Title

Data Privacy for the Grid: Toward a Data Privacy Standard for Inverter-Based and Distributed Energy Resources

Journal

IEEE Power and Energy Magazine, 21(5)

ISSN

1540-7977

Authors

Currie, Robert
Peisert, Sean
Scaglione, Anna
et al.

Publication Date

2023

DOI

10.1109/mpe.2023.3288595

Data Privacy for the Grid

Toward a Data Privacy Standard for Inverter-Based and Distributed Energy Resources

Robert Currie, Sean Peisert, Anna Scaglione, Aram Shumavon, and Nikhil Ravi

The traditional approach to planning the distribution grid has focused on reliability in the context of gradual and reasonably predictable load growth. Forecasts of load growth, combined with asset management practices, were used by system planners to identify upgrades to the system to maintain or improve reliability. The decisions, typically based within load flow analysis tools, included considerations about contingency scenarios and corporate forecasts (i.e., top-down predictions at a summary level of what will happen in a particular area that could impact load growth and behavior). Today, this traditional approach no longer fits *all* purposes.

At the top of the list of reasons to question conventional wisdom about planning are the rapid growth in Distributed Energy Resources (DER) and the electrification of transportation and residential heating, which have the potential to radically alter the characteristics of the load on the system, both in terms of magnitude, duration, and timing with corresponding impacts on the need for distribution grid capacity. In addition, all these edge resources have embedded intelligence as well as network communication capabilities, which are becoming faster and more reliable. In turn, a large amount of data is becoming available on customer demand, behavior, technology adoption and a variety of DER devices connected through smart inverters. The collection of data from the edge device networks is aimed at improving both the planning and the operation of the grid; however, sharing them creates several cyber security and data privacy issues since distribution-level data include Personally Identifiable Information (PII). It is extremely important that this data be shared in a manner appropriate for Critical Energy Infrastructure Information. This article's goal is to highlight the main technologies that can be harnessed to define industry standards on solid scientific ground and the need to tailor them to address the emerging energy sector data needs. While covering different approaches to address data security, the focus is on a statistical framework called Differential Privacy, which has emerged as the most reliable way to *open the data* to different stakeholders while at the same time preventing leakage of sensitive data attributes and PII. This method has not been codified as a grid standard and was not considered in the North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection development, a cyber security

framework defined for the identification and protection of critical cyber assets to support the reliable operation of Bulk Electric System.

1. Why is Grid Data Needed?

To accommodate distributed energy resources (DER) both in front of and behind the meter, it is important to understand customer behavior and technology performance. This knowledge will also help to leverage demand response (DR), especially with electric vehicles; DR will complement grid storage and reduce congestion. Improved data sharing and visibility on the power grid are necessary for planning the interconnection of grid-scale renewables and storage, including where capacity is available and where DER can add value. This dynamic impacts a range of existing or emerging activities (including those shown in) that are necessary to support the decarbonization of the economy.

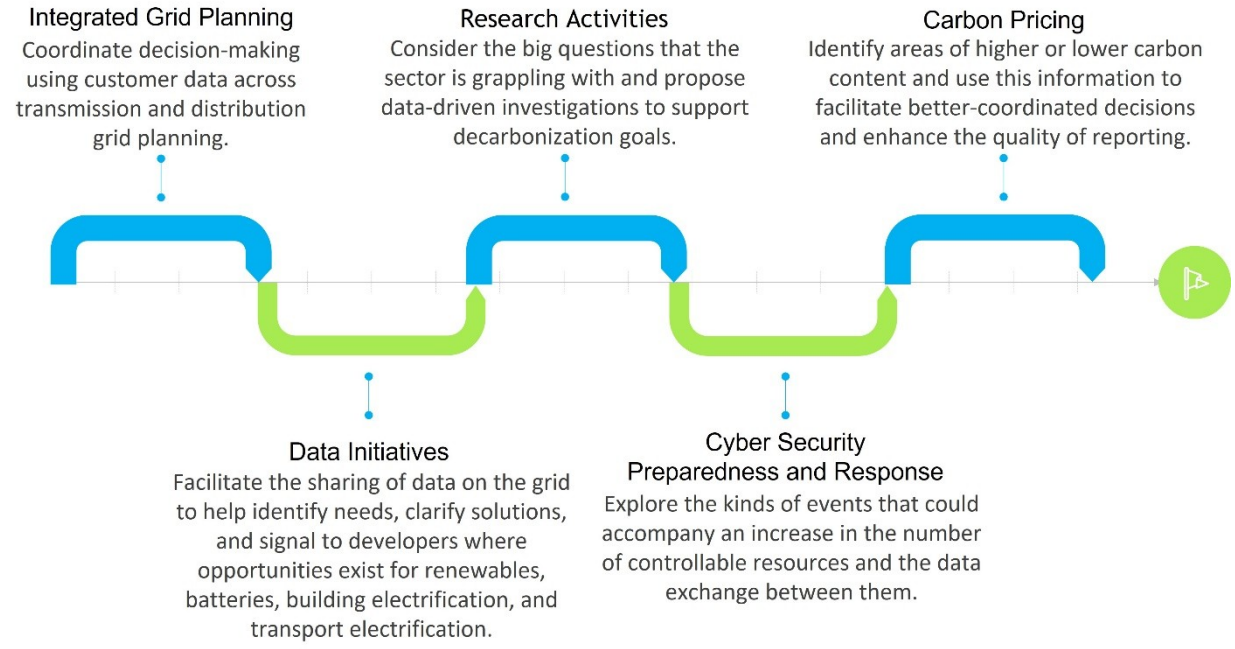

Figure 1 Activities necessary to support the decarbonization of the economy.

