https://doi.org/10.1016/j.future.2019.12.017

A Privacy-Preserving Approach to Prevent Feature Disclosure in an IoT Scenario

Serena Nicolazzo[1], Antonino Nocera[2*], Domenico Ursino[3], and Luca Virgili[3]

[1] Sky Italia
[2] DIII, University of Pavia
[3] DII, Polytechnic University of Marche
[*] Contact Author
serena.nicolazzo@skytv.it; antonino.nocera@unipv.it; d.ursino@univpm.it;
l.virgili@pm.univpm.it

Abstract

In this paper, we propose a privacy-preserving approach to prevent feature disclosure in a multiple IoT scenario, i.e., a scenario where objects can be organized in (partially overlapped) networks interacting with each other. Our approach is based on two notions derived from database theory, namely k-anonymity and t-closeness. They are applied to cluster the involved objects in order to provide a unitary view of them and of their features. Indeed, the use of k-anonymity and t-closeness makes derived groups robust from a privacy perspective. In this way, not only information disclosure, but also feature disclosure, is prevented. This is an important strength of our approach because the malicious analysis of objects' features can have disruptive effects on the privacy (and, ultimately, on the life) of people.

Keywords: Privacy-Preserving Approach; Internet of Things; Feature Disclosure Prevention; Multi-IoTs Scenario; Object Grouping Scheme for Privacy Management

1 Introduction

In the last few years, we are assisting to the enormous increase of the number of sensors and devices, which are becoming extremely pervasive and used in most contexts of daily life. At the same time, objects are developing awfully smart and social skills. All these aspects are revolutionizing the Internet of Things (hereafter, IoT) [1]. As a proof of this, more and more researchers are beginning to study the behavior of things, to talk about their profiles and their social interaction [2], and to manage objects almost as if these were humans. As a result of these investigations, several architectures implementing these ideas have been proposed, and are currently being proposed, in the literature. Social Internet of Things (hereafter, SIoT [3]), Multiple IoT Environment (hereafter, MIE [4]) and Multiple Internets of Things (hereafter, MIoT [5]) are only three of the latest architectures with these characteristics.

Such an evolution of the IoT scenario puts researchers in front of several issues that can become important opportunities if correctly addressed. A major example is the huge interest the researchers have shown in security and privacy in IoT. Indeed, in the recent years, many approaches to the definition of security solutions in the context of smart objects have been proposed, such as solutions for intrusion detection [6, 7], access control [8, 9] and privacy [10]. In the context of privacy in IoT, one of the most relevant challenges regards the capability of preserving the privacy of users, who are employing a set of smart objects connected with each other and, possibly, with objects belonging to other users. In such a scenario, characterized by the pervasive presence of smart objects, a lot of user's data can be produced by the smart objects she is using. This scenario appears even more complex if we consider that objects are becoming increasingly autonomous when they perform their tasks. Among these, one of the most important and crucial for the whole IoT is the interaction with other objects. In order to refine and improve this capability, objects may use and propagate information about the features they can provide. This information allows other objects to improve the selection of the preferred contacts and to enhance their querying capability. However, if properly combined with other data, it can provide sensitive information about the user, which she had no intention of disclosing. Knowing the features of more objects adopted by users, the amount of sensitive knowledge about her that can be derived dramatically increases.

To give an example of what we stated above, let us consider a scenario in which a person is in a hospital because she is suffering from gastrointestinal disorders. To carry out diagnosis, she must undergo several analyses in different departments of the hospital. The simplest and fastest of these analyses can be performed through smart objects. For example, in one department, the patient could be connected to an insulin meter, in another one she could be connected to a heart rate meter, and so forth. Knowing that a patient is connected to a specific device (for instance, the insulin meter) already discloses important sensitive information about her (in particular, that she could suffer from diabetes or some pancreatic disease). Knowing also that she is connected to more devices that are simultaneously used for the test of the condition of a specific organ (for example, the devices used to diagnose pancreatic disorders, such as diabetes, pancreatitis or pancreatic cancer), the amount of sensitive information disclosed becomes much more serious because we know in detail what are the possible diseases that doctors suspect may affect the patient.

