the anonymization of consumer data in a district-level smart grid using the *k*-anonymity approach. The data utilized in this study covers the demographic information and associated energy consumption of consumers. The anonymization process is implemented at the prosumer level, considering their importance in sharing flexibility and distributed generation at the low voltage grid, and the fact that they need to interact with each other and the grid while keeping their data private. The proposed approach is tested under three anonymization scenarios: prosecutor, journalist and marketer. The smart grid data are investigated mostly under the prosecutor scenario with three risk levels: lowest, medium and highest. The results of the *k*-anonymity approach are compared to *k*-map and *k*-map + *k*-anonymity. No difference has been found between the three investigated approaches for the selected data set. Since, the aim of the k-anonymity is to not transform the information about any individual record among those k-1 individual, the recorded type and the number of attributes play a key role for the anonymization process. One of the risk is the using continuous attributes in the anonymization process which may cause the information lose in the anonymization process such as near real time energy consumptions. Hence we have focused on to anonymization of the consumers' demographic information, rather than their energy consumption.**

***Keywords—Anonymization; k-anonymity; Smart grid; District electricty; Low-voltage grid.***

## I. INTRODUCTION

The energy recently became one of the most important challenges for the policy makers and researchers [1]. This challenges is related to technological changes in the power generation systems, which resulted in a new energy participants called *prosumers*. A prosumer is defined as both an energy con*sumer*, and an energy *pro*ducer [2]. The existence of the prosumers is also contributed to development of the micro grids in the entire electricity grids. Since the traditional grids are basic, nonadoptive and single way flow systems, they are not being able to handle both the peak demand request, and to organize the energy generation among the micro grids [3-4]. Hence the traditional grids started to transform to smart grids. This transition of course requires several modifications and transformations in both communications protocols and the control technologies to satisfy the needs of the bi-directional information and energy flow [5]. This flow starts from device level with Home Area Networks (HANs), to Building Energy Management Systems (BEMS), Neighborhood Area Networks (NANs) [6]. These systems harmonically communicate with each other to transform the traditional grids into a smart grid by sharing the information about devices status, energy consumption, and generation and some other personal info. This information sharing process has both strength and weakness on the smart grids such as, the information share allows a level of integration among the micro grids, on the other hand, it also inevitably makes the micro grid users a target of intruders, which makes the system vulnerable against them [7-8]. The risk of intruders hence leads to a variety of severe consequences such as, leakage of the prosumers personal info, energy cuts, fake communication signal etc. All of these external interference potentially creates an unreliable and insecure power system operations. Therefore, the security of the smart grid becomes one of the biggest challenge to provide enough level of security. SGID [9] defined a guideline for the assessments of the smart grids' security in different level such as, device, building, aggregator and distribution service operator (DSO) levels to prevent from the loss of confidentiality, integrity and availability. Since the ethical issue is one of the key issue in the context of the smart grid, the personal data identifiers has to be prevented from any unpermitted third parties, such as IP address, post code, name and surname, and email address information during the data flow. The ethic assessment defined by EC is clearly stated at EU directive 95/46/EC to protect personal data safely during any communication levels [10].

To protect data confidentiality and to prevent the loss of integrity, there is a need for further research on how best to anonymize the data for multi-level sharing, especially at the low-voltage grid level [11]. Since the needs for anonymization process is related to protection of the personal information. There is a huge number of implementation found especially in medical data protection [12]. As the scientific research on the medical research has a large number of an ethical and privacy issues. The dissemination of the medical record requires the protection against to privacy bridging threads. Hence, the anonymization plays a key role in this area to protect the private information like direct identifiers (name and phone number), quasi identifiers (date of birth, post code, gender) and sensitive attribute (DNA, health conditions such as cancer, HIV, mental health etc.) about the patient. Azarm-Daigle et al. [13] also stated

This work is part of MAS$^2$TERING project – co-funded by the European Union under the 7$^{th}$ Framework Programme (FP7). Grant No. 619682.