2. Emerging Trends in Cyber Security and Data Privacy

It is against this backdrop that this article discusses where cyber security and data privacy concerns are impacting the planning and operation of inverter-based and distributed energy resources. We begin by considering the following trends that are emerging as the industry grapples with these challenges:

a. Increasing Governmental Oversight

Regulators or governments require utilities to share more data. Due to the emerging nature of this type of activity, there are inconsistencies in the treatment of security and data privacy concerns. Some recent developments include the

Integrated Energy Data Resource program in New York and the mandated use of the Common Information Model for sharing grid data in the UK.

b. Aging Data Handling Workflows

The distribution section of the grid suffers from a lack of investment in the acquisition and management of data. Planning models, geospatial data, and operational data are often managed separately and suffer from a range of quality and accuracy issues. These are not simple issues to resolve and require investment and new processes and procedures at the utility, which takes time and funding. Collating and analyzing low-quality data can be challenging, as the data often requires skilled pre-processing. These difficulties may not be immediately apparent until the data is used for various purposes. In light of the developments that were discussed in Section 2.a, the industry will have to first confront the issues related to aging data handling workflows. For example, although data shared about the grid is intended to simplify the process of submitting planning applications for solar developers, it may require a significant amount of specialized knowledge to cleanse, process, and use such data effectively. As a result, developers may need to maintain constant communication with the utility to address any issues that arise. More generally, it may be possible to simultaneously address some of the problems related to the aging infrastructure via an industry-wide standardization effort. A positive development is happening in the communication networking industry, which is defining new standards for the Industrial Internet of Things (IIoT). The aim of these standards is to enable what is referred to as “Industry 4.0,” through a flexible set of communication protocols that can be adapted to the needs of different industrial control systems (ICS). An example is the so-called Lightweight Machine to Machine (LwM2M) communications protocol stack, which allows mapping virtually any sensor instrumentation into a standardized common description format. LwM2M includes the Constrained Application Protocol (CoAP), which is an interoperable simplified version of Hypertext Transfer Protocol (HTTP) for IoT devices, as well as the standard called Object Security for Constrained RESTful Environments (OSCORE), that standardizes the application-layer protection of CoAP. OSCORE provides end-to-end protection in communications with CoAP or CoAP-mappable HTTP clients and HTTP servers, incorporating another important access control mechanism that works together with the existing standard Datagram Transport Layer Security (DTLS) protocol.

c. Need for Statistical Safeguards

Safeguards are critical for open data development. In the last decade, the push for open data has inspired methods that use statistical safeguards to protect PII and address the implications of sharing the data on the privacy and security of the owners of the data. An example is the highly successful use of Differential Privacy for the U.S. 2020 Census. For instance, there are many reasons why data related to customer consumption and the power grid are not shared. Sometimes usage patterns uniquely identify individuals and their activities inside buildings,

sometimes data can reveal weaknesses in the grid that could help an attacker manipulate grid operation, and data may also be considered proprietary. At the same time, grid telemetry is extremely valuable for understanding grid operation, both for system operation and research purposes, including stability, optimization, planning, security, and more. Other data, like solar photovoltaic (PV) adoption, allow reidentification through satellite imaging of the residence where they are installed. Information about electric vehicle charging, including details on EVs and their participation in DR programs, as well as data on charging locations, can be exploited to launch cyber-attacks on the communication between and the control of electric vehicle charging infrastructure. Similar issues are associated with disclosing distribution systems information. The way the industry approaches these issues to facilitate coordinated decision-making requires statistical safeguards to be appropriately applied.

d. Cloud Infrastructure

Cloud Infrastructure is poised to play an increasing role in the collection and sharing of energy data. Outside the electric utility, sector clouds are becoming nearly ubiquitous, and they are considered the prevalent solution for data-intensive applications; inevitably, this trend is impacting discussion about privacy standards, as well as their enforcement in the utility sector. However, within the electric utility sector, and within the NERC CIP standardization efforts, it is not currently clear how to securely use cloud infrastructures and virtualization, and how utilities should work in a shared security model with the cloud providers. This lack of clarity impacts the ability of utilities to make progress on leveraging cloud computing capacity across all parts of their organization. One positive development for the industry to monitor, particularly at the distribution level, is the emergence of IIoT protocols that aim to establish end-to-end security measures. This could prove beneficial for cloud-based services used in ICS applications.