As a second example, let us consider a patient that simultaneously undergoes three tests, one for the measurement of blood sugar, one for measuring the level of hemoglobin in the blood and one for measuring her respiratory function. Just knowing that she is carrying out only one of these tests, we can hypothesize several diseases from which she may suffer (for example, we may hypothesize that she is using glucose meter because she is suspected of suffering from diabetes). But if we know that she is carrying out these three tests simultaneously, we might conclude that doctors suspect she might have lung cancer, considering that some forms of lung cancer involve important variations in blood sugar and hemoglobin.

In this paper, we aim at addressing this issue by proposing a privacy-preserving approach to prevent feature disclosure in an IoT scenario. Our approach is not focused on specific queries. Instead, as said before, it aims at preventing the disclosure of sensitive information of a user that can happen simply by examining the features of the devices she is employing. Taking also into account that utility and privacy is a major trade-off for privacy-preserving techniques, our approach aims at preserving

all existing information about user-object interaction. In fact, this information is extremely useful to support other applications and possible analyses on an IoT scenario. On the other hand, our approach is capable of protecting users' privacy by partially hiding objects' features still allowing their full exploitation in order to support objects' communication.

In more details, our approach leverages some traditional concepts from databases, such as k-anonymity [11] and t-closeness [12]. The basic idea of both these paradigms is to group data together so that the same piece of information is present in at least k records. This creates a sort of blurred cloud of data, in which it is not possible to successfully map the protected piece of information to a specific record among the k sharing it. Of course, when dealing with data distribution, it is possible to reduce the number of candidate records to be associated with a specific feature by exploiting the probability that a record contains that piece of information. The t-closeness paradigm overcomes this possibility by imposing criteria based on the probability distribution when selecting the admissible values used to k-anonymize a sensitive piece of information.

Our approach applies k-anonimity and t-closeness to build small conglomerates, hereafter referred as *groups* of objects, inside an existing network with the purpose of creating a single view of the objects present in each of them. The individuality of smart objects is preserved from a connectivity point-of-view, whereas their features are mixed inside each group. From the outside, a smart object presents itself by advertising the features available in the group it belongs to. Groups are built by solving a trade-off between privacy requirements and communication performance. k-anonymity and t-closeness are combined to make each group robust from a privacy perspective by properly selecting the number of features, their typology, and the number of objects as tuning parameters in order to meet the desired protection level.

Our approach is orthogonal to the existing strategies for the protection of communication channels and data exchange among objects, such as the ones described in [13, 14, 15, 16, 17]. Moreover, while many researchers have been developing frameworks to protect *object* interaction from both a security and privacy perspective, our approach focuses on the effects produced on the privacy of the *users* by the direct observation of the objects (and the corresponding features) they are employing. As a matter of fact, with the evolution of smart objects, techniques to allow the automatic interactions among them based on proximity or homogeneity have been developed [18]. As stated above, such strategies can be improved by using object scopes and features; therefore, enabling feature advertising is an important point and a key aspect for improving object interactions in the IoT. This consideration, combined with the observation that the knowledge of object features is an important vehicle to privacy leakage, leads to the need of a stable solution that enables these interactions in a privacy-preserving way.

Our proposal refers to such a scenario and presents a solution in this setting. In its design we also take into account the most recent developments on IoT research. It has been proved that it is more realistic to model an IoT scenario as a set of connected networks, instead of only a unique network of objects. This is due to the number of involved objects, their smartness and social interaction capabilities, as well as the possibility that each portion of the object network may desire to hide part or most of data exchanged inside it [5]. The usage of a multi-network representation of our scenario is a key point in our proposal. Indeed, *(i)* each identified group corresponds to a network of the system; *(ii)* each object can be modeled by means of a node; *(iii)* relationships between objects of the same group can be represented by means of arcs inside the corresponding networks (they are called

inner arcs); *(iv)* relationships between objects of different groups are modeled as arcs linking nodes of different networks (they are called cross-arcs). The possibility to have a direct, natural and immediate multi-network representation of our scenario allows for an easy mapping with properties, operations and concepts of multi-network contexts [19, 20, 21]. Thanks to this, in our analyses, we can benefit from the wide variety of results found for multi-network systems in the past literature.

