Distributed Energy Resource Management Systems: Preserving Customer Privacy through K-Anonymity

Mohammed Alsaid, Nirupama Bulusu
Computer Science
Portland State University
Portland, OR, USA
mohamm.alsaid@gmail.com, nbulusu@pdx.edu

Midrar Adham, Robert B. Bass
Electrical and Computer Engineering
Portland State University
Portland, OR, USA
midrar@pdx.edu, robert.bass@pdx.edu

2023 IEEE Power & Energy Society General Meeting (PESGM) | 978-1-6654-6441-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PESGM52003.2023.10252650

***Abstract*—The smart grid represents the next generation of electricity distribution systems that utilizes recent technological innovations. It uses digital communication between its components and entities to attain more automation, self-sufficiency, and reliability. One of the many concerns in smart grid digital communication discussions is the possibility of violating customers' privacy. Violating customers' privacy imposes a significant barrier as smart grid desirable attributes are tightly tied to customers' participation. Employing privacy models can address concerns regarding information privacy in smart grid digital communication. In this work, we provide an approach to utilizing K-anonymity to ensure data within the system excludes Personally Identifiable Information. Results suggest that a dynamically generated generalization hierarchy minimizes information loss incurred by the anonymization process.**

***Index Terms*—Anonymization, privacy, K-anonymity, Distributed Energy Resource, DERMS**

I. Introduction

The traditional power distribution concept is becoming outdated, predominantly because it has yet to keep pace with recent technological advancements. Arguably, it is the most complex system ever created. Nevertheless, this comes with disadvantages as well as virtues. Empirical data bring to light the irreversible side effects of the traditional approach. Indeed, the evidence of carbon emissions produced by power generation is undeniable.

The research community has been exploring the new concept and its shortcomings. In particular, the cyber-security and privacy of Smart Grid (SG) subsystems have been widely studied areas of research. This work extends that foundational work to provide security and privacy in a SG implementation.

Applying K-anonymity to SG components is not a partially novel notion in that there are many similar works. For instance, Stegelmann and Kesdogan proposed a privacy-preserving smart metering architecture [1]. This approach provides means for collecting energy consumption information without violating consumers' privacy. However, smart metering is only one component of the much broader concept of SGs, which Energy Grid of Things (EGoT) Distributed Energy Resource Management System (DERMS) attempts to address.

Similarly, Yuce et al. studied solutions for consumer data privacy in a district-level microgrid [2]. They obtained privacy guarantees using k-anonymity for consumers' demographic and associated energy consumption information. This approach differs from the EGoT DERMS approach as the level of operation is much broader and attempts to apply k-anonymity on the trust layer.

And, Li et al. proposed an approach focusing on demand response in microgrids using vehicle-to-vehicle technology [3]. The approach adds a privacy-preserving attribute to their auction scheme by applying k-anonymity to achieve location privacy guarantees.

All the previously examined works consider applying anonymization to some aspects of SG implementations. This work is not different in that it also explores the K-anonymity application for the EGoT DERMS. Rather, it differs from other works in applying the anonymization technique to the trust layer. None of the discussed works explore anonymization combined with a trust layer concept.

II. Background

A. The Energy Grid of Things

The EGoT DERMS is Portland State University (PSU)'s implementation of a SG system. It employs Service-Oriented Load Control (SOLC) methods for Demand-Side Management (DSM), and was designed with interoperability in mind. There is heterogeneity in the types of protocols supported by smart appliance manufacturers. Hence, the EGoT DERMS relies on IEEE 2030.5 as the primary protocol for communication between entities when possible [4]. The protocol allows for the maximum degree of flexibility that can be leveraged to accommodate the largest number of off-the-shelf products [5]. Figure 1 presents a conceptual view of the overall structure of EGoT DERMS.

In the EGoT DERMS, customers who own smart appliances are called Service Provisioning Customers (SPCs), and aggregators are referred to as Grid Service Providers (GSPs). Ideally, each GSP can dispatch a large number of Distributed Energy Resources (DERs) such that the grid services it provides to a Grid Operator (GO) are impactful [6]. The motivation is that SPCs own grid-enabled appliances, DERs, that can provide various DER services to a GSP. The GSP can use these appliances in large numbers to provide grid

This work was supported by US DOE OE0000922.