that one of the major in the healthcare system occurs during the cross-organizational data sharing process. This process has to consider confidentially of the patient records. As the data regulation and law about the patient data and ownership of the data have to be considered well. Even if there is a requirement to share some information about the patient to satisfy the interoperability. The level of the shared information has to be defined well, and the rest of the information should be in an anonymized form. Recently the anonymization and privacy of the data called *k*-anonymity [14-17]. Further, Gokila and Ventkateswari [18] also highlighted that the publishing process is not only a problem of the medical it is an issue of organizational ethical and data protection issue. Hence, it has to be considered all ethical issues related to dissemination of any personal record.

Sweeney [15] has stated that one of the most robust data protection tool is the usage of *k*-anonymity method. The *k*-anonymity tries to capture private tables (personal details) for re-identification. Further, the algorithm demands every tuple in the data record released be indistinguishable related to no fewer than k correspondents [19].

The re-identification process with k-anonymity is mostly driven by the type of data selected for this process. Hence, several cyber-attack scenarios can be considered to identify the re-identifications process. El Emam and Dankar [20] defined two possible intruders attack on the personal details of the prosecutors and journalist. Hence they used a k-anonymity using cross classification on the quasi identifiers of both person types to reduce the risk of the possible recognitions.

Schrittwieser et al. [21] proposed a *k*-anonymity based fingerprinting anonymization process for the medical health record. The proposed method creates a unique fingerprints of the data set for the partially anonymized data set. The main idea of the proposed approach is to generate the group of the data which has similar level of the anonymity.

Further Sun et al. [22] proposed an extended *k*-anonymity approach on sensitive attributes. The method is based on the adaptation of the *p* sensitive attributes in *k*-anonymity model. Since, the usage of the *p* sensitive attribute may not provide a secure anonymization. The sensitive data attributes can be re-clustered into an adoptive number of the variables.

The new challenge of the anonymization process is now on the smart grid information both on the disseminated records and live information on the existing smart grid devices. Therefore, a k-anonymity based smart grid data anonymization process is presented in this study.

The paper is organized as following: the section two presents the anonymization techniques; the section three presents the *k*-anonymity; the section four presents the proposed methodology and the section five present the preliminary testing and results; and finally the conclusion is presented in section six.

## II. BACKROUND

Anonymization is a type of sanitization process to prevent confidentiality, which transform the data set into an unidentified form. In general, there are two possible scenarios to utilize the anonymization process on them, which are called interactive scenario condition and non-interactive scenario condition [15]. An interactive scenario condition is akin to the use of a statistical database. The non-interactive scenario condition covers the publishing data to run subsequent external queries. These two approaches are illustrated in the fig.1 [15].

The "Interactive Scenario" approach suggests that a user queries data that is kept on a server – and the data is not generally released to other parties. In this approach, the types of queries that can be submitted to the data can be controlled – which increases the potential privacy of the data. However, due to this limit, complex queries cannot be directly supported on the data, making it difficult to undertake more complex analysis and aggregation of data with other sources. This approach assumes that as new data becomes available, the data owner has the ability to control access to it, and to restrict/limit queries that can be submitted to previous versions of the data.

The "Non-Interactive Scenario" involves publishing the data externally – so users can run any queries they prefer whilst keeping the data locally. The benefit of this approach is that an external user can store the data locally and process it in any way desired. A key distinguishing characteristic of this approach is that once the data has been released, it is no longer within the control of the data owner. It is therefore necessary to ensure that enough information is removed from the released data to prevent any privacy breaches.

The choice of the approach determines how much data is "exposed" to external third parties and the complexity of queries that can be carried out on the data. Two general approaches to support anonymization include [18]:

(i) Randomization involves modifying the content of the data set;

(ii) Suppression involves removing values associated with particular attributes to limit possible disclosure.

The approach selection is dependent on needs of the anonymized data to be utilized for further analysis. Hence it affects the anonymization process. This is dues to not to lose the required data with anonymization process. Therefore, the application after anonymization process effects the selection of the anonymization method.

