7-18-2022

A Privacy-Preserving Strategy for the Trust Layer of the Energy Grid of Things Distributed Energy Resource Management System

Mohammed Abdullah Alsaid
Portland State University

A Privacy-Preserving Strategy for the Trust Layer of the Energy Grid of Things Distributed Energy Resource Management System.

by

Mohammed Abdullah Alsaid

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Electrical and Computer Engineering

Thesis Committee:
Nirupama Bulusu, Chair
Banafsheh Rekabdar
John M. Acken
Robert B. Bass

Portland State University
2022

Abstract

Emergent from the shadows of the traditional grid flaws, the Smart Grid (SG) idea was born and led by government mandates toward cleaner energy production. The SG represents the next generation of electricity distribution systems that subsume recent technological innovations. It uses digital communication between its components and entities to attain more automation, self-sufficiency, and reliability. Unfortunately, this relatively new concept is not flawless; the intrinsic reliance on increased digital communication spreads open attack paths for adversaries. Therefore, finding solutions that address information exchange vulnerabilities has become imperative.

The Energy Grid of Things (EGoT) is Portland State University's (PSU's) implementation of a Distributed Energy Resource Management System (DERMS). The EGoT DERMS requires access to customers' information to achieve operational objectives. The system's access to customers' information needs to be restricted such that it does not violate customers' privacy. Applying privacy protection models such as K-anonymity to EGoT DERMS sub-components safeguards that privacy.

This thesis work proposes a strategy to ensure communication in the EGoT DERMS is privacy-preserving and secure. Specifically, it provides an approach to applying the

Mondrian Algorithm to ensure data within the system excludes Personally Identifiable Information (PII) and provides means for securing the communication according to industry standards (IEEE 2030.5). Results suggest that the generalization hierarchy derived for the EGoT DERMS exhibits an Identical Generalization Hierarchy structure. Guarantees of sameness manifested in the test feeder topology would not hold in real-world scenarios.

Dedication

This work is dedicated to my family for their unwavering support through a tumultuous journey. To those who witnessed the start of the journey but not the end, may they rest in peace.

Acknowledgements

Achievements mean little when they cannot be shared with others. I want to extend my thanks to those individuals whose support made this dream a reality. First, I would like to thank my ever-so-patient advisors. To Dr. Robert B. Bass, you have my deepest thanks. Your guidance and support were vital to my success. To Dr. Nirupama Bulusu, thank you for your support and time. Your lessons were invaluable and extremely helpful. To Dr. John M. Acken, your wisdom and assistance made me a better scholar. Your honesty and constructive criticism have always motivated me.

Secondly, I would like to thank my committee members, Dr. Nirupama Bulusu, Dr. Banafsheh Rekabdar, Dr. John M. Acken, and Dr. Robert B. Bass. Not only for your time and patience but rather for the intellectual insights you brought that helped in my development as a student and a researcher.

I want to also extend my thanks to my colleagues that aided me through my journey. I want to thank my Midrar Adham for his invaluable insight and support. Not many people are blessed enough to have friends as colleagues. Also, I would like to express my thanks to Abdullah Barghouti, Sonali Fernando, Shahad Alomani, and Tylor Slay. Your aid has made a challenging task more manageable.

Finally, none of this would have been possible or even mattered without my parents and siblings' love and support. You have my heartfelt gratitude, and I am forever indebted to you.

Contents

List of Tables

List of Figures

Acronyms

ACL Access Control List

API Application Programming Interface

CA21 California Rule 21

CDTA Central Distributed Trust Aggregator

CSIP Common Smart Inverter Profile

DCM Distributed Control Module

DER Distributed Energy Resource

DERMS Distributed Energy Resource Management System

DLC Direct Load Control

DR Demand Response

DSM Demand-Side Management

DTM System Distributed Trust Model System

DTMC Distributed Trust Model Client

EGoT Energy Grid of Things

ESI Energy Service Interface

FDI False Data Injection

GO Grid Operator

GSP Grid Service Provider

IGH Identical Generalization Hierarchy

kV kilovolts

LAN Local Area Network

MVoT Metric Vector of Trust

NCP Normalized Certainty Penalty

PII Personally Identifiable Information

PSU Portland State University

SG Smart Grid

SOLC Service-Oriented Load Control

SPC Service Provisioning Customer

SVM Support Vector Machine

TMDG Trust Model Data Generator

TMS Trust Model Simulator

WAN Wide Area Network

1 Introduction

The traditional concept of power distribution is becoming outdated, predominantly in the sense that it has not kept pace with recent technological advancements. Arguably, it is the most complex system ever created. Nevertheless, this comes with more disadvantages than 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 [1][2].

Emergent from the shadows of the traditional grid flaws, the Smart Grid (SG) idea was born and led by government mandates toward cleaner energy production. The SG represents the next generation of electricity distribution systems that subsume recent technological innovations. It uses digital communication between its components and entities to attain more automation, self-sufficiency, and reliability. Unfortunately, this relatively new concept is not flawless; the intrinsic reliance on increased digital communication spreads open attack paths for adversaries.

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

1.1 Problem Statement

The success of any SG hinges on its customers. For instance, an increase in Distributed Energy Resource (DER) participation within a system boosts its ability to counterbalance disruptive events. That is attributed to grid operators having greater control over the demand side of the grid. Some studies have examined the importance of adding incentives to encourage Demand Response (DR) program participation [3, 4]. Similarly, one must ensure there are little to no discouraging factors that affect customers' participation. One very significant barrier to customers' participation is the prospect of violating their privacy. Any DR program depends on a large amount of information exchange between its components. With that in mind, the problem this work attempts to address is how to preserve customers' privacy in EGoT DERMS without compromising security or operational objectives.

1.2 Objectives of Work

This work encompasses developing an information exchange interface for conveying trust information within a Distributed Trust Model System (DTM System). The interface shall be designed to preserve privacy and provide security. In the DTM System, local Distributed Trust Model Clients (DTMCs) transmit their local trust information to a Central Distributed Trust Aggregator (CDTA) for further processing. This information exchange must be secure and protective of customers' privacy; otherwise, participation in SG customer programs would diminish, affecting the system's self-sufficiency attribute.

The EGoT DERMS is an implementation of a SG that focuses on security according

to industry standards and the protection of customers' private information. For security, preventive measures are provided by the IEEE 2030.5 protocol, while detective measures utilize a trust layer.

This thesis work has two major objectives. The first objective is to augment the system's trust layer with privacy features by employing anonymization algorithms. The second objective is to design an interface whereby various components can transmit this anonymized information.

2 Background

2.1 EGoT DERMS

The idea of the Smart Grid is not entirely novel. Consequently, there have been many implementations with many novel ideas. Additionally, some countries have adopted some of the management strategies derived from the concept [5]. Moreover, many manufacturers have started to integrate protocols that aid the control of electrical resources to be easily used as DER [3].

Demand-Side Management (DSM) is a broad term that addresses utilities' monitoring, planning, and implementation of activities to influence customers' electricity usage in a way that affects the utilities' load shape. The operational objectives of these utilities' activities typically include peak shedding, load shifting, and several others [6]. Direct Load Control (DLC) is the traditional approach for executing DSM activities. In the DLC approach, the general theme revolves around utilities taking responsibility for directly managing customers' DERs who enroll in DR programs [7]. If customers desire to opt out, they need to reach out to the utility directly. Alternatively, Service-Oriented Load Control (SOLC) is a modern approach to DSM. Contrary to the DLC approach, SOLC confers the responsibility of DER management to the DER owner. The primary benefit is that once the customer participates,

the owners retain complete control over their DERs, which means they can participate or opt out of any program whenever they choose to.

The EGoT DERMS is PSU's implementation of the Smart Grid. It employs SOLC methods for DSM. It 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. The protocol allows for the maximum degree of flexibility that can be leveraged to accommodate the largest number of off-the-shelf products [3]. Figure 2.1 presents a conceptual view of the overall structure of EGoT DERMS.

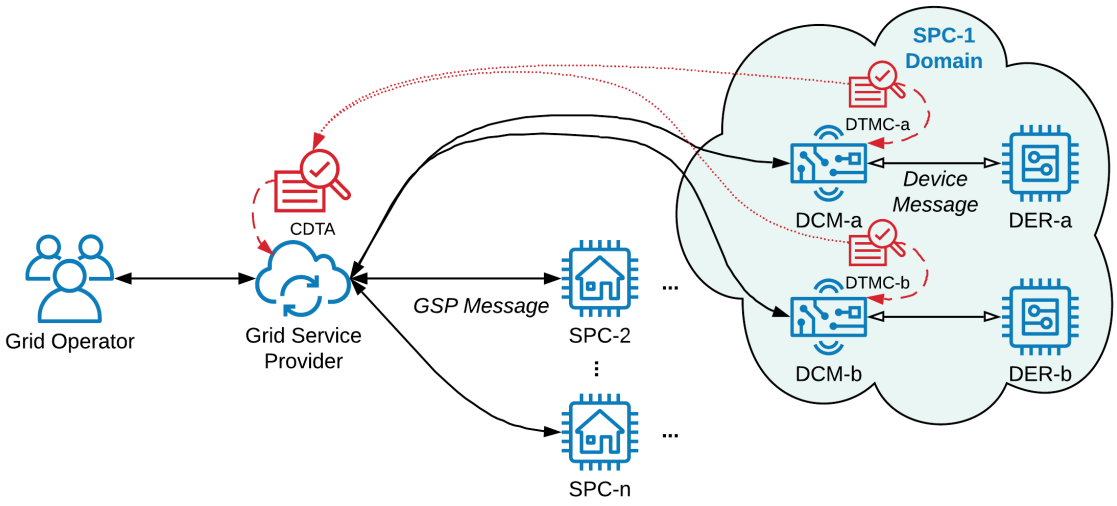

Figure 2.1: An overview of the system architecture. The trust layer (shown in red) comprises the DTM System.

One can conceptually divide EGoT DERMS into two different segments: aggregation and dispatch. Moreover, one can further divide the dispatch segment into the homeowner and utility sides. Customers who own smart appliances are called Service Provisioning Customers (SPCs) whereas the aggregators are called Grid Service Providers (GSPs). Ideally,

each GSP can dispatch a large number of DERs such that the services it provides are impactful [8, 9]. The motivation is that SPCs own controllable appliances, DERs, that can provide various DER services to a GSP. The GSP can use these appliances in large numbers to provide grid services that meet Grid Operators' operation objectives depending on the state of the grid.