978-1-6654-6441-3/23/$31.00 ©2023 IEEE
Authorized licensed use limited to: PORTLAND STATE UNIVERSITY LIBRARY. Downloaded on October 02,2023 at 19:27:52 UTC from IEEE Xplore. Restrictions apply.

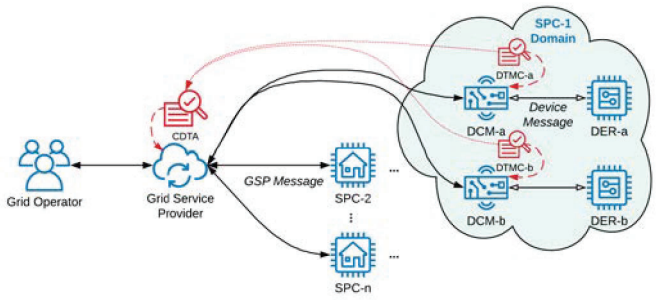

Fig. 1. An overview of the system architecture. The trust layer (shown in red) comprises the DTM System.

services that meet grid Operators' operation objectives, as needed depending on the state of the grid.

GOs are entities that manage the grid to achieve operational objectives. The operational objectives can be to either to maintain operations within the physical constraints that must be honored to prevent damage to grid components and equipment; or attain operational goals associated with stable, reliable, economical delivery of power at nominal conditions. To do so, GOs seek grid services from GSPs to meet their operational objectives. Note that the GSP provides **grid** services to the GOs. It does that by using the offered **DER** services provided by the DERs.

Due to the variability of DER manufacturers and the heterogeneity of the protocols they obey, there must be a mechanism for interoperability. Interoperability is accomplished through software and hardware support. Distributed Control Modules (DCMs) in the system are tasked with expanding DER functionalities such as the support of IEEE 2030.5, scheduling, and network communication. Therefore, DCMs are the realization of hardware and software support for interoperability [7].

B. The Trust Model

Trust is a notion with multiple definitions derived from various disciplines. Generally speaking, it is the degree of reliance an entity can place on another to achieve an objective [8]–[10]. This definition is relevant to distributed systems such as the EGoT DERMS where reliability plays a crucial part [11]. Most importantly, the trust model provides a detective, passive role for the EGoT DERMS. Namely, it monitors communication between actors without interfering. The trust model is referred to as the DTM System.

The DTM System comprises two types of actors: many Distributed Trust Model Clients (DTMCs) and one corresponding Central Distributed Trust Aggregator (CDTA). The DTMCs are components placed adjacent to DCMs, as shown in Figure 1. These DTMCs monitor their respective DCMs without interfering with the DCMs functionalities in an effort to measure trust in the system. Each DTMC measures the trust by monitoring the DCM communication with other actors in the system, specifically the DER and the GSP. The DTMC is able to quantifying the local trust of the DCM, DER, and GSP by developing a communication fingerprint of each actor, which is referred to as a Metric Vector of Trust (MVoT). Finally, DTMCs send their local trust information, the MVoTs, to the CDTA where the distributed trust is aggregated and an overall trust of the EGoT DERMS is computed.

C. Common Smart Inverter Profile v2.0

Among the many propositions put forth in the Common Smart Inverter Profile (CSIP) standard is the topological grouping of DERs. Figure 2 illustrates the topological and non-topological groupings as described in CSIP. The Figure depicts a topology tree on which several service points are located. Note that only several paths are highlighted, and the rest are omitted for clarity purposes. The topological location of each node is the result of concatenating all its ancestors. This location also represents the physical location of each node. For example, node D1, which corresponds to a feeder, is physically connected to substation C1. Notice that each node in the Figure is a group itself. In addition, the grouping needs not to be topological. For instance, Group-Z shown in the Figure, does not conform to the topology. Instead, it is placed according to the utility needs. Given that the EGoT DERMS adopts IEEE 2030.5, it is only natural to adopt the CSIP grouping. However, note that the non-topological groups are not considered in this stage of EGoT DERMS and, by extension, are outside the scope of this work.