The outline of this paper is as follows. In Section 2, we examine related literature. In Section 3, we describe the proposed model in detail, whereas in Section 4, we illustrate our privacy-preserving object grouping scheme. In Section 5, we describe our security model. In Section 6, we investigate the relationships between privacy constraints and network performances. In Section 7, we propose a discussion about the peculiarities of our approach. Finally, in Section 8, we draw our conclusions and have a look at possible future developments of our research efforts.

2 Related Work

Like all the areas of networked computing, the IoT presents particular challenges to security and privacy, due to the interconnected nature of the Internet. It means that Internet resources can be attacked from everywhere at every moment. The threats that can affect IoT entities are numerous, such as attacks targeting communication channels, physical threats, denial of service, identity fabrication, and so on [22]. This has led several researchers to develop countermeasures for addressing security and privacy issues specific to the IoT [23, 13, 24, 14, 25]. In particular, in [23], the authors present an overview of security principles, as well as of technological and security challenges; then, they propose countermeasures for securing the IoT. One of the main challenges in this research field is that proposed solutions must cope with the restrictions and limitations in terms of components, devices, computational and power resources characterizing the IoT [26]. On the one hand, the pervasive nature of this technology provides its users with more opportunities to enhance their interactions and to have access to advanced features fostering the creation and consolidation of social relationships. However, on the other hand, it poses new severe technical challenges [27, 28, 29, 30].

Many researchers have adopted Blockchain based strategies to overcome resource availability in the IoT and to propose solutions to privacy and security issues [16, 31, 32, 33]. Specifically, in [16], the authors propose an approach using Blockchain to build a decentralized security and privacy-preserving model. This approach has been thought for smart-home scenarios, in which there is the possibility of having a dedicated high-resource device playing the role of miner. The approach described in [31], instead, uses Blockchain to build a network of gateways, to which smart objects can connect. In this way, even though older devices can be not equipped with resources necessary to implement security and privacy-preserving protocols, they can communicate through the gateway network to overcome their limitations. A further step towards the protection of privacy in the IoT is described in [32]. Here, the authors address data confidentiality in the IoT by combining Attribute-Based Encryption (ABE) with Blockchain to achieve integrity, non-repudiation and confidentiality in IoT communications. Another interesting idea in this context is the one described in [33], in which an approach to build SVM models using data from the IoT, but preserving user privacy, is provided. To reach its goal, this approach uses a Blockchain-based solution in which data collected by smart sensors are first encrypted by means of a homomorphic cryptosystem. Then, each sensor shares encrypted data by using Blockchain as

distributed public ledger. Finally, a modified SVM algorithm working on encrypted data is adopted to train a classifier using such data.

Still in the context of data protection in the IoT, many researchers propose applications leveraging Fog Computing. For instance, in [15], the authors describe an approach to protect privacy of users when data aggregation strategies leveraging Fog Computing delegation are adopted. The peculiarity of this approach, with respect to other well known solutions, such as those described in [34] and [35], relies on the capability of aggregating data from heterogeneous smart devices in a privacy-preserving way. The importance of investigating privacy and security issues when delegating IoT services over Fog Computing solutions is discussed in [36] and [37]. Both these papers provide evidence of the high-impact issues brought about by the adoption of Fog Computing to improve IoT operability.

Other works focus on data confidentiality, i.e., on the objective that data is secure and available only to authorized users. In [13], the authors present an architecture for the IoT security, caring that sensors do not reveal collected data to neighboring nodes. They assure data confidentiality through encryption technologies, which prevent data stealing threats. Furthermore, the authors of [14] focus on how data will be managed, stating that, to ensure protection throughout the process, there must be policies on how to manage several kinds of data, as well as some policy-enforcement mechanisms.