Grid Operators (GOs) are entities that manage the grid to achieve operational objectives. The operational objectives can be to either 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. Pay attention that the GSP provides **grid** services to the GOs. It does that by using the offered **DER** services provided by the DERs.

The EGoT DERMS system follows a server-client architecture where the servers are hosted by GSPs. GSPs function as distributed aggregators. GOs can subscribe to the grid services offered by GSPs, which in turn propagate the new service objectives down to the DERs based on their topology and geographical location. Hence, GSPs provide aggregation and dispatch of DERs.

The bi-directional communication between GSP and DERs need to be formally defined. In EGoT DERMS the communication is defined in a rules-based manner that governs the behavior of the parties involved. These governing rules are collectively named the Energy Service Interface (ESI). Such a concept is necessary to ensure operational objectives are

met and to meet accountability requirements [10]. That is, when a request is initiated and accepted, guarantees of service delivery have taken place within the constraint established by the ESI [11].

As mentioned earlier, 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 [12].

2.1.1 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 [13, 14, 15]. This definition is relevant to distributed systems such as the EGoT DERMS where reliability plays a crucial part [16, 17]. 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 DTMCs and one corresponding CDTA. The DTMCs are components placed adjacent to DCMs, as shown in Figure 2.1. 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 measure the local trust of the DCM, DER, and GSP by observing the communication fingerprint of each actor. Finally, DTMCs send their local trust information, which is referred to as a Metric Vector of Trust (MVoT), to the CDTA where the distributed trust is aggregated and an overall trust of the EGoT DERMS is computed.

2.1.2 Smart Energy Profile Application Protocol

IEEE 2030.5 is a protocol that defines an application layer that runs atop the TCP/IP protocol suite [18]. It provides functions for utilities to manage end-user resources to deliver grid services. The provided functions include demand response, load control, and many more. However, due to the rules of the ESI, the EGoT DERMS uses only a small subset of the provided functions. Specifically, it uses the mandatory function sets of the protocol for the server-side, which amounts to eleven function sets. And, the EGoT DERMS uses the flow reservation function set to provide DER services to the GSP.

2.1.3 Common Smart Inverter Profile v2.0

California state initiatives introduced California Rule 21 (CA21), which provides efforts towards accommodating renewable energy. Specifically, CA21 includes requirements for smart inverters. The SunSpec Alliance proposed the Common Smart Inverter Profile (CSIP) v2.0 standard to aid DER manufacturers and aggregators in compliance with CA21 proceedings and IEEE 2030.5 [19]. One of the CSIP objectives is promoting a "plug & play" level of interoperability.

Among the many propositions put forth in the CSIP standard is the topological grouping of DERs. Figure 2.2 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, as shown in Figure 2.3. 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 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 proposed 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 thesis work.

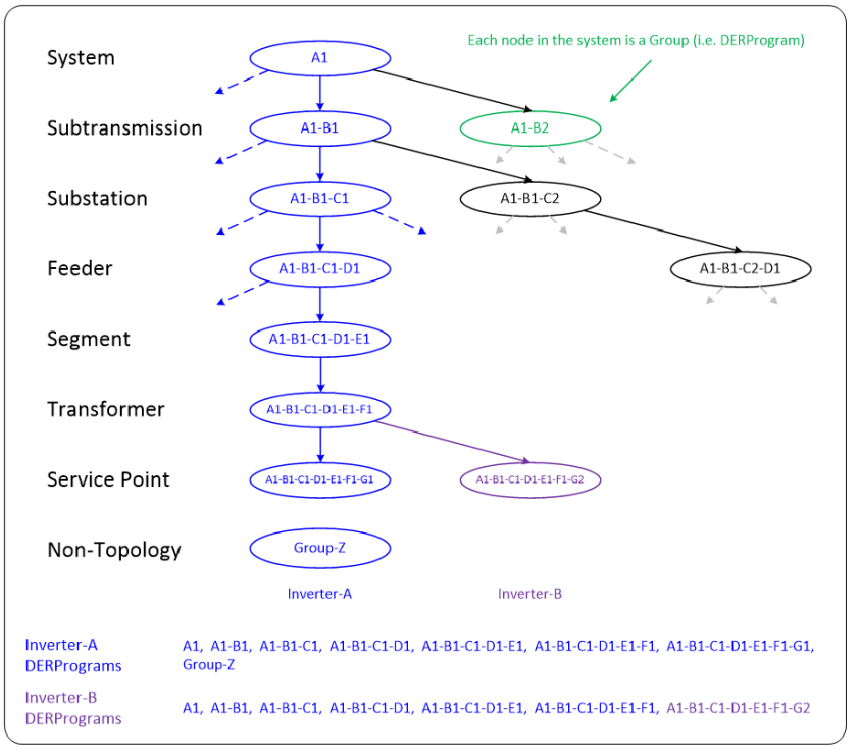

Figure 2.2: Topological grouping as described in CSIP v2.0 [19].

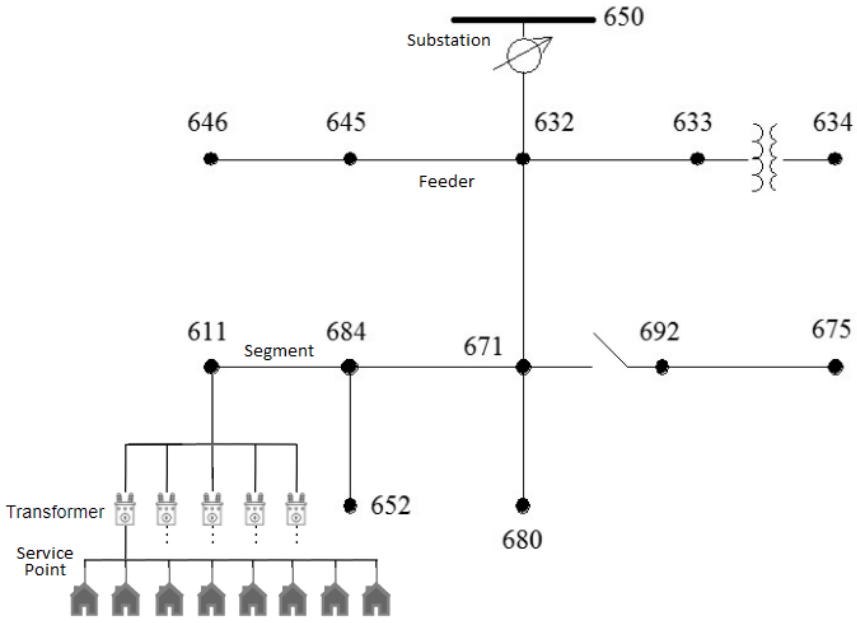

Figure 2.3: The electrical representation of topological groupings shown in Figure 2.2.

As previously mentioned, SPCs can use their DERs to provide various DER services to the GSP. Table 2.1 shows the various DER services, their purpose, and their location extent. Such DER services include energy scheduling, reservation, regulation, emergency, and frequency services. DER decision regarding dispatching DERs depends on the location of the DER and the type of target grid service it tries to meet. It must be noted that both these factors are interlinked. For instance, a frequency response service is used when there is an under-frequency or over-frequency event. The location extent of such an event can cause a catastrophic impact on the entire system. In contrast, the voltage service has a narrower extent. A significant deviation in voltage beyond the permissible limits of $\pm$ 5%, as specified by ANSI C57.96, could impact specific system components, such as some of the feeders or transformers, as opposed to the whole system. Hence, the voltage service limits are local, unlike the frequency response service.

Grid-DER Service	**Purpose**	**Location Extent**
Energy Schedule	Ensure adequate energy resource supply.	System, Subtransmission, Substation, or Feeder
Reserve	Reserve source or load capacity.	System, Subtransmission, Substation, or Feeder
Regulation	Support area control error (ACE).	System
Emergency	Support recovery of a collapsed electrical power system.	Substation, Feeder, or Segment
Voltage Support	Detect and correct voltage excursions outside of defined limits.	Feeder, Segment, or Transformer
Frequency Response	Detect and arrest sudden frequency deviations outside of defined limits.	System

Table 2.1: The six Grid-DER Services, their purposes, and topological location extents.

2.1.4 IEEE 13 Node Test Feeder

The IEEE 13 node test feeder model is a distribution system model with a nominal voltage of 4.1 kilovolts (kV) [20]. In other words, the primary voltage of any given distribution transformer in the model is also 4.1 kV. As the name suggests, it is a testing feeder model used for study and research purposes. Kersting proposed the IEEE 13 node test feeder design as an intentionally overloaded and unbalanced system. It is unbalanced because distribution systems, in reality, are all unbalanced, which adds an authentic component to studies using the model. Also, the unbalanced characteristic means that each node in the system is configured to be either a three-phase, two-phase, or single-phase node.

The IEEE 13 node test feeder model used at PSU can model up to 1000 DERs. These DERs use household demand profiles, water heaters, and EVs. This model is employed in this thesis to generate a representative topology, with which topological IDs are generated.

2.2 K-Anonymity

Many organizations aim at publishing microdata for research purposes (demographic, health, and other domains). However, such microdata may contain Personally Identifiable Information (PII) that breaches the privacy of their customers. For example, combining the published data with publicly available external data sets can pinpoint individuals even though the obvious PII of the microdata 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 [21].

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 [22]. Consequently, it is not easy to find an optimal solution in a reasonable time. 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.

2.2.1 Basic Definitions

Prior to discussing the various proposed models for achieving K-anonymity and its variants, one must define some appropriate terminology.

Quasi-identifiers are attributes in the data that can be used to re-identify records by joining the anonymized data set with external data sets. More formally, they are a set of attributes $\{X_1, X_2, ..., X_n\}$ in a table **T** that describe sensitive information. Deciding on what attributes should be considered as quasi-identifiers relies on domain knowledge.