D. K-Anonymity

Many organizations aim to publish microdata for research purposes, such as demographic, health, and other data domains. However, such microdata may contain Personally Identifiable Information (PII) that breaches the privacy of customers, patients, and citizens. For example, combining published data with publicly available external data sets can pinpoint individuals even though the microdata's obvious

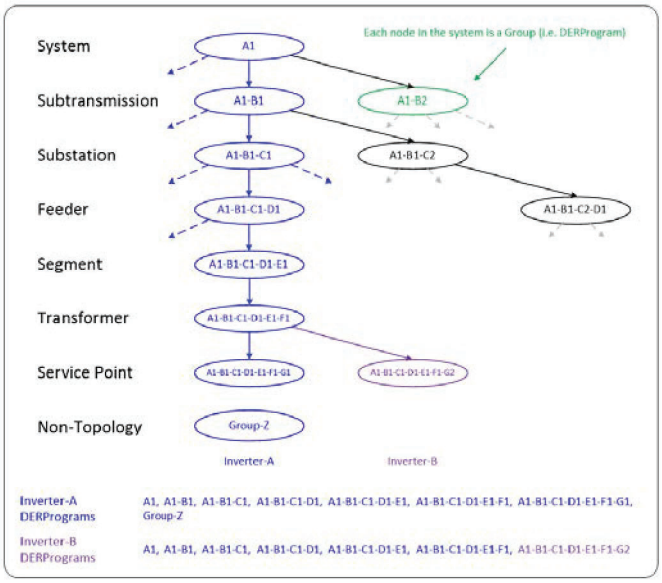

Fig. 2. Topological grouping as described in CSIP v2.0 [12].

978-1-6654-6441-3/23/$31.00 ©2023 IEEE
Authorized licensed use limited to: PORTLAND STATE UNIVERSITY LIBRARY. Downloaded on October 02,2023 at 19:27:52 UTC from IEEE Xplore. Restrictions apply.

PII was removed. Sweeny demonstrated this in 2002 by re-identifying individuals from public health records, which resulted in exposing the health records of Massachusetts governor William Weld [13].

Sweeny proposed K-anonymity to protect individuals' privacy and reduce the chances of launching successful re-identification attacks. The key idea is based on aggregating records in the data such that each record has at least $K-1$ identical records (K is a user-defined number of identical records desired). K-anonymity is conditioned on producing valuable anonymized data to fulfill the purpose of publishing data to advance research.

The problem of optimal K-anonymity is classified as an NP-hard problem even with simple restrictions [14]. Consequently, finding an optimal solution in a reasonable time is not easy. An optimal solution means the data set is anonymized optimally according to various metrics. Due to the inherent hardness of the problem, it is crucial to identify efficient methods of finding/approximating a good enough solution: a solution that does not cause significant information loss.

III. Methods

A. The Mondrian Algorithm

LeFevre, DeWitt, and Ramakrishnan proposed a multi-dimensional model for k-anonymization and a greedy algorithm for k-anonymization [15]. This Mondrian algorithm aims to approximate the optimal anonymization contrasted with finding it. Essentially, it finds a solution by partitioning the instances with respect to all quasi-identifiers in a Mondrian manner. That is, all partitions used are axis-aligned. The proposed approach has a far better complexity than previously proposed methods for achieving K-anonymity. The fact that it relies on a greedy algorithm gives us the benefit of achieving anonymization in $O(nlogn)$ time complexity.

The Mondrian assigns a penalty cost for each tuple **T** in the anonymized view **V**. The most straightforward penalty metric applicable is the discernibility metric (C_{DM}). It computes the penalty based on the number of tuples in each equivalence class. The metric is defined as:

$$C_{DM} = \sum_{E \in EquivClasses} |E|^2 \tag{1}$$

LeFevre et al., however, proposed an alternative metric for calculating the cost penalty called the *Normalized average equivalence class size metric* (C_{avg}). C_{avg} is defined by the following:

$$C_{avg} = \frac{NumOfRecords/NumOfEquivClasses}{K} \tag{2}$$

Both metrics penalize classes with more records. While classes with fewer records might be desirable in some cases, the metrics do not capture the distribution in the quasi-identifier attributes space [16]. A more accurate metric, the Normalized Certainty Penalty (NCP), accounts for the cardinality of the equivalence classes and the scope of the quasi-identifier attributes space [17]. NCP can be defined for numerical attributes as follows, where C is the equivalence class, and A is a numerical attribute:

$$NCP_A(C) = \frac{max_A^C - min_A^C}{max_A - min_A} \tag{3}$$

Equation 3 contains a definition of NCP that would not work for categorical attributes as the concept of distance is non-existent. For such a case, the metric can be defined as follows:

$$NCP_A(C) = \frac{size(u)}{|A|} \tag{4}$$

where $|A|$ is the number of distinct values of attributes of the categorical A, u is the closest common ancestor in the generalization hierarchy for the attribute value, and $size(u)$ is the number of leaves in the sub-tree of u. Additionally, NCP can be converted into a percentage by dividing the NCP value over the number of values in the data set; such a percentage is more comprehensible and thus used as the primary metric for information loss in this work. Finally, keep in mind that all attributes in EGoT DERMS topological IDs are categorical, which means Equation 4 is the equation used to compute the penalty.

B. Generalization Hierarchy

The Mondrian algorithm utilizes a generalization hierarchy to generalize, or suppress, attribute values. This reliance on generalization hierarchy aligns with the topological load groupings in distribution systems. For example, every load has a topological location that describes its associated substation, segment, feeder, and service point to which it is connected, as described in Figure 2. This topological location is used as an identifying value for loads in an electrical and distribution system. However, this work uses the distribution part of the topological hierarchy to create the IDs, starting from the substation down to the service point. Such topology can be morphed and used as a generalization hierarchy for the Mondrian algorithm. Figure 3 shows an example hierarchy constructed for the service point (substituted by DERs instead) attribute in the ID. For this example, both K and H have the same value of five.

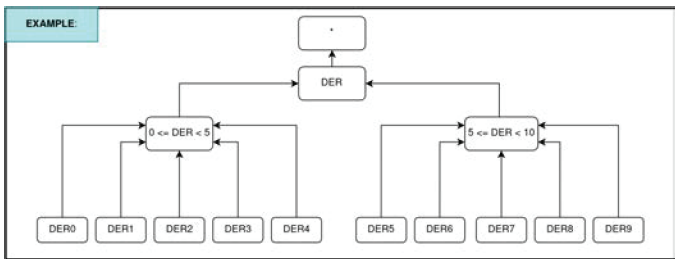

Fig. 3. An example of DER attribute hierarchy in the system, where K is 5, and H is five.

978-1-6654-6441-3/23/$31.00 ©2023 IEEE
Authorized licensed use limited to: PORTLAND STATE UNIVERSITY LIBRARY. Downloaded on October 02,2023 at 19:27:52 UTC from IEEE Xplore. Restrictions apply.

IV. Results

A. K-Anonymity

Since a GSP has access to records of all DCMs registered with its DERMS, such a dataset was used to produce anonymized data. This data set was generated according to the IEEE 13-node feeder design, which is a test feeder used as the system is still in development. Figure 4 shows the visible effects of the H value as the records are aggregated in groups of twos, threes, and fives, which is greater or equal to two and satisfies the 5-anonymization requirement.

Effects of the Mondrian Algorithm on a data set with K = 2

substation	segment	transformer	DER
substation 0	segment 9	transformer 2	DER 0
substation 0	segment 2	transformer 1	DER 1
substation 0	segment 1	transformer 3	DER 2
substation 0	segment 8	transformer 0	DER 3
substation 0	segment 7	transformer 4	DER 4
substation 0	segment 7	transformer 4	DER 5
substation 0	segment 7	transformer 4	DER 6
substation 0	segment 7	transformer 4	DER 7
substation 0	segment 7	transformer 4	DER 8
substation 0	segment 7	transformer 4	DER 9
substation 0	segment 7	transformer 4	DER 10
substation 0	segment 7	transformer 4	DER 11
substation 0	segment 7	transformer 4	DER 12