Even though our approach shares some common aspects with the proposals described above, its objective is different. Indeed, most of the approaches above aim at protecting data and avoiding unauthorized access to it. To carry out this task, they operate on the communication channel among objects; some of them also provide facilities to perform privacy-aware data aggregation. Our proposal can be considered as an application on top of existing and consolidated strategies to obtain security and confidentiality in the physical communication channel among objects. Indeed, it focuses on a scenario in which objects directly advertise their capabilities and features (by using existing technologies to interact with other objects in a secure way) to foster the creation of new links in the network. Feature advertising is very common in networking as it is used by the network administrator to detect services running on a device, along with the corresponding versions. This strategy can be also investigated to improve the IoT by means of UPnP scans, through which objects can exchange their descriptions as a response to an HTTP request in an XML document, or by means of Banner Grabbing [38].

Feature description has been adopted in some application scenarios to improve the use of the IoT by exploiting the social-side of this network, in order to filter contents and contacts, thus evolving towards the concept of opportunistic IoT [39] and, therefore, to classify objects data and information for improving their interactions [40]. Also for these approaches, the knowledge of the features and the kind of information that an object can produce is a very important aspect and has been used in different applications, such as service discovery in the IoT [41].

It is worth underlying that, the impact to privacy of both service discovery and feature disclosure in the IoT has been already an important subject of study. Indeed, the recent scientific literature on the IoT includes numerous proposals of privacy protecting schemes in this context [42, 43, 44]. In particular, the authors of [42] illustrate a new approach to private authentication and service discovery in the IoT. This approach ensures the mutual privacy for both the device delivering the service and the one exploiting it. It can also guarantee that the service is authentic (unforgeable service). In [43], the author proposes a solution to the problem of privacy-preserving service discovery and access control. This strategy is, then, successfully deployed in a smart-home scenario. Another interesting

IoT	Internet of Things		SIoT	Social Internet of Things
MIE	Multiple IoT Environment		MIoT	Multiple Internets of Things
n_i	the i^{th} node		P_i	the profile of n_i
ϕ_i	the set of the features exposed by n_i		G_k	the k^{th} group
min_k	the minimum number of nodes of G_k		max_k	the maximum number of nodes of G_k
φ	a feature		NS_k	the set of the nodes of G_k
NS_k^P	the set of the nodes permanently associated with G_k		NS_k^T	the set of the nodes temporarily assigned to G_k
Φ_k	the set of the features exposed by G_k		WZ	the Welcome Zone
$\mathcal{M}$	a MIoT		N	the set of the nodes of $\mathcal{M}$
A	the set of the arcs of $\mathcal{M}$		A_I	the set of the i-arcs of $\mathcal{M}$
A_C	the set of the c-arcs of $\mathcal{M}$		$\mathcal{I}_k$	the k^{th} IoT of $\mathcal{M}$ corresponding to the group G_k
$\overline{\mathcal{I}}$	the IoT of $\mathcal{M}$ corresponding to the Welcome Zone		$\mathcal{G}_k$	a graph representing $\mathcal{I}_k$
N_k	the set of the nodes of $\mathcal{G}_k$		A_k	the set of the arcs of $\mathcal{G}_k$
σ_c	the score of the node n_c		π_c	the priority of the node n_c
τ_c	the time elapsed since n_c participated to its current group		i_c	the importance of n_c

Table 1: The main abbreviations used throughout this paper

evaluation of the privacy and security flaws, when enabling distributed service discovery in the IoT, is presented in [44].

While all these approaches strive to protect the identity of both *the object* offering a service and the one receiving it, our approach focuses on a different privacy threat. Indeed, although, by adopting the strategies described in this section we could improve the security of object interactions and the protection of service delivery, an attacker can still have access to the basic information about which features and services are available. As explained in the Introduction, also this simple knowledge can lead to disruptive privacy threats as it can be used to infer information about the habit, behavior or status of the corresponding object owners. This is an important application-level privacy flaw that must be considered and faced, and, to the best of our knowledge, our approach is a first attempt in this direction.