An **Equivalence Class** constitutes the set of tuples $x_1, x_2, ...x_n$ in the table **T** that share the same value for the attributes $X_1, X_2, ..., X_n$. For example, let table **T** contain attributes X_1, X_2, and X_3 defined as shown below. Notice that the values for a given attribute in the

records are the basis for creating equivalence classes, which are bolded in Table 2.3 for clarity. For instance, the values for the provided example result in two equivalence classes over attribute X_1, which correspond to the bolded values in the first row (**1** and **3**). Since two records have a value of 1 for the attribute, they belong to the same class. The remaining record has a value of 3 for the attribute and should be in another class. The dimensions of the anonymization model dictate how equivalence classes are constructed and are briefly described in a separate subsection.

X_1	X_2	X_3
1	4	3
3	6	6
1	2	6

Table 2.2: An example table **T**.

Attribute used to generate equivalence classes	Resulting equivalence classes
X_1	{(**1**,4,3), (**1**,2,6)}, {(**3**,6,6)}
X_2	{(1,**4**,3)}, {(3,**6**,6)}, {(1,**2**,6)}
X_3	{(1,4,**3**)}, {(3,6,**6**), (1,2,**6**)}

Table 2.3: An example of resulting equivalence classes on the different attributes in Table 2.2

K-Anonymity Property A data set is said to be k-anonymous when every record occurs in the data set at least k times. In other words, the size of each equivalence class in the data set is at least k.

K-Anonymization is a view **V** of the table **T** in which the records are suppressed and/or generalized to satisfy the k-anonymity property with respect to the *quasi-identifiers*.

Attribute Disclosure Is a concern for datasets that contain PII. It is independent of *Identity Disclosure*, where a record is linked to a particular individual. Rather, it expresses the case where an *attribute value* is associated with an individual [23].

Models dimension Concern the privacy model's ability to re-code tuples into anonymized tuples. Some models are single-dimensional where a re-coding function ϕ can only be applied to each attribute separately. Other models are multi-dimensional, which utilize re-coding functions that can be applied to the entire tuple when anonymizing.

As technologies get proposed, they face the test of time, and hopefully, they do so without failing determinately. The K-Anonymity algorithm certainly is no different. It underwent the test of time, and researchers have found better means for achieving privacy. Some models build on K-anonymity, intending to minimize risks inherent to K-anonymity. The following sections briefly discuss some of the proposed models and their driving motivation.

2.2.2 ℓ-diversity

A k-anonymized table is a view of the original data set where each record is indistinguishable from K-1 records for some quasi-identifiers. At first glance, this seems like it gives a good measure of privacy. However, this has two glaring issues demonstrated by Machanavajjhala et al. [24]. The first case where K-anonymity fails is when there is little diversity in the values of sensitive attributes. Little diversity in values allows for attribute disclosure, enabling attackers to exploit homogeneity in values to infer the attribute values of their target. For instance, let a target be within an equivalence class with three other

duplicates and only three sensitive attributes. If there is no diversity between the sensitive attributes, our target's sensitive attributes values are known.

The second flaw of k-anonymity concerns an attacker employing background knowledge to discern their target's record within the data set. In this case, the attacker works backward to narrow down the set of possible records for the target based on one known sensitive attribute. The attacker continues to repeat the process until they zero down to a single record with a high percentage of certainty.

The previously discussed cases illustrate that k-anonymity does not guarantee privacy. Instead, the simplicity of the model allowed for broader adoption and popularity. This motivated researchers to derive a privacy model beyond K-anonymity, where privacy is guaranteed. Machanavajjhala et al. proposed ℓ-diversity, which enforces stronger restrictions on the model. Under ℓ-diversity, the number of different sensitive attribute values must be at least ℓ. This can be achieved according to diversity metrics such as *Distinct ℓ-diversity, Entropy ℓ-diversity, and Recursive (c, l)-diversity*.

2.2.3 t-closeness

Similar to *ℓ-diversity*, t-closeness aims at improving upon previous models. t-closeness attempts to build on K-anonymity and minimize the risks associated with ℓ-diversity. The main privacy risk associated with ℓ-diversity is that attackers can still disclose attribute values under special cases. The first case occurs when sensitive data values are skewed and achieving diversity becomes difficult. The second case occurs when the values for sensitive data are semantically similar such that attackers can instead limit the value to range

as opposed to a particular value. This led Li et al. to propose t-closeness, an extension of ℓ-diversity, where the privacy model accounts for sensitive attribute value distribution within an equivalence class [25].

2.3 K-Anonymity Applied to Smart Grids

Applying K-anonymity to SGs components is not entirely a novel notion in that there are similar works. For instance, Mark Stegelmann and Dogan Kesdogan's approach proposes a privacy-preserving smart metering architecture [26]. 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 EGoT DERMS attempts to address.

Similarly, Yuce et al. studied solutions for consumer data privacy in a district-level microgrid [27]. It obtains 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.

Finally, Donghe Li et al. proposed an approach that focuses on demand response in microgrids using vehicle-to-vehicle technology [28]. The approach attempts to add 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

SGs implementations. This work is not different in that it also explores the K-anonymity application for the EGoT DERMS. Nonetheless, it differs from other works in that it attempts to apply the anonymization technique to the trust layer. None of the discussed works explore anonymization combined with a trust layer concept as defined in the EGoT DERMS.

2.4 Threat Model

For critical infrastructure systems such as the SG, one must enumerate the weaknesses and risks anticipated by the adopted design. Doing so requires the forethought of expected adversaries and attack types allowed by design. Limiting the security posture of a critical system to the mere adoption of recommended standards renders the system vulnerable with little insight into its weaknesses [29].

Drafting threat models for systems serve as a mechanism for systematically finding vulnerabilities. Creating a system threat model necessitates iteratively identifying assets, examining the interactions between the system components, and enumerating threats. The outcome of the process is security requirements that can be the basis of the system security.

To identify assets for a system, one must list all critical resources in the system. Critical resources that require protection can be tangible such as data integrity, or intangible resources like trust in the system. Further, identifying threats requires creating a taxonomy of the assumed adversaries' probable goals. Such taxonomy is beneficial as one can derive security requirements from it. Lastly, note that not every threat has to be eliminated; instead, have its associated risk is managed. In other words, determine whether one should mitigate the risk

or accept it depending on how severe the threat is.

An example of common SG attacks includes the False Data Injection (FDI) attack. In FDI attacks, adversaries inject false data into the sensors to corrupt the integrity of sensors readings [30]. In principle, the adversary wants the difference between the erroneous and actual data to be minute enough to go undetected yet big enough to introduce calculation errors into the state variables.

Anwar, Mahmood, and Shah proposed an approach for cyber-attack detection in a SG [31]. The proposed approach involves using supervised machine learning to detect faults and FDI attacks. Specifically, it entails training a Support Vector Machine (SVM) model on a training dataset containing normal and abnormal behavior. The proposed work does not consider preventive security schemes such as authentication or authorization measures. Such measures are provided through the IEEE 2030.5 protocol in the EGoT DERMS.

2.5 Summary

This section provides background for key concepts pertaining to the thesis work. The EGoT DERMS is an implementation that emphasizes interoperability. This section provides a description of the various actors in the EGoT DERMS and their defined roles. The standards and protocols that make up the backbone of the EGoT DERMS operation, CSIP and IEEE 2030.5, are briefly introduced. The trust model is an additional layer added to the EGoT DERMS system to augment the security mechanisms. Threat modeling is an iterative approach to systematically finding threats and managing risks for critical infrastructures.

K-Anonymity is a model for removing PII from data sets prior to publishing or operating on the data. Other improvements to the K-anonymity model include ℓ-diversity and t-closeness, which address some K-anonymity shortcomings. Applying K-anonymity to SGs is not new. However, this work investigates its application to detective security measures in the EGoT DERMS as means of privacy protection.

3 Design Methodology

3.1 The EGoT DERMS Threat Model

As mentioned in the Threat Model section of the background, it is essential to draft a threat model of the system before developing premature security policies. During the phase of drafting implementation profiles for the EGoT DERMS, an application diagram was drafted as a means to identify possible threats to the EGoT DERMS system actors. Figure 3.1 below contains a diagram of the initial threat model.

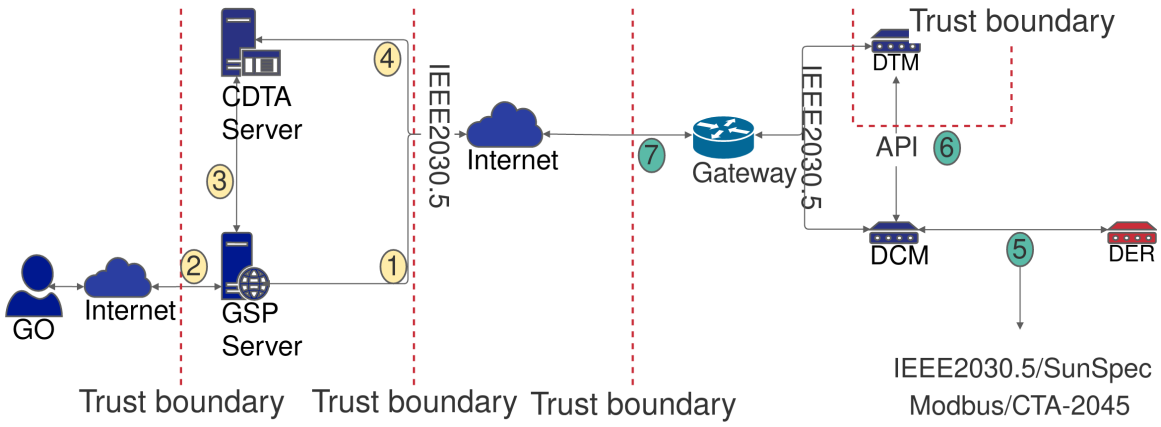

Figure 3.1: An application diagram for the system. It shows the data flow in the system.

As shown in Figure 3.1, communication between actors is routed through the internet. Note that communication between the GO and the GSP is outside the scope of this work.

Moreover, notice that communication between the DCM and DTMC actors is local. Specifically, DCMs and DERs communication is direct and does not go through the router in cases where serial communication is used. Notice that dashed red lines signify trust boundaries where trust levels change. That is, the trust levels in communication within SPC's Local Area Network (LAN) are different from communication that is routed through the internet.

Similarly, trust levels vary between DTMCs and other actors, which is demonstrated by the trust boundary surrounding the DTMC. Finally, numbers in circles highlight data exchange positions (i.e., data crosses a trust boundary). Green circles represent data exchange within a LAN network, whereas yellow circles represent communication that goes through a Wide Area Network (WAN). Note that the communication between the CDTA and GSP is regarded as WAN communication. Both servers can be within the same site, hosted by the same entity/utility but on separate sites, or hosted on a cloud server.