=>

substation	segment	transformer	DER
substation 0	segment 0-5	transformer 0-5	DER 0-5
substation 0	segment 0-5	transformer 0-5	DER 0-5
substation 0	segment 5-10	transformer 0-5	DER 0-5
substation 0	segment 5-10	transformer 0-5	DER 0-5
substation 0	segment 5-10	transformer 0-5	DER 0-5
substation 0	segment 7	transformer 4	DER 5-10
substation 0	segment 7	transformer 4	DER 5-10
substation 0	segment 7	transformer 4	DER 5-10
substation 0	segment 7	transformer 4	DER 5-10
substation 0	segment 7	transformer 4	DER 5-10
substation 0	segment 7	transformer 4	DER 10-15
substation 0	segment 7	transformer 4	DER 10-15
substation 0	segment 7	transformer 4	DER 10-15

Fig. 4. Sample of 2-anonymization effects on IEEE 13 node feeder data. The table on the left contains records sampled from the IEEE 13 node feeder topology, whereas the table on the right contains the records after anonymization.

Since the Mondrian Algorithm provides multi-dimensional K-anonymity, it takes into account recording all sensitive attributes when anonymizing. For instance, all the attributes in the highlighted record in Figure 4 were used when constructing equivalence classes. Suppose the algorithm did not account for all attributes, which means it is not multi-dimensional. Without the DER attribute, the record could easily be part of the equivalence classes below it in the resulting table due to matching values in all attributes except the DER. However, the algorithm would suppress the DER value to achieve the K-anonymity property, which not only would increase the NCP penalty, but the record would be the first in an equivalence class by itself. This leads to other records being suppressed such that the table meets the K-anonymity property and more penalties.

B. Information Loss

The anonymization degree, which in this case is denoted by K, exerts influence on the information loss observed in the resulting data set. Figure 5 demonstrates the information loss observed when the algorithm is run on the 13-node test feeder data set under various K-values. Figure 5 shows that as K grows in size, the penalty grows slowly, approaching 20%. This behavior is because the algorithm finds fewer and fewer ways to partition data that contains five attributes, with four attributes containing value variations. The fifth attribute, the substation attribute, is intentionally treated as an insensitive attribute, primarily due to the need for having accessibility to such information for the DERMS.

Figure 6 shows a plot of information loss against variable values of K. In Figure 6, only the first half of the data set

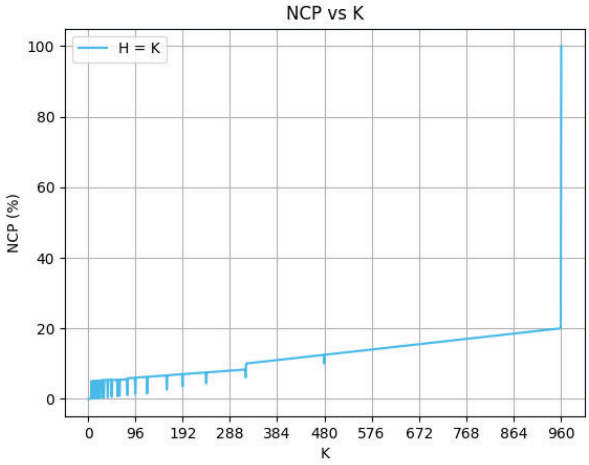

Fig. 5. Plot of NCP against different K values for IEEE 13 node feeder data.

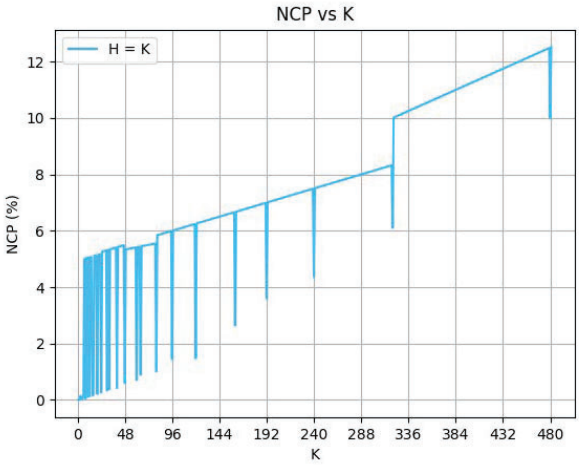

Fig. 6. Plot of NCP against different K values for half of IEEE 13 node feeder data.