3 The proposed model

In this section, we illustrate the model that we adopt to represent and handle the actors operating in our approach. In order to increase the readability of this section and of the next ones, in Table 1, we report the main abbreviations used throughout this paper.

Our model uses the following main concepts:

- *Node.* It represents a smart object and has a profile, which allows its interaction with other nodes in an anonymous way. The profile of a node consists of an identifier, which does not report information about the specific features of the object (in order to guarantee anonymity), and of the set of the features provided by the group it belongs to. A node has also associated all the information needed for the communication with other nodes (such as the MAC address, the IP address, etc.).

  Throughout this paper, we will use the symbols n_i to denote a node and ϕ_i to indicate the set of the features exposed by it.

  Furthermore, since there is a biunivocal correspondence between a smart object and the corresponding node, in the following, we will use these two terms interchangeably.

- *Group.* It is a set of smart objects characterized by heterogeneous features to comply with the

principle of t-closeness. A group has a minimum and a maximum number of nodes. In the following, we will use the symbols:

- G_k, to denote the k^{th} group;
- min_k and max_k, to represent the minimum and the maximum number of nodes of G_k;
- NS_k, to indicate the set of the nodes of G_k;
- Φ_k, to denote the set of the features exposed by G_k.

In turn, NS_k consists of two subsets, namely:

- NS_k^P, i.e., the set of the nodes permanently associated with G_k;
- NS_k^T, i.e., the set of the nodes temporarily assigned to G_k.

- *Welcome Zone* (hereafter, WZ). It is a staging area where nodes are put during their startup phase, when they require to join our system. It can be seen as a special group of nodes in which no feature is exposed. Furthermore, it contains a reference to all the other groups operating in our system.

- *MIoT* (Multi-IoT). It represents the environment where smart objects operate and through which they exchange messages. From a physical viewpoint, a MIoT consists of a network of smart objects that can communicate with each other either directly (if there exists a direct link between them) or indirectly (if there is the need to pass through other intermediate nodes). The network handles two basic kinds of communication, namely:

  - *Point-to-point*: it consists of a private message between two nodes of the MIoT that cannot be accessed by any other node.
  - *Broadcast*: it consists of a public message delivered inside a group or inside the Welcome Zone that can be seen by all the corresponding nodes.

  From a logical viewpoint, according to the model proposed in [5], a MIoT can be modeled as a set of Internets of Things (hereafter, IoTs):

$$\mathcal{M} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_m, \overline{\mathcal{I}}\} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_m, \mathcal{I}_{m+1}\}$$

  Here, each IoT $\mathcal{I}_k$, $1 \leq k \leq m$, corresponds to a group, whereas $\overline{I} = I_{m+1}$ corresponds to the Welcome Zone. A graph $\mathcal{G}_k = \langle N_k, A_k \rangle$, $1 \leq k \leq m+1$, can be associated with each IoT of $\mathcal{M}$. In this case:

  - N_k is the set of the nodes of $\mathcal{G}_k$; there exists a node n_i for each smart object associated with $\mathcal{G}_k$.
  - A_k is the set of the arcs of $\mathcal{G}_k$. Our model assumes that there always exists an arc between two nodes of the same group or between two nodes of the Welcome Zone.

Finally:

$$\mathcal{M} = \langle N, A \rangle$$

Here:

- $N = \bigcup_{k=1}^{m+1} N_k$;
- $A = A_I \cup A_C$, where $A_I = \bigcup_{k=1}^{m+1} A_k$ and $A_C = \{(n_j, n_q) | n_j \in N_k, n_q \in N_l, k \neq l\}$.

A_I is the set of the inner arcs (hereafter, *i-arcs*) of $\mathcal{M}$; they link nodes belonging to the same group. A_C is the set of cross arcs (hereafter, *c-arcs*) of $\mathcal{M}$; they link nodes belonging to different groups and play an important role in our privacy-preserving protocol, as will be clear in the following. A node connected to at least one c-arc is called *c-node*; otherwise, it is called *i-node*. Actually, in our model, we can distinguish two main categories of c-nodes. The former refers to nodes that *temporarily* belong to a group G_k; indeed, just because they are not permanently assigned to G_k, they still continue to belong also to WZ[1]. The latter, instead, comprises nodes that have c-arcs towards nodes belonging to other groups.