Different categories of adversaries can launch attacks that targets EGoT DERMS system actors. However, given the complexity of such infrastructure, potential high-risk threats come from tech-savvy users and nation-backed adversaries. The category of tech-savvy users describes a group of users with malicious intentions but limited resources to launch devastating attacks. The motivation for the first category might be to conduct further reconnaissance of a specific target or game the system to reap the involved financial incentives without providing their DERs to GSPs. Tech-savvy users' skills might enable them to dissect the protocols to find undiscovered vulnerabilities and exploit them.

Nation-backed adversaries are adversaries with the requisite expertise and resources to

initiate destructive attacks on a grid. Unlike tech-savvy users, nation-backed adversaries have more resources, expertise, and motives to inflict real damage to the grid. The complexity of attacks they can instigate is much higher than the other category. State-level adversaries' goals might be to bring about financial loss, cause blackouts, or other political reasons.

3.2 K-Anonymity Applied to the EGoT DERMS

3.2.1 The Mondrian Algorithm

LeFevre, DeWitt, and Ramakrishnan proposed a multi-dimensional model for k-anonymization and a greedy algorithm for k-anonymization [32]. The 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 works by assigning 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:

$$C_{DM} = \sum_{E \in EquivalenceClasses} |E|^2 \tag{3.1}$$

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

$$C_{avg} = \frac{Number\_of\_records/Number\_of\_equivalence\_classes}{K} \tag{3.2}$$

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 [33]. 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 [34]. NCP can be defined for numerical attributes as follows, where C is the equivalence class, and A is a numerical attribute:

$$NCP_A(C) = \frac{max_A^C - min_A^C}{max_A - min_A} \tag{3.3}$$

Equation 3.3 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:

$$NCP_A(C) = \frac{size(u)}{|A|} \tag{3.4}$$

Where $|A|$ is the number of distinct values of attribute 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 3.4 is the equation used to compute the penalty.

3.2.2 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 2.2. This topological location is used as an identifying value for loads in an electrical and distribution system. As mentioned previously in Section 2.1, only the distribution side of the grid is considered. Hence, this thesis 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 3.2 portrays a generic scheme of the generalization hierarchy used for the EGoT DERMS.

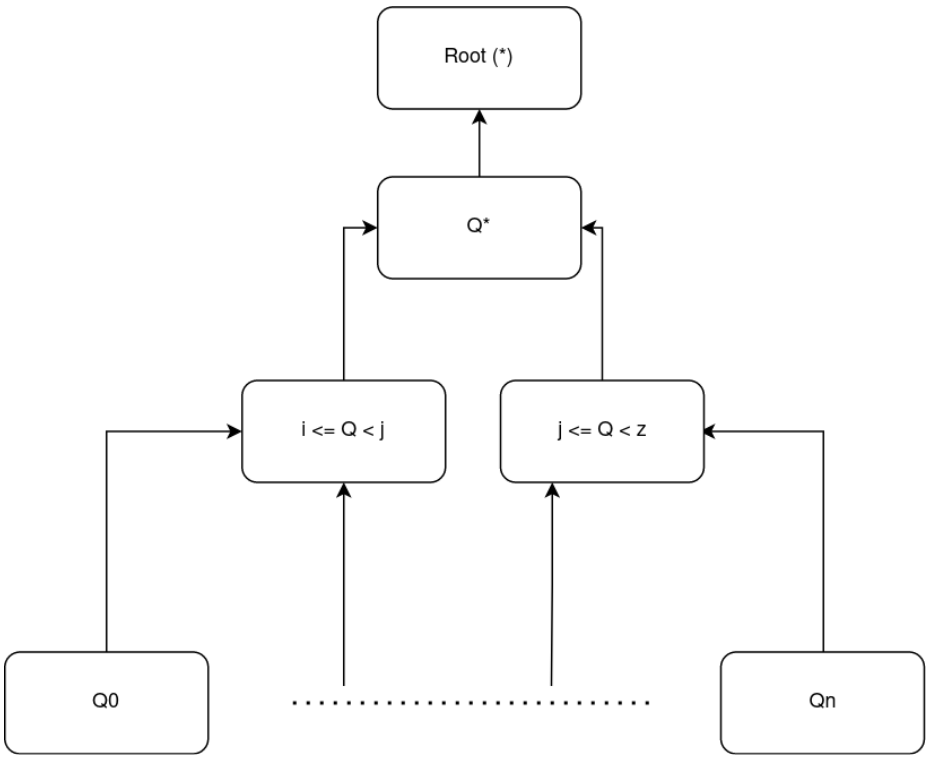

Figure 3.2: The generic scheme used as a generalization hierarchy in the system.

The i, j, and z shown in Figure 3.2 are calculated based on a value H that is derived from the value of K and the size of records present in the data set, whereas the Q corresponds to the quasi-identifier value. The H value is dynamically selected based on the value K. The hierarchy shown in Figure 3.2 is constructed for each attribute, i.e., each part of the topological location. Figure 3.3 below 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 5.

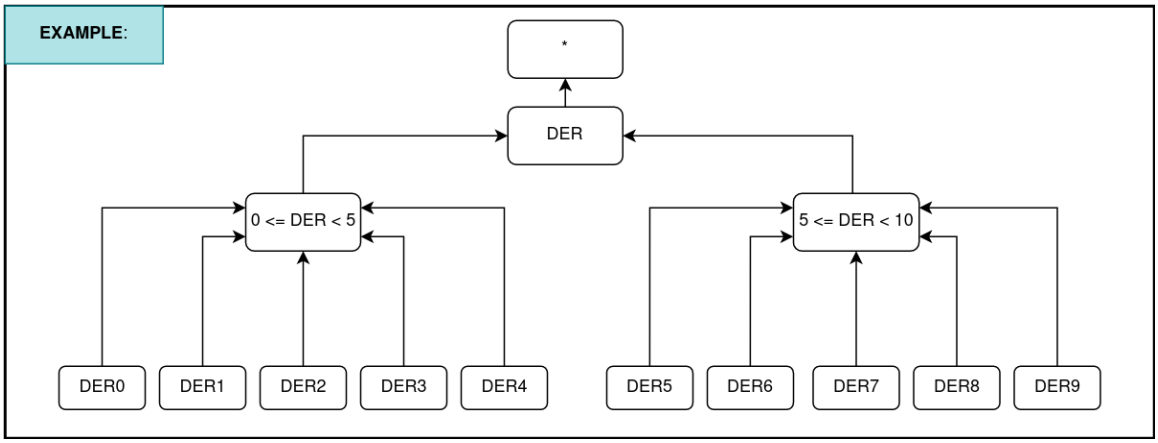

Figure 3.3: An example of DER attribute hierarchy in the system. Where K is 5, and H is 5.

3.3 Central Distributed Trust Aggregator Interface

3.3.1 Sequence Diagrams

The majority of information exchange between actors in the EGoT DERMS system is defined according to one or more standards. For instance, DCMs and DERs follow either the SunSpec Modbus or the CTA-2045 standard at this current stage of the EGoT DERMS development. Moreover, GO and GSP communication is governed by the interface provided by the GO. However, trust layer data exchange is not defined by any standards, specifically, the CDTA and DTMC communication by which DTMCs send trust data upstream for aggregation. Figure 3.4 showcases a sequence diagram of normal interactions between the system actors.

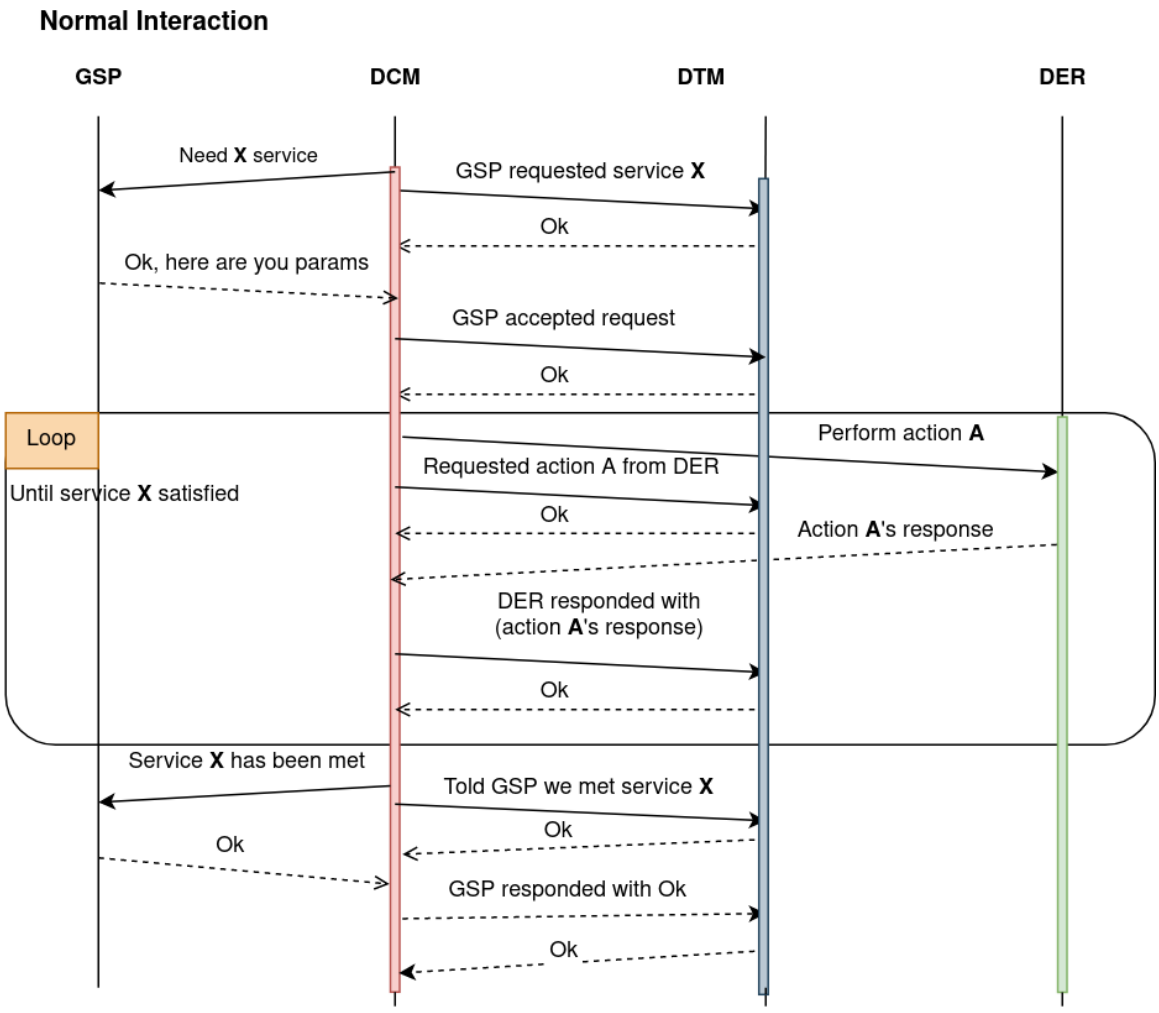

Figure 3.4: An overview of normal interaction between actors in EGoT DERMS.