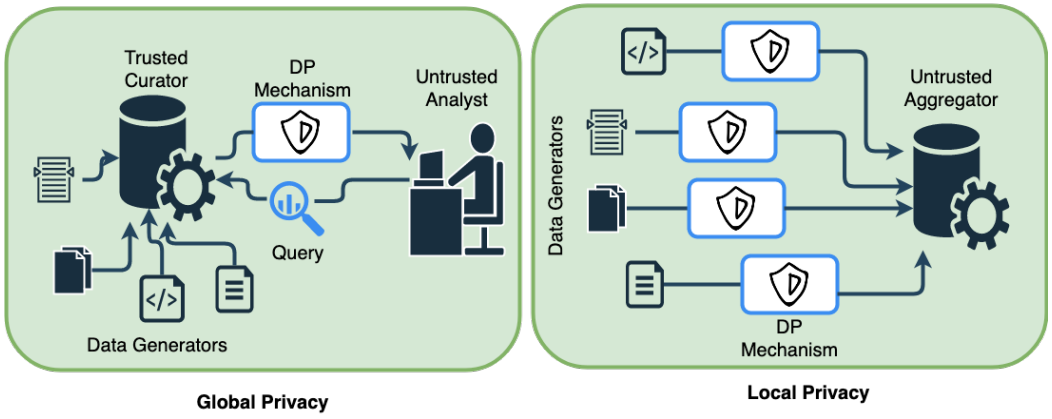

Figure 2: The prevalent vision of the grid is that of a Trusted Curator (a Utility or a contractor) that collects accurate data and controls access from third parties (left). Another possible scenario includes many data generators that want to protect their data (customers, different operators, etc.), and limit the information access by third parties' data aggregators (right). In both cases Differential Privacy (DP)

Mechanisms allow the publishing of information about data queries, blurring the actual query answers through the addition of random noise in a controlled manner to conceal the data in individual records that are queried.

e. Data Protection and Data Sharing Trade-Off

When considering data sharing with statistical protection there are two main configurations. In Figure 2 on the right, we describe the situation where customer data are protected before being aggregated by a server, which would provide the most protection to the consumer. However, the prevalent vision of the grid is that of a ***trusted curator*** as seen in Figure 2 (left), the electric utility itself, or a contractor. Utilities inherently silo data, making sharing very difficult, out of concern that sharing certain data can implicitly violate the privacy of customers. In addition, the default operational security paradigm for the past two decades has been to protect utility grids for national security reasons. While regulators are requiring utilities to share more data, fear of sharing data still exists for extremely relevant reasons, like privacy. As a result, there is a push and a pull for and against sharing that leaves utilities stuck in the middle.

This tension puts utilities in a quandary — there is great value in making data available to partners, peers, vendors, federal and state governments, and even researchers, but also risks in doing so. What are those risks? What degree of information sharing is acceptable? What controls on sharing are considered most acceptable to the various stakeholders, including legal, technical, and statistical controls? What controls are acceptable to the users of the data given that technical controls and anonymization approaches can make data essentially useless, as well? Next, we aim to capture our findings about requirements, best practices, and forward-looking recommendations to serve as useful source material for stakeholders and future standards development.

3. Why is confidentiality a concern?

As discussed in the previous sections, there are major concerns in sharing data. Many of those concerns are enforced by regulators or state bodies. In general, these findings stem from fears about how grid data could be misused as it changes hands. An example is the physical attacks on grid equipment such as the Metcalf substation incident or cyber-attacks. As a result, regulations generally indicate that such data should not be made public. More broadly, individual privacy sentiments across many sectors have increased in recent years, as have specific laws and regulations regarding individual privacy. Notable examples include the General Data Protection Regulations (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States, among others.

However, a lot of electric grid data is public. Anyone can determine the topology of a distribution grid by driving around and following electrical lines or by viewing satellite (e.g., Google Earth) or ground (e.g., Google Street View) imagery. One might argue that forcing an attacker to drive around neighborhoods to follow

transmission lines or sift through online imagery might raise the bar sufficiently to protect the grid. Some argue that electrical lines should be removed from Google Earth and Street View, but this overlooks the fact that information may still be accessible through alternative sources like Bing. Additionally, as demonstrated by sites like WikiLeaks, once data is made public, it is difficult to control its spread and it may remain accessible even if it is removed from specific platforms. Attempting to make public information private may well be akin to the Streisand effect and only amplify the degree to which that information spreads.

We would argue that once the proverbial cat is out of the bag, it is not going to be put back inside. Furthermore, pretending that grid data is not public is not simply "security through obscurity" but self-defeating. A motivated adversary would never be deterred by alternative means of finding such information with such a low barrier to the acquisition, and in the meantime, vital resources are being wasted on protecting data that does not need to be protected, and important activities for which the data is necessary cannot be accomplished. Having access to the relevant data is especially useful when it comes to the coordinated planning and preparation required to massively increase the scale of inverter-based distributed energy resources connected to the grid.