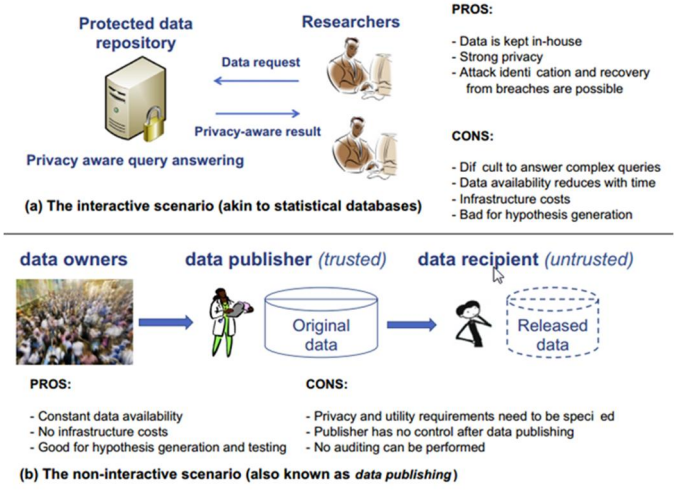


Fig. 1. The possible scenario conditions to implement the anonymization process [15].

To support privacy, the following aspects of the data need to be considered:

- **Direct identifiers**: Attributes that can explicitly re-identify individuals, such as name, mailing address, phone number, other national IDs, and email address.
- **Quasi-identifiers**: Attributes which in combination can lead to identity disclosure, such as demographics (e.g., gender, date of birth, and zip code)
- **Sensitive attributes**: Attributes which individuals are not willing to be associated with (i.e. sensitive information that could reveal a particular condition about an individual)
- **The choice of the algorithm** used to support privacy needs to ensure that the following types of disclosers are prevented:
- **Identity disclosure**: An attacker can associate an individual with their record in a published dataset. This is often a key objective in many anonymization approaches.
- **Membership disclosure**: An attacker can infer with high probability that an individual's record is contained in the published data.
- **Attribute disclosure**: It occurs when an individual is associated with information about their sensitive attributes, i.e. an attacker is able to gain access to a value associate with such an attribute.

Several anonymization techniques exist in literature; e.g., *k*-anonymity, *l*-diversity, *p*-closeness. However, the most popular privacy model for protecting customer demographics is *k*-anonymity [14] and has therefore been adopted in this project.

*k*-anonymity requires each record in a dataset $D$ to contain the same values in the set of Quasi-IDentifier attributes (QIDs) with at least $k-1$ other tuples in $D$ [12]. Recall that quasi-identifiers are typically innocuous attributes that can be used in combination to link external data sources with the published dataset. Satisfying *k*-anonymity offers protection against identity disclosure, because it limits the probability of linking an individual to their record, based on QIDs, to $k^{-1}$. The parameter $k$ controls the level of offered privacy and is set by data publishers.

Another privacy model that has been proposed for demographic data is *k*-map [12], which is similar to *k*-anonymity but considers that the data linking is performed based on larger datasets (called population tables), from which the published dataset has been derived. Thus, *k*-map is less restrictive than *k*-anonymity, typically allowing the publishing of more detailed personal information, which helps data utility preservation. On the negative side, however, the *k*-map privacy model is weaker (in terms of offered privacy protection) than *k*-anonymity because it assumes that:

- The attackers do not know whether a record is included in the published dataset;
- Data publishers have access to the population table.