was plotted. This was done to zoom into the behavior the algorithm displays when K is relatively small. Note that there are recurring periodic dips up to the half-point where K is equal to one-half the size of the data set. These frequent dips are caused by the greedy algorithm finding and picking new, better partitions that result in less information loss. Such behavior indicates the existence of H values that produce optimal structure such that it minimizes information loss with respect to the IEEE 13-node feeder topology.

C. Information Loss

As discussed earlier, results shown in Figure 5 and Figure 6 indicate the existence of some H values that minimize information loss. Figure 7 and Figure 8 demonstrate the performance of a simple heuristic used to reduce the information loss for the used scheme and the generalization hierarchy used in this work. The heuristic relies on finding such H in advance to pick the best H value for a given K. Finding

978-1-6654-6441-3/23/$31.00 ©2023 IEEE
Authorized licensed use limited to: PORTLAND STATE UNIVERSITY LIBRARY. Downloaded on October 02,2023 at 19:27:52 UTC from IEEE Xplore. Restrictions apply.

H in advance requires one to empirically sample H values based on the results obtained by naively setting H equal to K. The dependence on prior knowledge of the underlying is a significant limitation of the heuristic described here.

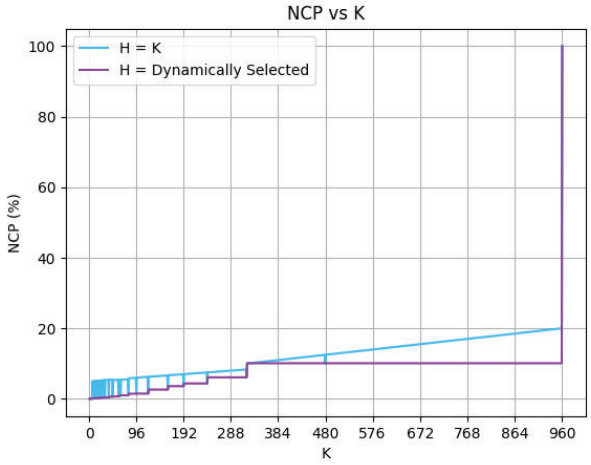

Fig. 7. Plot of NCP against different K values using two different heuristics. Choosing H to equal K results in higher overall penalty incurrence than dynamically selecting H.

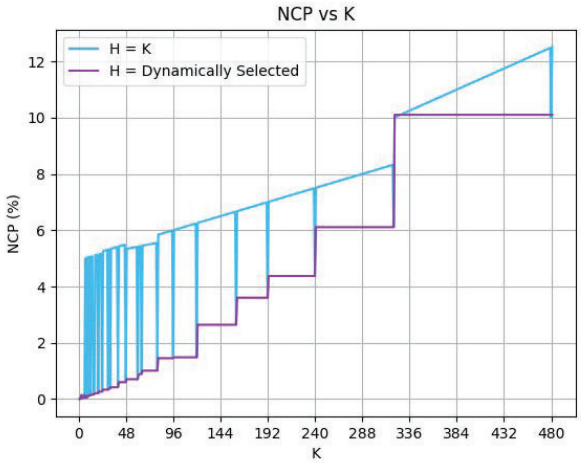

Fig. 8. Plot of NCP against K values using different heuristics for half of the data from Figure 7. A staircase effect can be observed when H is dynamically selected where each step corresponds to dips in the naive heuristic.

V. Conclusion

The EGoT DERMS adopts a Service-Oriented Load Control approach to manage Distributed Energy Resource. This architecture relies on heavy digital interaction between the system actors to achieve its operational objectives. The digital information exchange could potentially infringe upon customers' privacy. Guarantees of privacy promote customer participation, which boosts the system's ability to counterbalance disruptive events using large aggregations of DERs.