4. Current Practices

As a result of their efforts to hide data, many utilities are left with two paths. One is that utilities are periodically forced to hand over large datasets, for example, because of regulatory audits or a state-level initiative. In such circumstances, ironically, this attempt to protect and not share data ends up exposing even more of the raw data itself. While sharing the data with regulators may be considered "safe" from a national security perspective, it may only exacerbate the privacy problem (e.g., it might contain PII) and may also leave utilities more exposed to regulatory penalties and scrutiny. Further, sharing this data implies trust in the recipient. However, trust of *intent* is one thing, and trust of *competence* — or at least greater competence than all possible adversaries — is another. The failure of organizations, including the U.S. Office of Personnel Management, the Central Intelligence Agency, and the National Security Agency, to protect classified information demonstrates that trust in competence is inadequate. Even with the vast security protections taken by these organizations, sharing data still requires implicitly trusting any stakeholders with access, particularly the system administrators and anyone with physical access to the system containing the sensitive data. Even with all the legal contracts one could wish for, such implicit trust requirements increase the risk to and liability of an institution for accepting responsibility for hosting data, or conversely the risk to the owners and stakeholders who are interested in seeing that data remain confidential.

The other approach that is emerging as a best practice for sharing grid data, is the so-called "15/15 Rule." This rule states that any aggregation of customer data is considered anonymous if it contains at least fifteen customers and if no single customer's data comprises 15% or more of the total values in the aggregated answer. However, the 15/15 rule has been shown to offer *no analytical privacy guarantee*. For instance, an adversary could strategically execute common aggregate queries (like calculating the average power load in a feeder) multiple times and apply simple algebraic manipulations to deduce the existence or absence of individual data records and the specific information they contain.

Most traditional ***anonymization*** or ***sanitization*** approaches generally work very much like the "15/15 Rule" in that they mask certain fields in a set of records and/or aggregate data in a way that seeks to find privacy in the safety of numbers (15 in the case of the 15/15 Rule). Techniques that do this include k-anonymity and several related variations. However, all such techniques have repeatedly been shown to fail to preserve privacy by suffering from "linkage attacks" in which even supposedly anonymized records in the database are linked with external sources of information that can expose sensitive details. Notorious examples of this include identifying Massachusetts Governor William Weld in the Personal Genome Project data, and the de-anonymization of portions of the Netflix Prize dataset by linking the private data with publicly available Internet Movie Database data.

There is a common misconception that sharing synthetic data, which mimics the trends and patterns of actual data, would protect individual privacy. If such synthetic data reproduce the ensemble averages of the data they are emulating, then clearly it is equivalent to sharing such averages which, as we said, is not how one can truly safeguard privacy.

5. Potential Solutions

a. Share the Right Data

The solution that "*more data sharing is better*" is not advisable because it often is not. It is unnecessary to share a lot of data as it does not materially benefit potential use cases, and the privacy and confidentiality risks outweigh the benefits of sharing. Moreover, any shared data must also be well-curated, and this again speaks to sharing the *right* data, not just *more* data. There is a set of queries that provides the necessary information for stakeholders to optimize their regulation or business objectives. Sharing the right data is important to optimize the mechanisms for opening data while using statistical safeguards.

b. Transparency of Data Collection and Use

One of the core tenets of regulations like GDPR and CCPA mentioned earlier is not just that sharing private information about individuals should be limited but that when it is shared, there should be *transparency* about *what is shared*, *with whom it*

is shared, and *how it will be used*, *stored*, and *eventually deleted* once shared. Data do have to be shared more broadly to support grid planning and to meet decarbonization targets. However, perhaps more could be done to promote more awareness for the end consumer on what data is linked to them and what the purpose is for sharing that data or performing analyses on it. Transparency would also allow the research community to vet methods and practices. In California, the California Public Utilities Commission (CPUC) and the California investor-owned utilities (IOUs) both acknowledge that these practices should be conveyed in an "understandable language." However, the current criteria are very open-ended. In contrast, companies like Google and Apple are now including "privacy nutrition labels" in their app stores. These labels used by Google and Apple, based on earlier work by Lorrie Cranor at Carnegie Mellon University, provide an easy-to-read, standard information template on what data is being collected and the purpose of collecting that data. The electric utility industry could consider a similar approach that is tailored to the type of information and analysis of energy systems.

c. Multiparty Computation

Secure multiparty computation and homomorphic encryption are techniques for computing over encrypted data. Unlike approaches like network encryption and full disk encryption that protect data in transit and at rest, respectively, these techniques protect while in use, and as a result, data never need to be decrypted at all. Both techniques have made significant strides over the past ten years and have also been applied to securing and ensuring the privacy of underlying data in analysis processes including those used in the financial sector and government policy. However, such techniques generally remain substantially - sometimes orders of magnitude - slower than cleartext computation. They can also require custom code modification and re-compilation of data analysis code. Therefore, both performance and usability challenges can be significant. Thus, at least for the foreseeable future, such approaches don't appear to represent a primary solution for large-scale data analysis and machine learning. As a result, while software-based encryption techniques can be useful, they seem unlikely to represent the path forward for securing data analysis soon. However, secure multiparty computation can be accomplished using *hardware trusted execution environments (TEEs)*, which can be performant solutions for modern data-driven computing such as machine learning and graph analysis.

TEEs are portions of certain modern microprocessors that enforce strong separation from other processes on the CPU, and some can even encrypt memory and computation. Common commercial TEEs today include ARM's Confidential Compute Architecture, Intel's Secure Guard Extensions (SGX), and AMD's Secure Encrypted Virtualization (SEV). TEEs can be used to improve security over traditional enclaves at minimal cost to performance in comparison to computing over plaintext. TEEs can isolate computation, preventing even system administrators of the machine in which the computation is running from observing the computation or data being

used or generated in the computation. The broad interest in leveraging TEEs to protect data is emphasized by the creation of the Linux Foundation's Confidential Computing Consortium, and the fact that all three major commercial cloud providers — Amazon Web Services, Google Cloud Platform, and Microsoft Azure — all have some sort of TEE functionality.