## III. METHODOLOGY

In this study, a *k*-anonymity based approach has been utilized in the scope of MAS$^2$TERING project which as district level smart grid electricity management and control project funded by European Commission Framework Project 7. The main idea of the anonymization process for the MAS$^2$TERING project is to protect the prosumers' confidential information from intruders. The anonymization is utilized on the metered data supplied from Customer Energy Management System (CEMS) agents to the energy aggregator system to protect the prosumers demographics information. Both the interactive and non-interactive scenarios described above are supported by the behavior of the anonymization component of the aggregator agent. The rendered behavior provides database access and (restricted) query replies to anonymize data for non-trusted agents. The anonymization component of the aggregator exhibits three main activities:

- **Managing communications** – the agent listens for incoming messages from clients. Some of the message types accepted include the specification of metrics for the anonymization privacy and utility and meter data. Additionally, the agent can respond immediately to restricted queries to data.
- **Monitoring internal agent state** – the agent monitors its internal state manifested by several parameters such as the number of pending data set anonymization requests, typical performance of the anonymization loop and any time based default triggering, age of received meter readings that have not been added to the last anonymization output.
- **Performing the anonymization** – using the current setting for protection and utility metrics a near optimum anonymized dataset is generated. The agent aims to reach the required level of protection while maintaining the parameters requested for utility.

The updates form meters are typically at 30 minute intervals consisting of accumulated electrical energy consumption and can be sent in real time or buffered and the anonymization transformation takes the order of seconds for a small dataset. Testing will be carried out with much larger datasets to assess performance. However, the generation of the anonymized database is typically asynchronous as mentioned above.

To fulfil the above operations for the project, several anonymization tool box have been investigated to select the most appropriate one. The primary protection model to be realized for the project is the *k*-anonymization which guarantees not to reveal any of the entity belongs to data set.

The following libraries that support *k*-anonymization were evaluated for the scope of project: ARX-Framework which is an open source framework supporting that provides an API for transforming data sets using generalization, suppression, aggregations among others [23], and UTD - anonymization

toolbox, which is similarly an open source library provisions supporting a wide range of anonymization algorithms over several protection methods. Since, the ARX framework has defined workflows and more compete process support with the provisions of utility and re-identification, the ARX library was selected for use in MASTERING as an implementation engine. The workflow in ARX accommodates the need to achieve the balance between utility and the privacy level rendered and the processes can be iterative. The process begins with the specification of configuration including parameters such as (k-Anonymity) parameters and the generation of a generalization hierarchy. The anonymization process is then invoked followed by analysis of the solution to determine if the required privacy metrics have been met. An optimal transformation is considered to be the transformation that results in minimal information loss according to some metric. Analysis consists of examining the transformed data set to determine re-identification risks and the process can be repeated if the required parameters are not upheld. The transformation process generates a 'transformation lattice'. The main information for transformation node is a Boolean which states if the node is anonymous. The framework also offers the ability to retrieve statistics about the transformation [25]. Further, a comprehensive comparison of both ARX and UTD anonymization toolbox is also presented in table 1.

TABLE I. Comparison of features of anonymization libraries.

| Library / Feature | ARX Framework | UTD - Anonymization Toolbox |
|---|---|---|
| K-Anonymity | Plain k-anonymity | datafly,Mondrian Multi-dimensional K-Anonymity |
| Additional Support | l-diversity, t-closeness, delta disclosure privacy and presence and well as semantic models | Incognito, Incognito with l-diversity, Incognito with t-closeness, Anatomy |
| Utility Analysis | Yes | - |
| Re-identification Risk Analysis | Yes | - |
| Input | File, database drivers, Microsoft excel, object based API | Unstructured text file |
| Output | File, database drivers, Microsoft excel, object based API | Unstructured text file |
| Configuration | Object based API, configuration file | Structured text file XML, command line options |
| Activation | Object based API | Command line |
| Open source | Yes – API and tool | Yes - API |
| Documentation | Good | Adequate |
| Tools | API, GUI, Multi-Platform anonymization Tool | API |

## IV. EXPERIMENTS

To illustrate the performance of the proposed ARX based anonymization process on the smart grid, a sample personal data set is populated and merged with a smart grid energy consumption data due to the lack of full smart grid data set. This data set then is utilized in ARX toolkit using an API. However, the dataset has been loaded into the ARX graphic user tool for the purposes of reporting. Various output artefacts have been generated from the implemented test code or manually duplicated and imported into the user interface tool. Specifically, an abstraction hierarchy was generated for the input. That hierarchy is range-based for level 1 abstraction, and the ranges are fairly arbitrary, so is the level 2 abstraction. The test abstraction hierarchy and input data is shown in Figure 2.