As a final point, we observe that, while i-arcs are automatically built by our system once a group is formed, c-arcs are built by nodes. Specifically, c-arcs can be created either to connect a node of the WZ temporarily assigned to a group with the other nodes of this group, or to connect nodes belonging to different groups. Concerning this last aspect, it is worth underlying that, in our solution nodes can still interact with each other by using the classical strategies defined in the IoT literature, such as node proximity or node homogeneity [45].

4 The proposed privacy-preserving object grouping scheme

In this section, we illustrate our object grouping scheme. In particular, in Subsection 4.1, we provide a general overview of the behavior of our scheme. In Subsection 4.2, we describe the node-level operations, whereas, in Subsection 4.3, we present the group-level ones. Finally, in Subsection 4.4, we illustrate the information delivery protocol.

4.1 General overview of the proposed scheme

As stated in the Introduction, the objective of our approach is to protect the privacy of the users of smart objects in a MIoT when feature advertising guides object interactions. As explained in the Introduction, to prevent privacy leakage, our approach borrows some concepts, namely k-anonymity [11] and t-closeness [12], from databases.

In our scenario, we implement these notions by creating groups of objects so that each object can participate to the MIoT by using the features of its group as a business card. Intuitively, any object can be a mean to reach the content available inside a group of objects if they can interact with each

[1]Recall that, in our approach, WZ is modelled as an IoT of $\mathcal{M}$.

other. As a consequence, if all the communications happening inside the group are made anonymous, observers cannot know which nodes of the group can provide content related to a specific features.

Our scheme consists of two main operation categories, namely *Node-level operations* and *Group-level operations*. The former includes the two fundamental actions that a single smart object (i.e., a node in our model) can perform inside the MIoT, namely *join* and *leave*. The latter refers to operations performed by all the nodes of a group to preserve the MIoT liveness. In more details, it consists of the following actions: *Formation of a group*, *Remediation of a group* and *Resize of a group*.

As depicted in Figure 1, each node can enter our system by means of a join operation. Our system is equipped with a staging area, i.e., the Welcome Zone, in which nodes are welcomed. Nodes joining WZ send hello messages to advise other nodes of their presence in WZ.

To satisfy privacy requirements, we impose a minimum number of nodes in the WZ before group formation can start. When this constraint is satisfied (see Section 4.2 for further details), smart objects exchange messages about their features through the information delivery protocol proposed in Section 4.4. This was designed to guarantee the anonymity of the source of each available feature. A group can be formed if, in WZ, objects and their features comply with specific criteria. These are defined by taking both the k-anonymity and the t-closeness paradigms into account.