Another important point is that these methods are still based on controlling access and therefore rely on trusting the recipient of the information. As we said, establishing such trust is not an easy task.

d. Differential Privacy

Differential privacy is a statistical technique for protecting the privacy of algorithms, such as statistical database queries that allow opening the database to third-party queries. The basic idea is to release a *randomized* answer to database queries, using a mechanism to generate the random answer that is designed to guarantee that sensitive attributes of the data used in computing the query are hard to guess from the answer. Its most common mechanism works by adding pseudorandom "noise" to the results of query outputs to hide the presence of any individual record in the database being queried, no matter how targeted the query might be to reveal such information. While the attribute to be hidden is primarily the presence of a specific data record in the subset, the idea can be generalized to hide a particular class attribute. Given its statistical rigor, differential privacy was used by the 2020 U.S. Census and is also used by Apple, Google, Microsoft, and others for gathering sensitive information while protecting the privacy of the individuals or other elements contained in the database records. More specifically, as illustrated in Figure 3, consider two datasets identical in every sense except for one record – changed or missing – and an aggregate query such as average, standard deviation, histogram, maximum, etc. As shown in Figure 3, DP mechanisms guarantee that the corresponding randomized answers are nearly statistically indistinguishable with respect to individual records. Examples of DP mechanisms include the Laplacian and Gaussian mechanisms which involve treating the true query answer with an appropriately tuned noise drawn from the Laplacian and the Gaussian distributions, respectively.

Since the analysis of the data often requires a session that includes many queries it is important to ensure that privacy guarantees are offered for the ensemble of the queries in the session. Because the noise added in each query is independent, the joint privacy leakage is the sum of the privacy leakages in each query. Thus, to have a certain level of DP guaranteed throughout the session, the analyst shall allocate to each query a fraction of the total available budget.

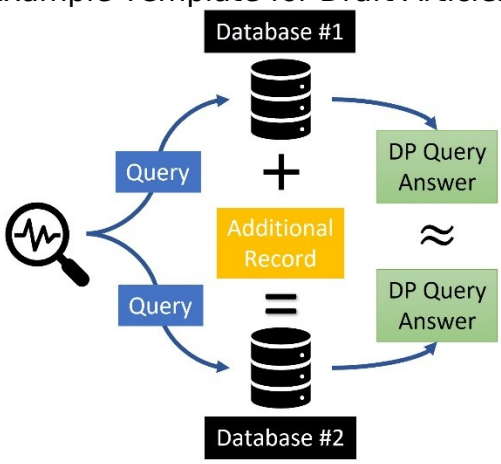

Figure 3 Differential Privacy schematic.

Hence, the statistical guarantee is based on a specific budget the analyst is endowed with, which expires well before all the queries combined would allow the analyst to accurately infer what should be hidden. It is a fundamental result of statistics that, eventually, with enough measurements, the additive noise embedded in them can be overcome and inferences become increasingly reliable. Containing the queries within a budget prevents the analyst from eventually guessing the information that needs to be kept private.

Differential privacy is thus a statistically sound, technical solution to mitigate privacy leakage while still enabling useful information sharing. Our results have demonstrated the effectiveness of leveraging differential privacy with advanced metering infrastructure (AMI) load time series data to generate differentially private synthetic load data that is consistent with the original (labeled) data while preserving privacy. At the same time, it was unclear what it would take to adopt a new technical approach to ensure the privacy of grid data. Would such a solution require regulatory approval? Or could a lengthy approval process be avoided because the differentially private output is already considered sufficiently "de-identified"? If the regulation needs to change, it is also unclear what demonstration would be sufficient to change that regulation. Note also that these statistical techniques *distort* the data, and one needs to be mindful in a safety critical infrastructure about the implication of such errors on the decision-making process that the information is supposed to aid. Striking the best trade-off between accuracy and privacy is something that cannot be left as an afterthought and should be codified in standards.

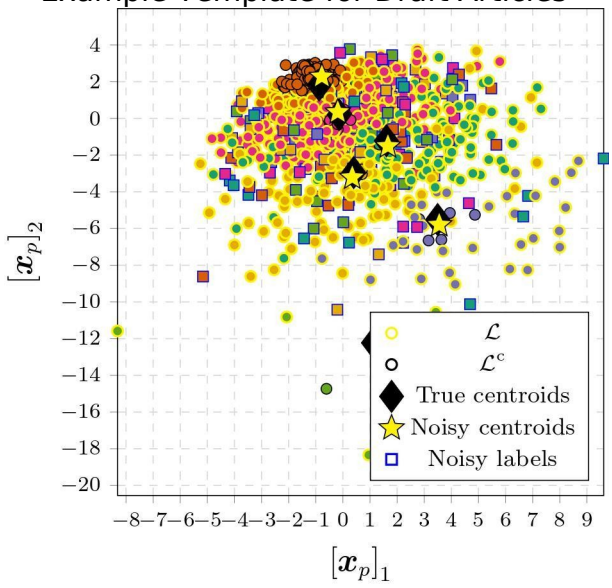

Figure 4 Scatter plot of the AMI Dataset with 1409 houses. Each house's daily load shape (of size 24 - one measurement per hour) is embedded in a 2D space using Multidimensional Scaling for illustration purposes. The true and the DP cluster centroids are shown using black diamond and yellow star markers, respectively. Of the 1409 data points, only a subset (L, shown with yellow borders) is chosen to potentially receive noisy labels. Among those, only points shown using square markers received noisy labels.