The ARX Framework API provides a range of analysis facilities which will be used in the agent implementation to post-process the personal info in the database to protect against any intruders attack. Further the ARX framework provides several metrics for analysis of re-identification of transformed data. The framework contains following risk models or scenarios for anonymization:

- **Prosecutor** scenario – information about an individual is known to be contained in the dataset;
- **Journalist** scenario – not known if a given individual's data is contained in the dataset; and
- **Marketer** scenario – the objective is to re-identify a large proportion of individuals in the dataset rather than specific individuals.

According to three models the given data set analyzed and results are presented in Figure 3-4.

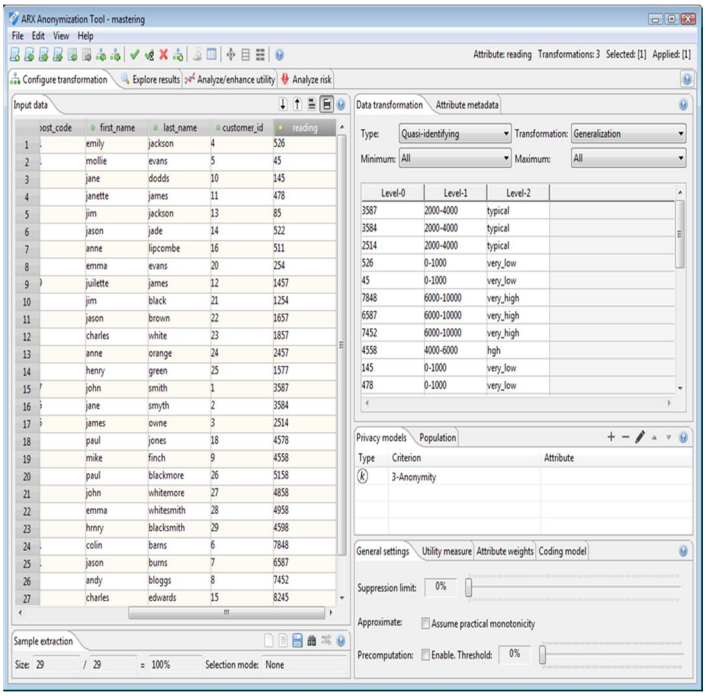



Fig. 2. The visualisation of the input test dataset in ARX toolbox.

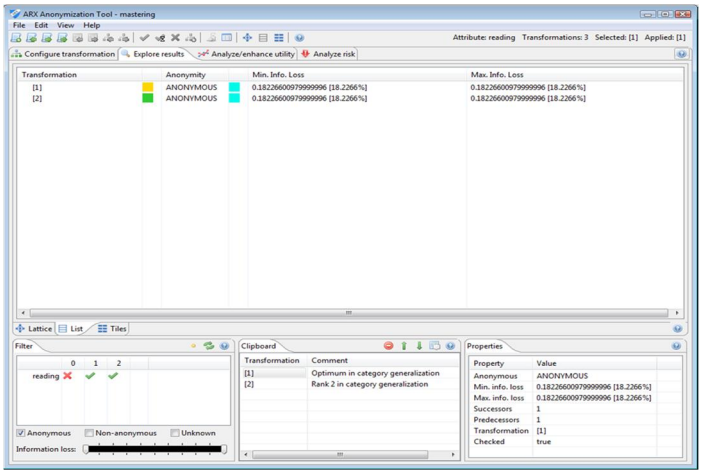
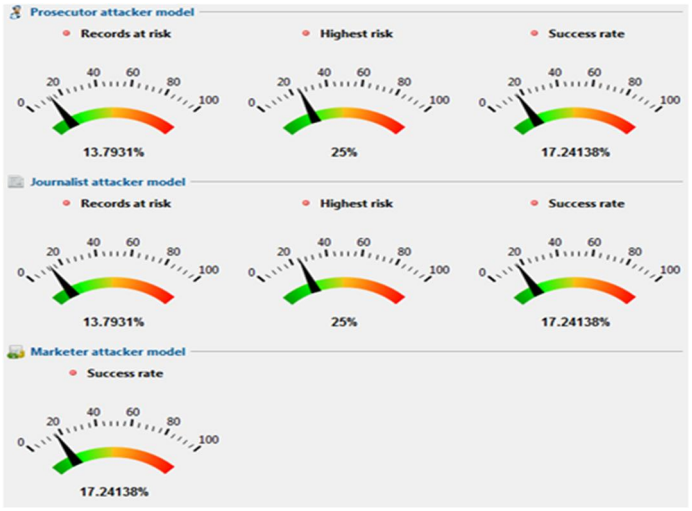