Figure 3.4 shows normal interactions between actors. Notice that the displayed interaction does not account for any privacy measures. For example, if the DTMC uses all the information passed by the DCM as shown in the figure, one can, with high certainty, infer lots of the DER characteristics based on its activity. Additionally, in the case of trust violation, the CDTA can easily zero in on the culprit actor, which could be a desirable trait in a system but comes at the cost of the privacy issue, as mentioned earlier.

K-Anonymity can be employed to address such contingencies. In particular, the GSP can

replace the topological IDs with anonymized ones. Keep in mind that the plain topological IDs are part of the service parameters returned by the GSP. Thus, the GSP would need to utilize the Mondrian algorithm sometime before the DCM requests services. An appropriate time for such a step can be periodic, upon changes in customers' participation or when a new customer has registered in the system. Figure 3.5 below outlines the same interaction shown in Figure 3.4 but with minor additions to accommodate the new changes.

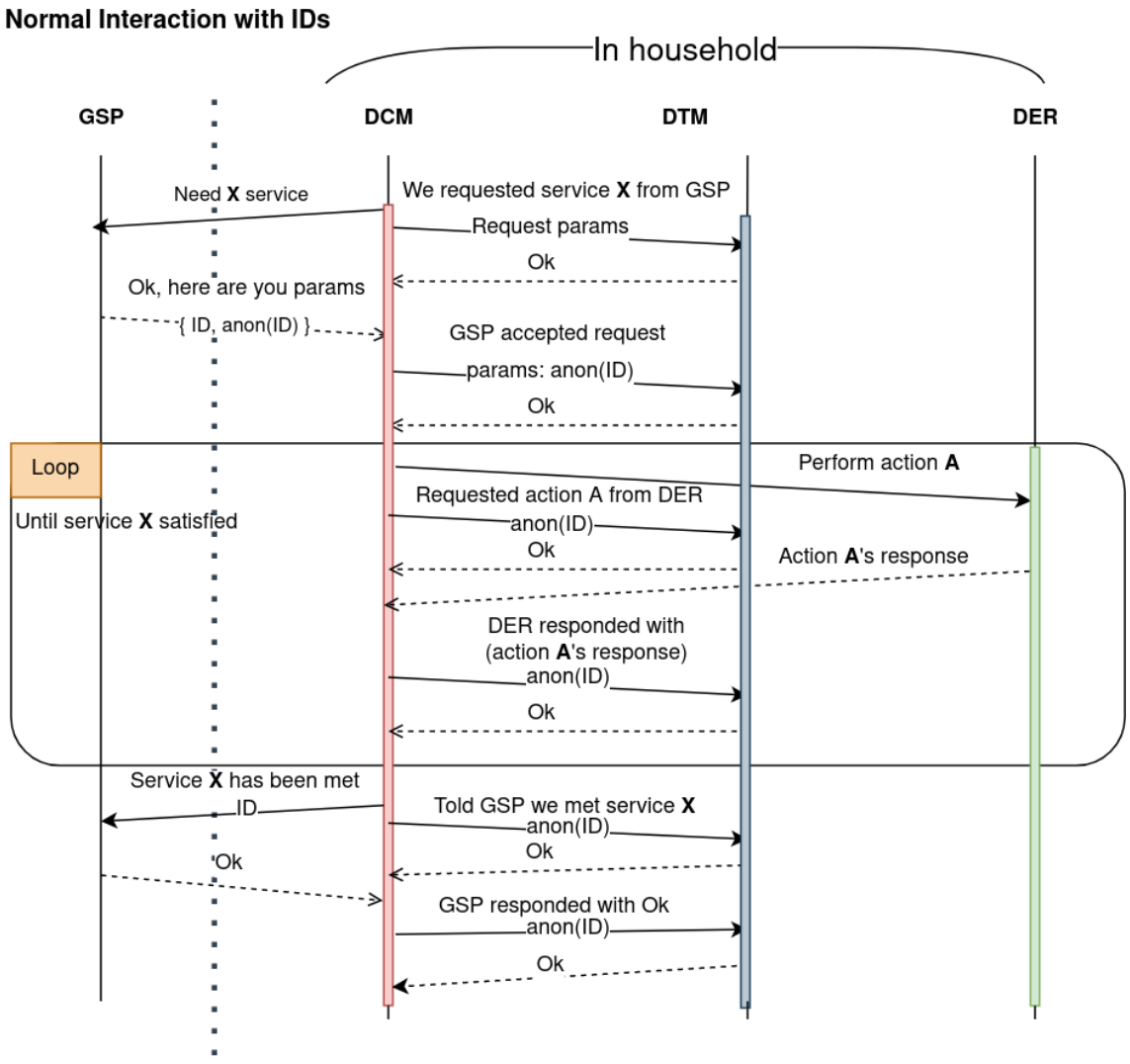

Figure 3.5: An overview of normal interaction between actors with the addition of IDs.

As previously mentioned, DTMCs use all outgoing and incoming communication involving DCMs to compute local trust within their respective SPCs. Then, DTMCs must send the resulting MVoTs upstream to the CDTA for aggregation. Figure 3.6 illustrates the drafted sequence diagram through which trust information can be transferred from local DTMCs to remote CDTAs.

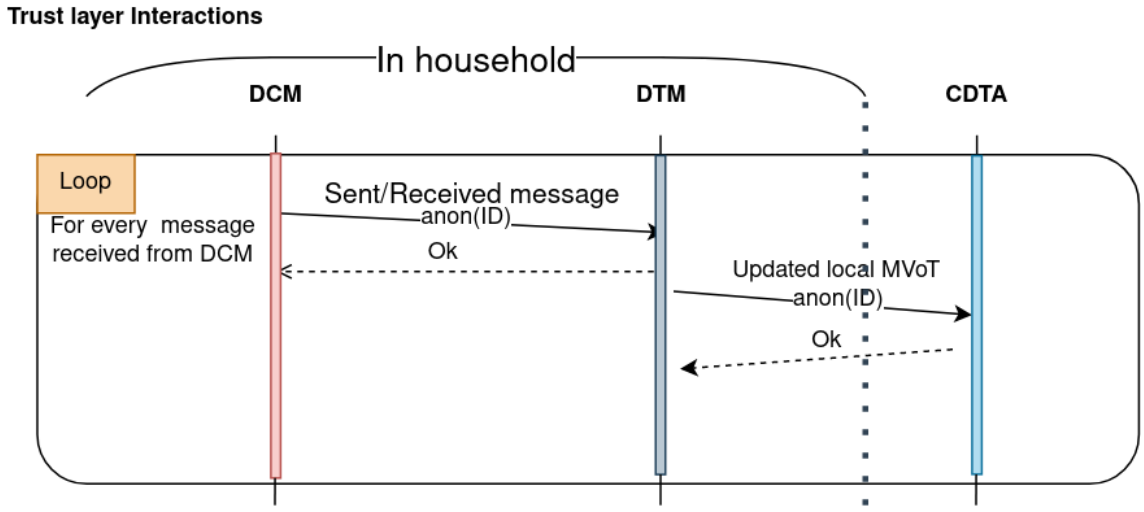

Figure 3.6: An overview of trust layer interactions in EGoT DERMS.

Finally, Figure 3.7 coalesces the diagrams with new additions; by way of explanation, it combines Figure 3.5 and Figure 3.6. Moreover, it showcases the periodic update from the DTMC to the CDTA upon each networking event with the DCM. The red dashed lines denote the boundaries of the SPC, which also translate to trust boundaries where every interaction within the SPC is LAN communication.