Evidence exists that, if applied strategically, DP is a very promising approach. Specifically, we observed that it is possible to communicate clustering results on load data and release the data centroids and labels in a differentially private manner, releasing the results with privacy guarantees and with minimum error. In Figure 4, we show the results of differentially private clustering (into 6 clusters) of daily load shapes belonging to 1409 consumers from 12 distribution circuits across California, USA. In this technique, we add optimal noise to all six centroids (true and the DP centroids are indicated using black diamonds and yellow stars, respectively) and the labels of a subset of houses (indicated with square markers). Clustering can be a first step to devising a differentially private methodology to generate synthetic traces. As it turns out, traces in each cluster fit well a multi-dimensional log-normal distribution, see Figure 5.

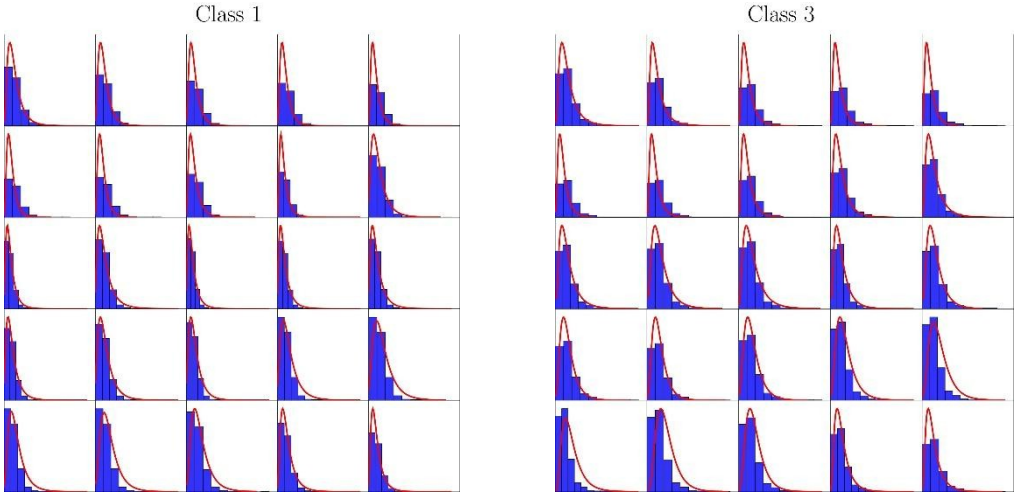

Figure 5: Histograms of the load profiles at each time interval, grouped by cluster. The red curves show the best-fit log normal.

Since in a logarithmic scale, such data are Gaussian, one can generate the synthetic random time series by randomizing the mean and covariance parameters of the distribution to enforce the desired DP guarantees on these statistical quantities, which in turn guarantees that the synthetic data are themselves differentially private. The results of this process are showcased in Figure 6, where we show (in gray) 15 synthetically generated load shape time series for each cluster. In the figure, the areas shaded in brown and green represent the patterns of real and artificially generated data, respectively. These shaded regions show a range of confidence in the data. When we compare the two shaded regions, we notice that they overlap to a large extent, meaning that the real and artificially generated data patterns are very similar to each other.