Fig. 3. The sample summary results from ARX user interface.




| Column | Identifier | Instance | Match type | Value/Confidence |
|---|---|---|---|---|
| age | Date | Date/Time | Attribute name | age |
| age | Date | Age | Attribute value | 100% |
| first_name | Name | First name | Attribute name | firstname |
| first_name | Name | First name | Attribute value | 88.88889% |
| first_name | Name | Last name | Attribute value | 77.77778% |
| last_name | Name | Last name | Attribute name | last name |
| last_name | Name | Last name | Attribute value | 80% |
| customer_id | Date | Age | Attribute value | 100% |

Fig. 4. Analysis metrics for the transformed dataset using K-Anonymization where $k = 3$.

TABLE II. COMPARISON OF ANONYMIZATION APPROACHES – RESULT METRICS.

| **Risk Level** | ***k*-anonymity** | ***k*-map** | ***k*-anonymity + *k*-map** |
|---|---|---|---|
| Lowest Prosecutor Risk | 12.5 | 12.5 | 12.5 |
| Record Affected by the Lowest Risk | 27.6 | 27.6 | 27.6 |
| Average Prosecutor Risk | 17.24 | 17.24 | 17.24 |
| Highest Prosecutor Risk | 25 | 25 | 25 |
| Record Affected by Highest Risk | 13.8 | 13.8 | 13.8 |
| Information Loss | 18.2 | 18.2 | 18.2 |

The smart grid data is most likely similar to the prosecutor scenario, where the personal identification should be anonymized with highest level of the conditions. Based on the results found with prosecutor scenario is included in the Table 2 to illustrate the anonymization risk using three anonymization approach; *k*-anonymity, *k*-map and *k*-anonymity + *k*-map.

The ARX toolkit uses a number of *data quality* measures associated with the anonymization process – and influenced by the amount of information loss associated with each attribute in the dataset. Hence, utility measures may either be based on equivalence classes (called single-dimensional) or based on the individual information loss of each attribute (called multi-dimensional). Several quality metrics have been proposed in literature, their aim in one way or another is to minimize the amount of information loss resulting from the generalization and suppression operations that are applied to produce the transformed dataset. According to table 2, all three techniques performed the same, and the information loss under this scenario is about 18.2%. The intruders attach risk is lower when the highest prosecutor scenario performed (25%). Further there is no differences between *k*-anonymity approach and other existing approaches in all possible cases

## V. DISCUSSION AND CONCLUSION

This study presents an anonymization application on the smart grid data in the scope of the MAS$^2$TERING project. Since the personal data protection became one of the most important issue for the EU states, EC issued a legislation rule regarding to any information, which provides the identification information about individuals, should be anonymized [26]. They also stated this anonymization can be addressed via k-anonymity, noise addition, permutation, differential, privacy, aggregation, l-diversity and t-closeness. As discussed in this paper, the k-anonymity based anonymization approach is the most fundamental and popular ones. Hence, a k-anonymity based anonymization approach is utilized on ARX framework via API.