This work proposes a privacy-preserving strategy for the EGoT DERMS trust layer. The method involves using K-anonymity to guarantee communication on the trust layer to exclude PII. Also, the strategy secures the communication channel according to IEEE 2030.5 specifications. Findings suggest that the generalization hierarchy for the 13-node feeder shows an Identical Generalization Hierarchy. Such guarantees of identicalness would not hold in a real-world setting.

References

[1] Mark Stegelmann and Dogan Kesdogan. Gridpriv: A smart metering architecture offering k-anonymity. In *IEEE 11th Int. Conf. on Trust, Security and Privacy in Comp. & Comm.*, pages 419–426, 2012.

[2] B. Yuce, M. Mourshed, Y Rezgui, and O. F. Rana. Preserving prosumer privacy in a district level smart grid. In *IEEE Int. Smart Cities Conf.*, pages 1–6, 2016.

[3] Donghe Li, Qingyu Yang, Dou An, Wei Yu, Xinyu Yang, and Xinwen Fu. On location privacy-preserving online double auction for electric vehicles in microgrids. *IEEE Internet of Things J.*, 6(4):5902–5915, 2019.

[4] Midrar A. Adham, Manasseh Obi, and Robert B. Bass. A field test of direct load control of water heaters and its implications for consumers. In *2022 IEEE Power & Energy Soc. Gen. Meeting*, pages 1–5, 2022.

[5] Manasseh Obi, Tylor Slay, and Robert Bass. Distributed energy resource aggregation using customer-owned equipment: A review of literature and standards. *Energy Reports*, 6:2358–2369, November 2020.

[6] Thomas Clarke, Tylor Slay, Conrad Eustis, and Robert B. Bass. Aggregation of residential water heaters for peak shifting and frequency response services. *IEEE Open Access J. of Power and Energy*, 7:22–30, January 2020.

[7] Mohammed Alsaid, Nirupama Bulusu, Abdullah Bargouti, N. Fernando, John Acken, Tylor Slay, and Robert Bass. Privacy-preserving information security for the energy grid of things. *SMARTGREENS*, 2022, 4 2022.

[8] Xinxin Fan, Ling Liu, Rui Zhang, Quanliang Jing, and Jingping Bi. Decentralized trust management. *ACM Computing Surveys*, 53(1):1–33, January 2021.

[9] Abe Singer and Matt Bishop. Trust-based security; or, trust considered harmful. In *New Security Paradigms Workshop*. ACM, October 2020.

[10] Zheng Yan. A comprehensive trust model for component software. In *Proc. of the 4th Int. Workshop on Security, Privacy and Trust in Pervasive & Ubiquitous Computing*. ACM Press, 2008.

[11] N. S. Fernando, J. M. Acken, and R. B. Bass. Developing a distributed trust model for distributed energy resources. In *IEEE Conf. on Tech. for Sust.*, 2021.

[12] SunSpec Alliance. Common Smart Inverter Profile (CSIP), March 16, 2018.

[13] Latanya Sweeney. K-anonymity: A model for protecting privacy. *International J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, October 2002.

[14] R.J. Bayardo and Rakesh Agrawal. Data privacy through optimal k-anonymization. In *21st International Conference on Data Engineering*, pages 217–228, 2005.

[15] K. LeFevre, D.J. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In *22nd Int. Conf. on Eng.*, pages 25–25, 2006.

[16] Gabriel Ghinita, Panagiotis Karras, Panos Kalnis, and Nikos Mamoulis. Fast data anonymization with low information loss. In *Proc. of the 33rd Int. Conf. on Very Large Data Bases*, page 758–769, 2007.

[17] Jian Xu, Wei Wang, Jian Pei, Xiaoyuan Wang, Baile Shi, and Ada Wai-Chee Fu. Utility-based anonymization using local recoding. In *Proc. of the 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, page 785–790, New York, NY, USA, 2006. Association for Computing Machinery.

978-1-6654-6441-3/23/$31.00 ©2023 IEEE
Authorized licensed use limited to: PORTLAND STATE UNIVERSITY LIBRARY. Downloaded on October 02,2023 at 19:27:52 UTC from IEEE Xplore. Restrictions apply.