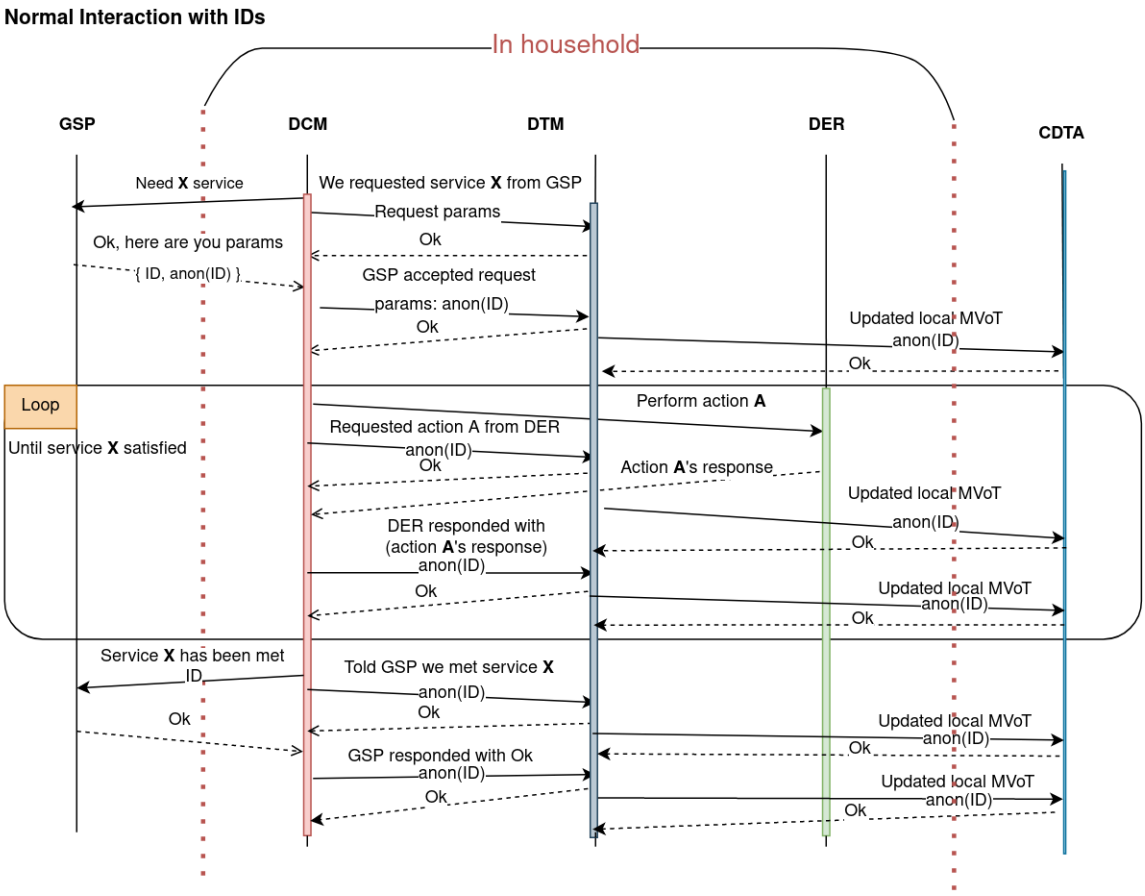

Figure 3.7: Complete sequence diagram for actors in the system, including trust layer actors.

3.3.2 Central Distributed Trust Aggregator API

Now that there is a plan of action for aggregating trust information represented as MVoTs, there needs to be an interface in place as a means for aggregating said information. A simple HTTP server was developed that exposes an Application Programming Interface (API) to meet such requirements. This API provides a single endpoint, /MVoT, with two available methods. The available methods include a POST method through which DTMCs can post local MVoT records. The second is a GET method, which returns all MVoT records collected and can be used for other features in the future, such as dashboard applications

and the like.

The CDTA stores the collected records in a relational database that uses the single-table design. Moreover, the body of POST requests is expected to have an XML content type. As such, the CDTA validates each posted record to ensure that it contains all the necessary parameters of an MVoT record. Figure 3.8 contains a list of MVoT parameters used to model trust in EGoT DERMS.

MVoTs	
PK	**mvot_id int NOT NULL**
	last_timestamp date NOT NULL
	registration_date date NOT NULL
	SDTT float
	RFC float
	TSLC float
	tx_time float
	comm_freq float
	certainty float
	avg_tx_time float
	trust_score float
	distrust_score float
	total_msgs int
	other_count int
	alert_count int
	timeout_count int
	count_expected_msgs int
	count_unexpected_msgs int

Figure 3.8: A list of MVoT parameters used to model trust in EGoT DERMS.

Note that the *mvot_id* field is the primary key for the database. It is a primary key that the database automatically increments, and it must not be part of the request body. Primarily, it is used for indexing purposes; therefore, if an *mvot_id* field was part of any POST request, the CDTA ignores that field and uses everything else when validating the request. Table 3.1

provides a brief definition for each MVoT parameter shown in Figure 3.8.

MVoT parameter	**definition**
anon_id	Anonymized ID of DER for that service. If the DTMC does not have this, it should use the default *:*:*:* ID (complete anonymization).
registration_date	The date of device registration
last_timestamp	The timestamp for the last message.
sdtt	The standard deviation of transit time for messages.
RFC	The relative factor of certainty
TSLC	Time passed since last communication
tx_time	The transit time for the last message
comm_freq	Calculated communication frequency
certainty	The current certainty levels.
avg_tx_time	The average transit time of observed messages.
trust_score	The calculated trust score (with the last network event factored in).
distrust_score	The calculated distrust score (with the last network event factored in).
total_msgs	The count of total messages sent or received by the DCM.
alerts_count	The count of alert messages raised in the past.
timeout_count	The count of message timeout encountered in the past.
count_expected_msgs	The count of expected messages received in the past.
count_unexpected_msgs	The count of unexpected messages received in the past.

Table 3.1: Table of brief definitions for MVoT parameters.

4 Results & Analysis

For every approach, there are advantages and disadvantages. With that in mind, this section explores the results obtained from the provided work discussed in this thesis. This section is divided into two subsections: K-Anonymity, and the CDTA-DTMC interface.

4.1 K-Anonymity

Since the GSP has access to records of DCMs registered in the system, such a dataset was used to produce anonymized data. This data set was generated according to the 13 node feeder design, which is a test feeder used as the system is still in development. Figure 4.1 showcases a side-by-side comparison between the original data set and the 2-anonymized resulting data set. The generalization hierarchy discussed previously was employed, and H here is 5.

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

Figure 4.1: Sample of 2-anonymization effects on IEEE 13 node feeder data. The table on the left contains records sampled from IEEE 13 node feeder topology, whereas the table on the right contains the records after anonymization.

Figure 4.1 shows the visible effects of the H value as the records have been aggregated in groups of 2s, 3s, and 5s, which is greater or equal to two and satisfies the 2-anonymization requirement. In the case where H is selected to be equal to K, the generalization hierarchy forces the algorithm to produce the results shown in Figure 4.2.

Effects of the Mondrian Algorithm on a data set with K = 5

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 0	segment 7	transformer 4	DER 13
substation 0	segment 7	transformer 4	DER 14

=>

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

Figure 4.2: Sample of 5-anonymization effects on IEEE 13 node feeder data. The table on the left contains records sampled from IEEE 13 node feeder topology, whereas the table on the right contains the records after anonymization.

When K is two, the segment part of the IDs had been anonymized in groups of 2s and 3s for the first five records. In contrast, when k is 5, the segment needs to be generalized to the highest level (suppression of the numerical part) for the same records. Otherwise, there is no way to aggregate those records such that they are indistinguishable from each other.

Since the Mondrian Algorithm provides multi-dimensional K-anonymity, it takes into account recoding all sensitive attributes when anonymizing. For instance, all the attributes in the highlighted record in Figure 4.1 and Figure 4.2 were used when constructing equivalence classes. Suppose the algorithm didn't account for all attributes, which means it isn't 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.

4.1.1 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 4.3 demonstrates the information loss observed when the algorithm is run on the 13 node test feeder data set under various K-values. Figure 4.3 shows that as k grows in size, the penalty grows to approach 20% slowly. 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 EGoT DERMS.

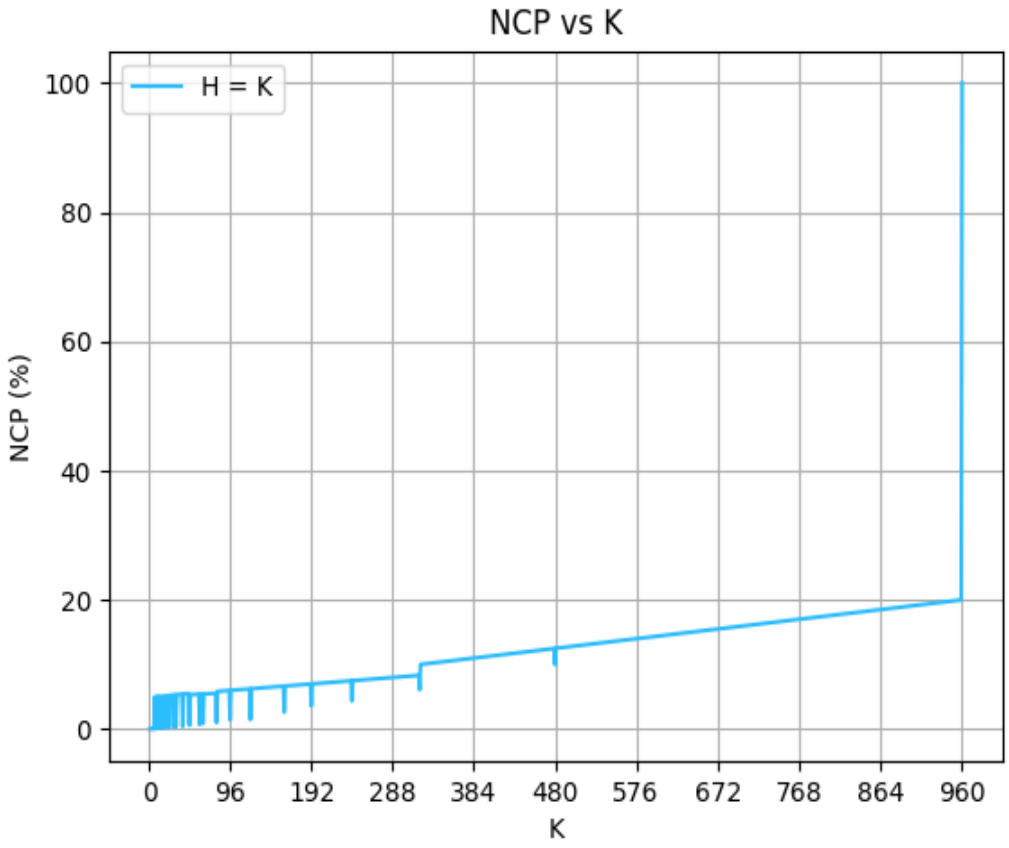

Figure 4.3: Plot of NCP against different K values for IEEE 13 node feeder data.

The penalty spikes to 100% once K reaches the size of the data set as there is no way of retaining information when K is equal to the total number of records. Note that only the substation information is retained at this elbow point, and everything else ends up suppressed instead of generalized to some level in the hierarchy, as shown by a small sample output in Table 4.1. This point is preceded by the moment when the algorithm has expended four out of five attributes to generalize.

substation	feeder	segment	transformer	DER
substation0	*	*	*	*
substation0	*	*	*	*
substation0	*	*	*	*
substation0	*	*	*	*
substation0	*	*	*	*
substation0	*	*	*	*
substation0	*	*	*	*
substation0	*	*	*	*
substation0	*	*	*	*

Table 4.1: Example output when K = size of the data set. Here suppression is used for all sensitive attributes.