The data set is observed under three possible scenarios as prosecutor, journalist and marketer conditions. The analysis is carried out under prosecutor scenario conditions which means the information about an individual is known to be contained in the dataset. According to the prosecutor scenario conditions, three anonymization algorithms are considered under three level of risk as low, medium and high risks. The lowest risk is found as 12.5% for all anonymization techniques (*k*-anonymity, *k*-map and *k*-anonymity + *k*-map).

The information lost under this scenario is found about 18.2% in all algorithms.

Since the anonymization process is a legal process in the light of directive 95/46/EC and other EU legal instruments, to anonymize the personal data in order to irreversibly prevent identification of personal information, there is no a standardization level to identify the level of the anonymization process [27]. In the proposed approach, the solution is implemented on smart grid. However it is applicable to implement on other industries areas too to prevent the personal information with some simple adaptation adjustments on the quasi identifiers and standardization of the anonymization levels

for these quasi identifiers. For example, the IP address, post code, user id and email address are very common identifiers in the area of the smart systems (smart grid and other advanced communication based systems). Thus, the anonymization for this types of quasi identifiers requires to define level of the anonymization process. Since the over anonymized information may not be utilize for another third parties which may need to utilize them for different purposes according to the different agreements such as the regional internet usage or annual census counts etc. The level of the anonymized data thus should satisfied all involved parties. Hence, the anonymization standardization in the software is important. Since El Hamam[28] proposed three general scenario for the personal data attacks as prosecutor, journalist and marketer conditions, which has been shown as an application in this study too, the information lost with these scenario based models are totally different. Hence during the selection of scenario models both the future implementations with the data usage and sharing level needs to be defined clearly with the proposed model. In future the possibility of the scenario option may be extended for different markets to define a flexible anonymization process with a targeted approach. This methodology can then be discussed in the scope of the standardization of the anonymization models and required quasi identifiers.

The next step of this implementation is to utilize K-Anonymity approach on a live pilot smart grid to evaluate the performance of the algorithm under different scenario conditions. Further, the number of attributes in the data set will also will increase to evaluate the performance of the *k*-Anonymity approach compare to other two to identify the performance boundaries.

## REFERENCES


[1] P. J. G. Pearson, and T. J. Foxon, "A low carbon industrial revolution? Insights and challenges from past technological and economic transformations," *Energy Policy*, vol.50, pp. 117-127, Aug. 2012.

[2] M. Mourshed, S. Robert, A. Ranalli, T. Messervey, D. Reforgiato, R. Contreau, A. Becue, K. Quinn, Y. Rezgui, and Z. Lennard, "Smart Grid Futures: Perspectives on the Integration of Energy and ICT Services," *Energy Procedia*, vol. 75, pp. 1132-1137, Aug. 2015.

[3] EC, *European Technology Platform SmartGrids: Vision and Strategy for Europe's Electricity Networks of the Future*, Office for Official Publications of the European Communities: Brussels, 2006. [Online], Available from: https://ec.europa.eu/research/energy/pdf/smartgrids_en.pdf. (Last access: 22 March 2016).

[4] H. Farhangi, "The path of the smart grid," IEEE Magazine on Power and Energy, vol. 8, no. 1, pp. 18-28, Jan. - Feb.2010.

[5] H. L., Chao, C.C. Tsai, P.A. Hsiung, and I.H. Chou, "Smart Grid as a Service: A Discussion on Design Issues," *The Scientific World Journal*, vol. 2014, pp. 1-11, Aug. 2014.

[6] A. P. García, J. Oliver and D. Gosch, "An intelligent agent-based distributed architecture for Smart-Grid integrated network management," In proc. of the IEEE 35th Conference on Local Computer Networks (LCN), 2010, Denver, CO, 2010, pp. 1013-1018.

[7] A.R. Metke, R.L. Ekl, "Smart grid security technology," In proc. of Innovative Smart Grid Technologies Conference Europe (ISGT), 2010.