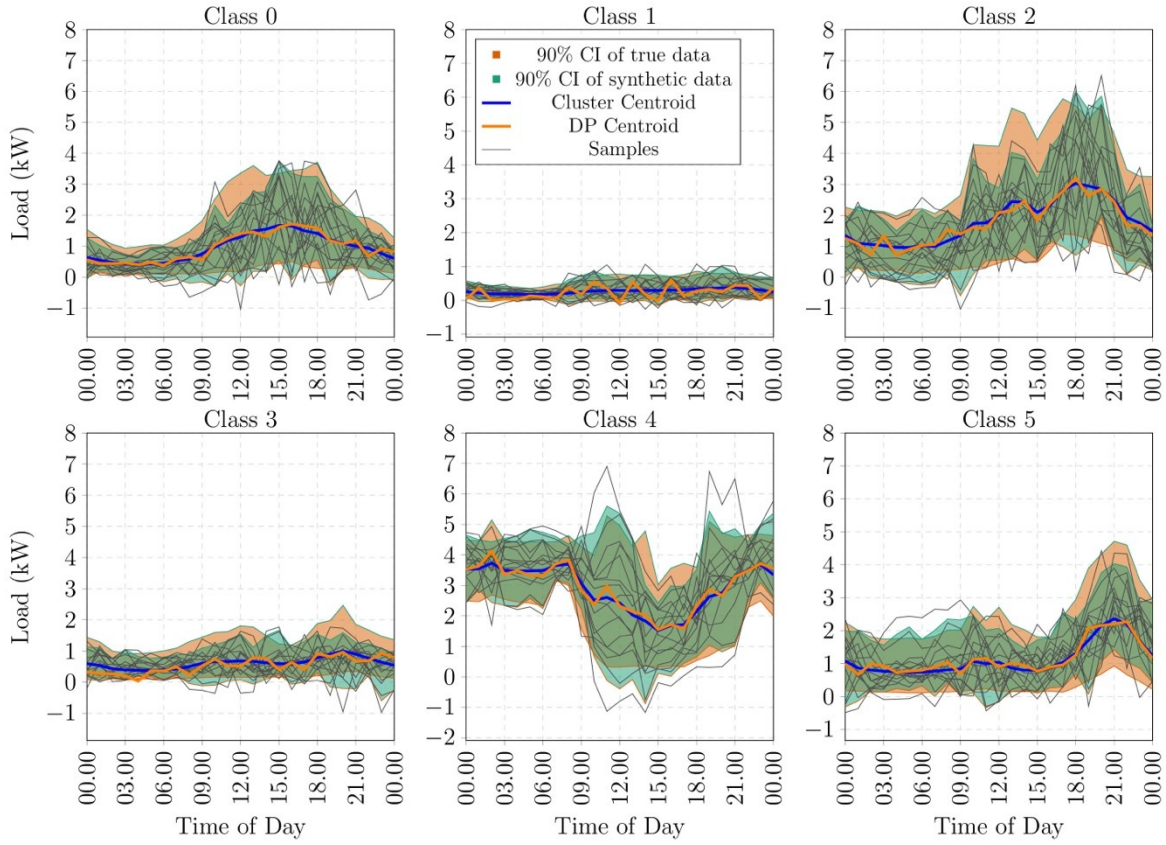

Figure 6: Differentially Private Synthetic Data Generation: Each cluster's 90% confidence interval is shown in the shaded brown region. The true and the DP centroids are shown in blue and orange curves, respectively. Using the DP centroids, we generate synthetic load shapes (shown in gray lines) for consumers in each cluster. The shaded green region shows the 90% confidence interval of the synthetically generated load shapes.

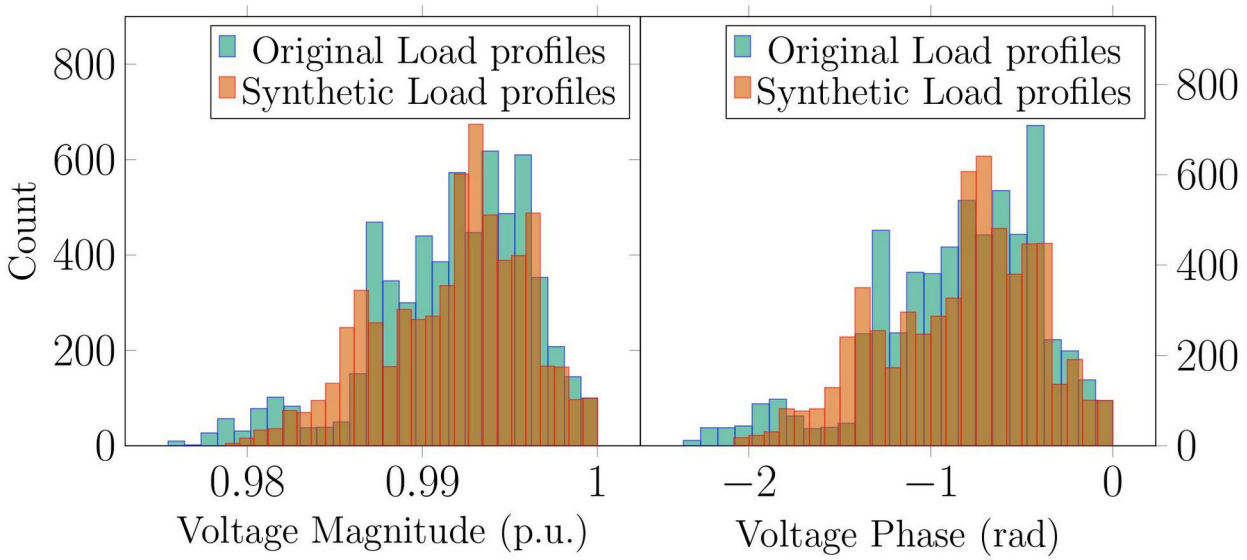

Figure 7: Histogram of the voltage magnitude and phase under true and synthetic load profiles for the IEEE 123-bus case.

One may wonder if such synthetic data is still useful for further analysis of the system. These synthetically generated load shapes were tested on two standard (the MATPOWER 141 and a modified balanced IEEE-123) distribution system test cases. Load shapes from households across all six clusters and their synthetically generated load shapes were used as load inputs of an Optimal Power Flow problem to obtain voltage magnitude and phase at each bus. The histograms of voltage magnitude and phase obtained under both cases with true and synthetic load profiles in Figure 7 and Figure 8 showcase that the results obtained for the synthetic loads provide a good match for those obtained for the true loads.