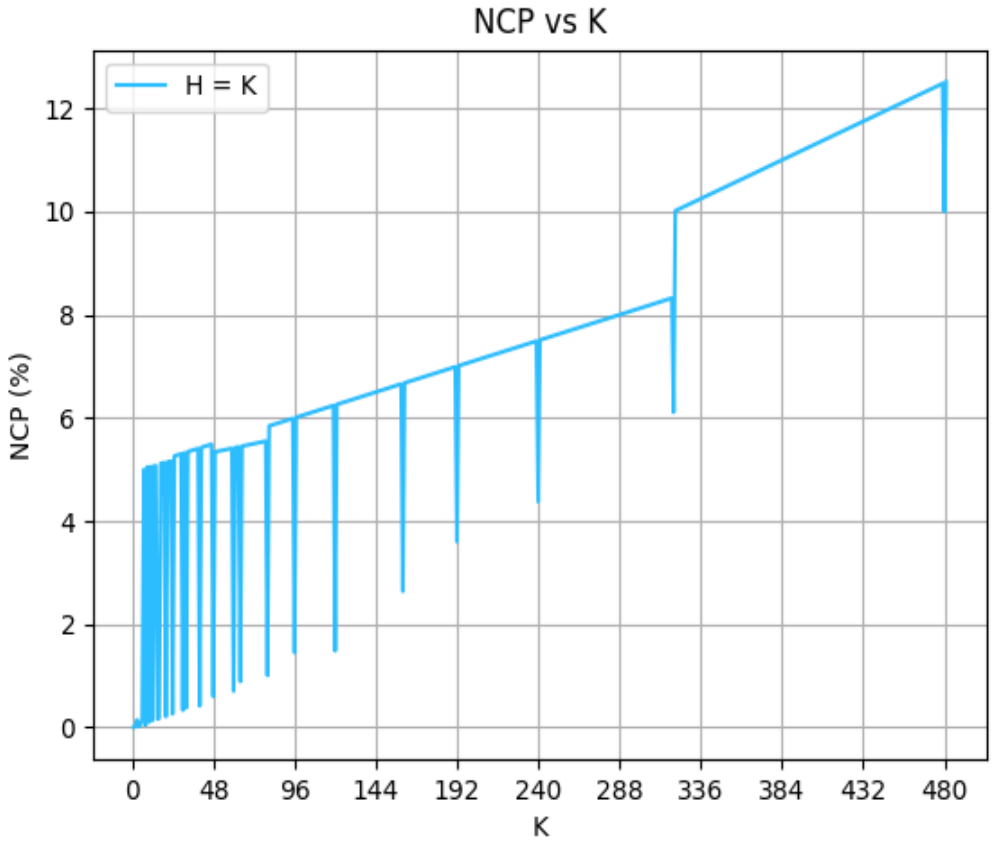

Figure 4.4: Plot of NCP against different K values for half of IEEE 13 node feeder data.

Figure 4.4 shows a plot of information loss against variable values of K. In Figure 4.4 only the first half of the data set 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.

4.2 CDTA Interface

Currently, some of the essential features of the EGoT DERMS are under development. For instance, the registration process, which allows new customers to join the system, is in the testing phase. There is a need to simulate devices to ensure simultaneous development. Case in point, the Trust Model Simulator (TMS) and Trust Model Data Generator (TMDG) were developed to aid in the concurrent development of the EGoT DERMS and the DTM System. Specifically, the objectives of TMS and TMDG are to generate and simulate messages representative of EGoT DERMS actors interactions.

The existence of MVoT data governs the CDTA interface. It means little whether the data is real or fake. The current system uses data generated by the previously discussed TMDG. Moreover, data obtained from the IEEE 13 node feeder was used to generate original and anonymized IDs. By complementing TMDG generated data with the resultant anonymized

IDs, representative instances of message exchange were obtained. These data become valuable for testing purposes. For instance, the following listings exhibit the interactions received by the CDTA for all available methods (i.e., GET and POST).

4.2.1 POST Request Example

```
<?xml version='1.0'?>
<MVoT>
    <anon_id>
        substation0:segment7:xformer2:DER 40-50
    </anon_id>
    <registration_date>
        1629393715.096357
    </registration_date>
    <last_timestamp>1629394315.096386</last_timestamp>
    <sdtt>0.1112444366039249</sdtt>
    <RFC>1.0</RFC>
    <TSLC>300.00001287460327</TSLC>
    <tx_time>0.17</tx_time>
    <comm_freq>0.0049999997595946</comm_freq>
    <certainty>0.0002321533514997</certainty>
    <avg_tx_time>0.131038961038961</avg_tx_time>
    <trust_score>0.0006344171455322</trust_score>
    <distrust_score>0.0</distrust_score>
    <total_msgs>3</total_msgs>
    <other_count>0</other_count>
    <alerts_count>0</alerts_count>
    <timeout_count>0</timeout_count>
    <count_expected_msgs>3</count_expected_msgs>
    <count_unexpected_msgs>0</count_unexpected_msgs>
</MVoT>
```

4.2.2 GET Response Example

```
<?xml version='1.0'?>
<MVoTs>
    <MVoT>
        <mvot_id>1302</mvot_id>
        <anon_id>*:*:*:*</anon_id>
        <registration_date>
            1629393715.096362
        </registration_date>
        <last_timestamp>
            1629537715.103173
        </last_timestamp>
        <sdtt>0.0108067067597848</sdtt>
        <RFC>1.0</RFC>
        <TSLC>300.00001192092896</TSLC>
        <tx_time>0.34</tx_time>
        <comm_freq>0.0033402776197895</comm_freq>
        <certainty>0.0011134258289796</certainty>
        <avg_tx_time>0.2538750830835711</avg_tx_time>
        <trust_score>0.53556014608144 38</trust_score>
        <distrust_score>0.0</distrust_score>
        <total_msgs>481</total_msgs>
        <other_count>0</other_count>
        <alerts_count>0</alerts_count>
        <timeout_count>0</timeout_count>
        <count_expected_msgs>481</count_expected_msgs>
        <count_unexpected_msgs>0</count_unexpected_msgs>
    </MVoT>
    ....
</MVoTs>
```

Currently, the CDTA authentication is done via self-signed certificates. Primarily this is because the EGoT DERMS is in its infancy stages. Further, some suggested security measures provided in IEEE 2030.5 can be employed to fortify the security posture of this communication channel. Finally, the provided endpoints facilitate information exchange in

the trust layer and further applications to be built on top. Such future applications will be addressed in the Discussion chapter.

5 Discussion

5.1 Information Loss

As discussed earlier, results obtained in Figure 4.3 and Figure 4.4 indicate the existence of some H values that minimize information loss. Figure 5.1 and Figure 5.2 demonstrate the performance of a simple heuristic used to minimize 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 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.

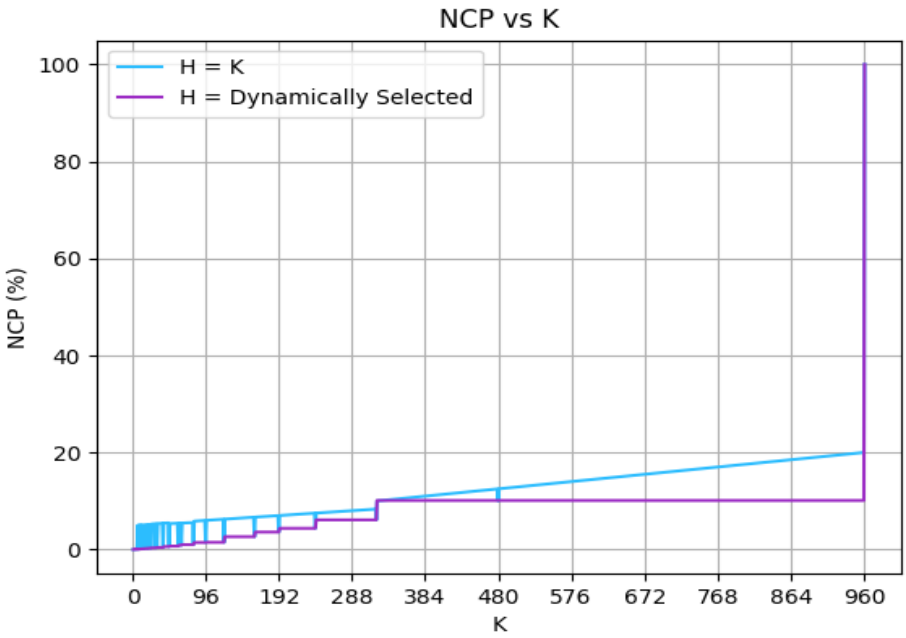

Figure 5.1: Plot of NCP against different K values using two different heuristics. Choosing H to equal K results in higher overall penalty incurrence than dynamically selecting H.

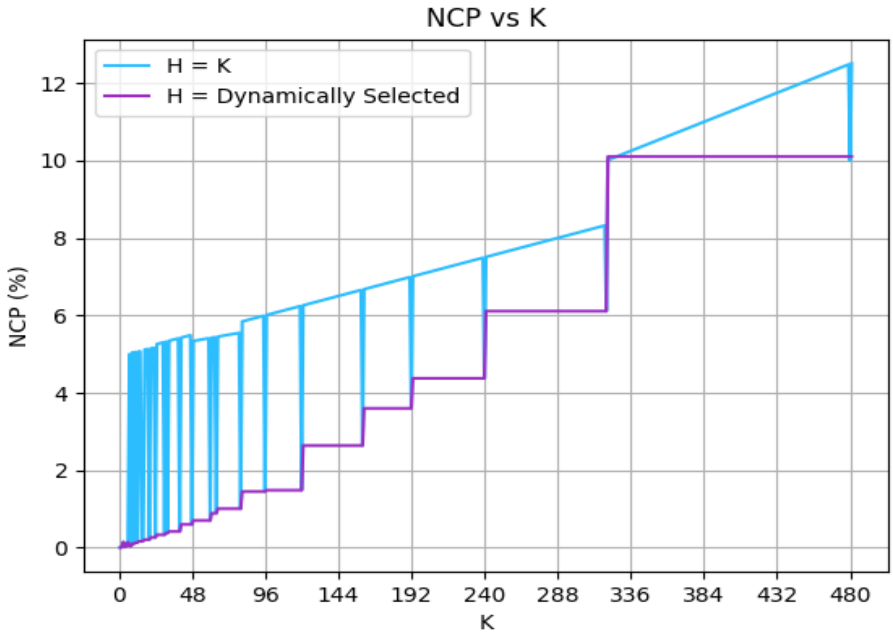

Figure 5.2: Plot of NCP against K values using different heuristics for half of the data from Figure 5.1. A staircase effect can be observed when H is dynamically selected where each step corresponds to dips in the naive heuristic.