[8] W. Weng and Z. Lu, "Cyber security in the Smart Grid: Survey and challenges," *Computer Networks*, vol. 57. No, 2013, pp. 1344-1371, Jan. 2013.

[9] SGID, *Introduction to NISTIR 7628 Guidelines for Smart Grid Cyber Security*, pp.1–597, Sep. 2010. [Online]. Available from: http://www.nist.gov/smartgrid/upload/nistir-7628_total.pdf. (Last access: 22 March 2016).

[10] EC, Directive 95/46/EC of the European parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Oct. 1995.

[11] I. L. G. Pearson, "Smart grid cyber security for Europe," *Energy Policy*, vol. 39, no. 9, pp. 5211-5218, Sep. 2011.

[12] A. Gkoulalas-Divanis, G. Loukides, J. Sun, "Publishing data from electronic health records while preserving privacy: A survey of algorithms," *Journal of Biomedical Informatics*, vol. 50, pp. 4-19, Aug. 2014.

[13] M. Azarm-Daigle, C. Kuziemsky, L. Peyton, "A Review of Cross Organizational Healthcare Data Sharing," *Procedia Computer Science*, vol. 63, pp. 425-432, Sep. 2015.

[14] P. Samarati. Protecting respondents' identities in microdata release. In IEEE Transactions on Knowledge and Data Engineering, 2001.

[15] L. Sweeney, "k-anonymity: a model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol.10, no.5, pp. 557-570, Oct. 2012.

[16] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *In proc. 22nd Intnl. Conf. Data Engg. (ICDE)*, pp. 24, 35, 2006.

[17] N. Li, T. Li and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," *In proc. of the IEEE 23rd International Conference on Data Engineering - Istanbul*, pp. 106-115, Apr.2007.

[18] S.Gokila and P.Venkateswari, "A Survey On Privacy Preserving Data Publishing," International Journal on Cybernetics & Informatics ( IJCI), vol. 3, no, 1, pp. 1-8, Feb. 2014.

[19] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, P. Samarati, *k-Anonymity,* In. Secure Data Management in Decentralized Systems T. New York: Springer, 2007, pp. 323-354.

[20] K. El Emam and F. K. Dankar, "Protecting Privacy Using k-Anonymity," *Journal of the American Medical Informatics Association*, vol. 15, no. 5, pp. 627-637, Sep.–Oct. 2008.

[21] S. Schrittwieser, P. Kieseberg, I. Echizen, S. Wohlgemuth, N. Sonehara, E. Weippl, *An algorithm for k-anonymity-based fingerprinting*. In Eds. IWDW 2011. LNCS, Heidelberg: Springer, 2012, vol. 7128, pp. 439–452, 2011.

[22] X. Sun, L. Sun and H. Wang, "Extended k-anonymity models against sensitive attribute disclosure," *Computer Communications*, vol. 34, no. 4, pp. 526-535, Apr. 2011.

[23] ARX, *ARX – Powerful Data Anonymization : A comprehensive software for risk- and utility-based privacy-preserving microdata publishing,* [Online]. Available from: http://arx.deidentifier.org/development/framework/. (Last Accessed: 25 April 106).

[24] UTD, UTD: Anonymization ToolBox, [Online]. Available from: http://www.cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php. (Last Accessed: 25 April 106).

[25] F. Kohlmayer, F. Prasser and K. A. Kuhn, "The cost of quality: Implementing generalization and suppression for anonymizing biomedical data with minimal information loss," *Journal of Biomedical Informatics*, vol. 58, pp. 37-48, Dec. 2015.

[26] EC, *Data Protection Working Party.* [Online], Available from: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. (Last Accessed: 25 April 106).

[27] Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T. and Estabrooks, C. A. Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. *Forum:Qualitative Social Research*, vol, 6, no1, Art29, Jan. 2005.

[28] El Hamam, K. *Guide to the De-Identification of Personal Health Information,* Florida, USA, CRC Pres, 2013, pp. 197-203.