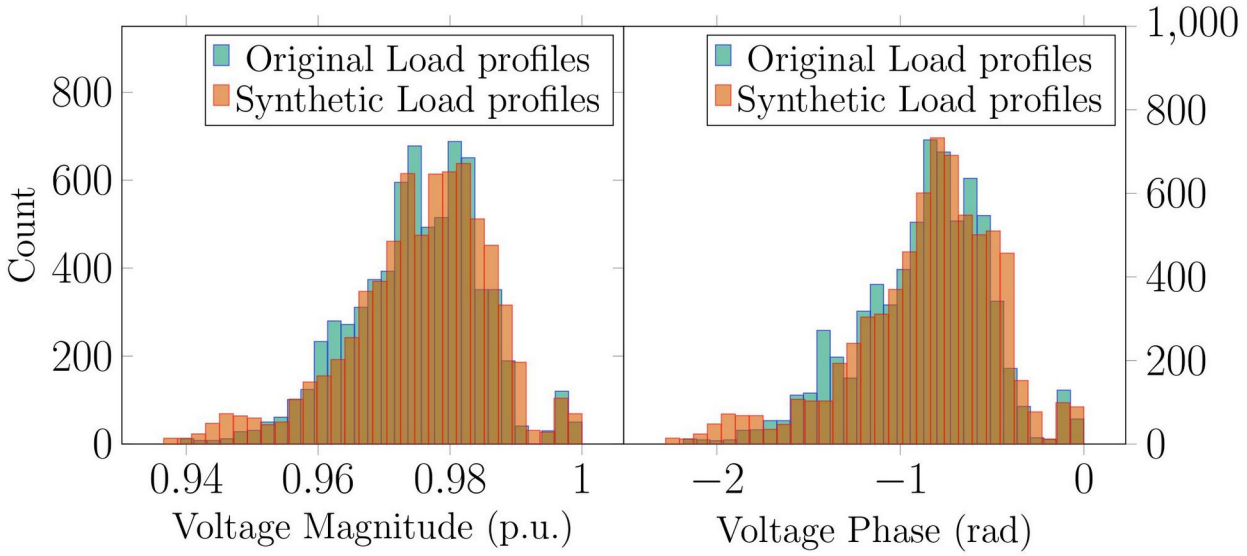

Figure 8: Histogram of the voltage magnitude and phase under true and synthetic load profiles for the MATPOWER 141-bus case.

While these results are promising, the pathway to obtaining accuracy appears to be inextricable from tailoring the mechanisms to the queries. To achieve greater utility of the shared noisy query answers, standards may have to go deeper in defining what and how one can share data or queries about them with DP guarantees. Unfortunately, the naïve approach of releasing data directly after adding noise is untenable because the noise that needs to be added to preserve privacy, without any form of query aggregation, is such that future analyses would be extremely inaccurate and, therefore, mostly misleading. This point relates to the one made previously on “sharing the right data” or, more precisely, sharing the right information about the data and designing statistical methods that are not only guaranteed to preserve privacy but also are designed to give the best accuracy possible, given the privacy constraints.

e. Putting Privacy-Preserving Techniques Together

Differential privacy can enable data sharing without exposing raw data, thus potentially enabling any untrusted user access to that data without risk. TEEs protect against untrusted computing providers and can be made to perform secure multiparty computation. Each approach has its benefits and can be deployed separately based on the risk model and can also be put together to provide more comprehensive guarantees about securing data from both platform providers and end users of the data in question. Although both are in production use in commercial industry, government, or both, neither differential privacy nor TEEs have stopped evolving and both will continue to become even more usable and performant. For energy data, it is important to incorporate domain expertise to clarify the type of analytical results that are most beneficial to different stakeholders and then determine the mechanisms that strike the best compromise between privacy and accuracy of the data queries.

In conclusion, the next steps include working with standards organizations that have the scope to address a critical mass of solar / inverter / distributed energy industry stakeholders; regulators; end-users of grid data, including grid planning and research; and technical experts in privacy-preserving methods such as the ones that we have discussed here.

For Further Reading

- Ravi, N., Scaglione, A., Kadam, S., Gentz, R., Peisert, S., Lunghino, B., E. Levijarvi & Shumavon, A. (2022). Differentially Private-Means Clustering Applied to Meter Data Analysis and Synthesis. *IEEE Transactions on Smart Grid*, *13*(6), 4801-4814.
- Cynthia Dwork. Differential Privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP)*, volume 4052 of Lecture Notes in Computer Science, July 2006. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/dwork.pdf
- U.S. Census Bureau, “Statistical Safeguards.” https://www.census.gov/about/policies/privacy/statistical_safeguards.html

Acknowledgments

This research was supported in part by the Director of the Office of Energy Efficiency and Renewable Energy, Solar Energy Technologies Office, and the Director of the Office of Cybersecurity, Energy Security, and Emergency Response, of the U.S. Department of Energy.

Biographies

Robert Currie and Aram Shumavon are with Kevala, Inc., San Francisco, California, USA 94133.

Sean Peisert is with Lawrence Berkeley National Laboratory, Berkeley, California, USA 94720.

Anna Scaglione and Nikhil Ravi are with Cornell University, New York, NY, USA 10044.