5.2 Alternative Approaches

An alternative method exists for achieving optimal K-anonymity for the used data set. For instance, as defined in this work, the PSU-modified IEEE 13 node feeder model uses distribution transformers that all have the same power rating. Such a case leads to a special case where the generalization hierarchy exhibits what is called an Identical Generalization Hierarchy (IGH). According to Mahana et al., while finding an optimal solution to the K-anonymity problem is NP-hard, IGHs present a unique case that can be solved efficiently [35]. However, it is essential to remember that the IEEE 13 node feeder is used for test purposes and would never occur in reality. Guarantees of sameness, while it holds for the used test data set, would seldom occur in the real world and would limit the contribution of this work in the future. Case in point, transformer power ratings vary within a distribution system, ensuring that the number of loads on transformers varies as well.

5.3 ACL-based Authorization

An Access Control List (ACL) is a mechanism for managing permissions associated with system resources [36]. Such a list can be used to add authorization for the EGoT DERMS. Ideally, an ACL is defined for each resource in the system. IEEE 2030.5 recommends the use of ACL-based authorization. For instance, IEEE 2030.5 provides an example ACL for EndDeviceList resource available by the GSP for its clients. Table 5.1 shows the provided example in the standard documentation. The Method field in aclDefaultAccess shown in the table is set to 0x00, indicating that only the GET method is available for the

clients. The AuthType is set to 0x1, which means no authentication, to indicate the use of the default mechanism (0x1 is the default value for the AuthType field). Lastly, the DeviceType is set to 0 to indicate that any device can access the resource. The example provided has no aclSpecificID as there are no further rules to enforce for specific clients, which means all clients have access to the resource. Similar to the value of aclSpecificID, the aclSpecificIDEntries is set to 0 as there are no entries to describe how the specific clients should access the resource. Note that Authorization is granted if all components of aclDefaultAccess are true (Method, AuthType, and DeviceType). If any of the components is false, the server should return 401 (unauthorized).

Attribute	**Comment**
aclDefaultAccess	Method: 0x01 (GET only)) AuthType: 0x1 (no authentication) DeviceType: 0 (any device type))
aclSpecificID	-
aclSpecificIDEntries	0

Table 5.1: Example ACL for the EndDeviceList resource as described in IEEE 2030.5 [18].

Attribute	**Comment**
aclDefaultAccess	Method: 0x1, 0x4 (GET and POST only) AuthType: 0x4 (self-signed authentication) DeviceType: 0 (any Device Type)
aclSpecificID	-
aclSpecificIDEntries	0

Table 5.2: Example ACL for CDTA resources as described in IEEE 2030.5.

To conform with the rest of the resources available by GSPs in the EGoT DERMS, the CDTA API should use the same authorization mechanisms suggested by the IEEE 2030.5 standard. Table 5.2 describes the suggested ACL associated with CDTA accessible resources

for DTMCs. The aclDefaultAccess is set to the GET and POST methods only, as they are the only ones available for the endpoint provided to the clients. The AuthType entry is set to 0x4 to indicate that clients authenticated via self-signed certificates can access the resource. Like the provided example, all devices should be allowed to access the resource. The remaining entries of the ACL are left to the default values as there are no other policies to enforce regarding specific clients. Ideally, this would change in the further stages of the system as only DTMCs and the GSP should be allowed to access the resource. For example, a DCM should not be allowed to post MVoT records to the CDTA.

5.4 Future Directions

One of the future applications for the DTM System is the development of a dashboard application. The dashboard would allow authority figures to assess the trust levels of the entire system and other statistical summaries regarding actor communication patterns. The GET method provided by the available endpoint put in place initial effort. For instance, a frontend dashboard application could query the endpoint to acquire recent MVoT entries before visually representing trust. Moreover, access to recent MVoT records makes way for other applications such as deep learning to enhance network security further. For example, having a data set that describes trust scores over different conditions can be optimized so that thresholds for alerts are dynamically chosen. Developing a model for picking optimized thresholds would be another possible application.

6 Conclusion

The EGoT DERMS adopts a SOLC approach to manage Distributed Energy Resources. Such architecture relies on heavy digital interaction between the system actors to achieve its 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.

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 excludes PII. Also, the strategy secures the communication channel according to IEEE 2030.5 specification. 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.

Appendix A: Source Code

One can find the source code for the adopted Mondrian algorithm and the CDTA interface on Portland State University PowerLab GitHub account. The logic for obtaining the anonymized IDs would be invoked by the GSP and can be found on https://github.com/PortlandStatePowerLab/EGoT_KAnonymity. The source code for the CDTA interface, which would run on the CDTA server, can be found on https://github.com/PortlandStatePowerLab/CDTA_API.

Bibliography

[1] E. Denny and M. O'Malley. Wind generation, power system operation, and emissions reduction. *IEEE Transactions on Power Systems*, 21(1):341–347, 2006.

[2] Qixin Chen, Chongqing Kang, Qing Xia, and Jin Zhong. Power generation expansion planning model towards low-carbon economy and its application in China. *IEEE Transactions on Power Systems*, 25(2):1117–1125, 2010.

[3] 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.

[4] T. Slay, J. M. Acken, and R. B. Bass. Incentivizing distributed energy resource participation in grid services. In *9th IEEE Conference on Technologies for Sustainability (accepted)*, 2022.

[5] R. Bayindir, I. Colak, G. Fulli, and K. Demirtas. Smart grid technologies and applications. *Renewable and Sustainable Energy Reviews*, 66:499–516, December 2016.

[6] C.W. Gellings. The concept of demand-side management for electric utilities. *Proceedings of the IEEE*, 73(10):1468–1470, 1985.

[7] M. Adham, M. Obi, and R. Bass. A field test of direct load control of water heaters and its implications for consumers. In *2022 IEEE Power and Energy General Meeting (accepted)*, 2022.

[8] 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 Journal of Power and Energy*, 7:22–30, January 2020.

[9] Kevin Marnell, Conrad Eustis, and Robert B. Bass. Resource study of large-scale electric water heater aggregation. *IEEE Open Access Journal of Power and Energy*, 7:82–90, February 2020.

[10] T. Slay and R. Bass. An energy service interface for distributed energy resources. In *IEEE Conference on Technologies for Sustainability*, 2021.

[11] T. Slay, G. Spitzer, and R. B. Bass. Proposed application for an entity component system in an energy services interface. In *9th IEEE Conference on Technologies for Sustainability (accepted)*, 2022.

[12] M. Alsaid , N. Bulusu, A. Barghouti, N. S. Fernando, J. M. Acken, T. Slay, and R. B. Bass. Privacy-preserving information security for the energy grid of things. In *11th International Conference on Smart Cities and Green ICT Systems (accepted)*, 2022.

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

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

[15] Zheng Yan. A comprehensive trust model for component software. In *Proceedings of the 4th international workshop on Security, privacy and trust in pervasive and ubiquitous computing*. ACM Press, 2008.

[16] N. S. Fernando, J. M. Acken, and R. B. Bass. Developing a distributed trust model for distributed energy resources. In *IEEE Conference on Technologies for Sustainability*, 2021.

[17] J. M. Acken, A. Barghouti, N. Fernando, and R. B. Bass. title here. In *2022 International Conference on Cyber Security and Resilience (accepted)*, 2022.

[18] IEEE. IEEE standard for smart energy profile application protocol (IEEE 2030.5-2018), December 21, 2018.

[19] SunSpec Alliance. Common smart inverter profile (CSIP), March 16, 2018.

[20] W.H. Kersting. Radial distribution test feeders. In *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, volume 2, pages 908–912 vol.2, 2001.

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

[22] R.J. Bayardo and Rakesh Agrawal. Data privacy through optimal k-anonymization. In *21st International Conference on Data Engineering (ICDE'05)*, pages 217–228, 2005.

[23] Athanasios Andreou, Oana Goga, and Patrick Loiseau. Identity vs. attribute disclosure risks for users with multiple social profiles. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM, July 2017.

[24] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1):3–es, mar 2007.

[25] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *IEEE 23rd International Conference on Data Engineering*, pages 106–115, 2007.

[26] Mark Stegelmann and Dogan Kesdogan. Gridpriv: A smart metering architecture offering k-anonymity. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 419–426, 2012.

[27] B. Yuce, M. Mourshed, Y Rezgui, and O. F. Rana. Preserving prosumer privacy in a district level smart grid. In *2016 IEEE International Smart Cities Conference (ISC2)*, pages 1–6, 2016.

[28] 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 Journal*, 6(4):5902–5915, 2019.

[29] Suvda Myagmar, Adam Lee, and William Yurcik. Threat modeling as a basis for security requirements. 08 2005.

[30] Mohiuddin Ahmed and Al-Sakib Khan Pathan. False data injection attack (FDIA): an overview and new metrics for fair evaluation of its countermeasure. *Complex Adaptive Systems Modeling*, 8(1), April 2020.

[31] Adnan Anwar, Abdun Naser Mahmood, and Zubair Shah. A data-driven approach to distinguish cyber-attacks from physical faults in a smart grid. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, page 1811–1814, New York, NY, USA, 2015. Association for Computing Machinery.

[32] K. LeFevre, D.J. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In *22nd International Conference on Engineering*, pages 25–25, 2006.

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

[34] Jian Xu, Wei Wang, Jian Pei, Xiaoyuan Wang, Baile Shi, and Ada Wai-Chee Fu. Utility-based anonymization using local recoding. In *Proceedings of the 12th ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining, page 785–790, New York, NY, USA, 2006. Association for Computing Machinery.

[35] Waranya Mahanan, W. Art Chaovalitwongse, and Juggapong Natwichai. Data anonymization: a novel optimal k-anonymity algorithm for identical generalization hierarchy data in IoT. *Service Oriented Computing and Applications*, 14(2):89–100, February 2020.

[36] Stefan Miltchev, Jonathan M. Smith, Vassilis Prevelakis, Angelos Keromytis, and Sotiris Ioannidis. Decentralized access control in distributed file systems. *ACM Comput. Surv.*, 40(3), Aug 2008.