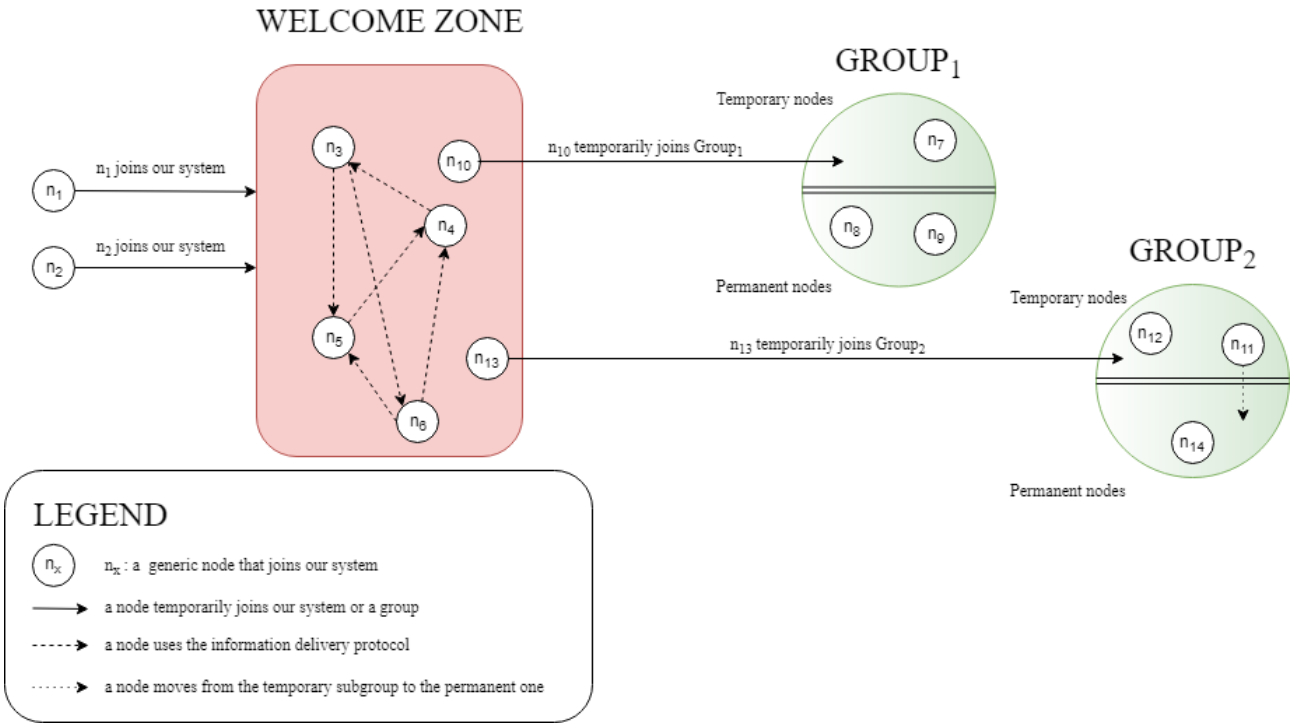

Figure 1: Overview of our approach

Over time, new nodes can register to the system and join (even temporarily) existing groups or take part to the formation of new ones. Furthermore, a node can leave its current group and, eventually, the system. Once again, objects use protocol messages to communicate their intentions (e.g., leaving the current group); in this case, group-level operations (such as the remediation and the resize of a

group) are triggered in response to them. These last operations have been conceived to manage the variation of the number of nodes inside groups over time.

As a final aspect, considering that rising messages with specific features as a subject can also lead to a privacy leakage, our approach provides a querying mechanism allowing for a privacy-preserving retrieval of information in such a complex system. It basically consists of two kinds of message, namely *Intra-group Query* and *Extra-group Query*, and of a communication protocol. Nodes can retrieve information from their group or from the MIoT network. The former task is achieved by using intra-group messages; the latter, instead, adopts special extra-group messages.

It is worth mentioning that group formation is only based on the arrival order of the nodes in WZ. Of course, this implies that a group can potentially contain heterogeneous nodes. However, the nodes of a group share a consistent number of features, because of the requirements of our privacy model. This aspect will be explained in Section 4.3.1. Anyway, the node homogeneity requirement is not crucial in our context; in fact, our objective is different and regards the creation of relatively small blurred clouds of nodes to protect the features exposed by each of them. From a technical point of view, the connections among nodes are handled by the MIoT, which provides the basic networking functionalities (private point-to-point communication and broadcast messages). Whenever a node joins the system, it actually registers its connectivity information (MAC address, IP address, etc.) to the MIoT. An important point is that we need to guarantee the possibility for nodes to directly interact with each other inside the group because we want to map each node to the features exposed in the whole group. For this reason, we impose the full connectivity of the nodes inside each group. Once again, all the communications (and, hence, the use of the corresponding connection links) is handled by the MIoT.

As a final point, group formation is the strategy adopted to implement our privacy model. However, we also preserve the original nature of an IoT by guaranteeing that nodes can still get in touch and interact according to existing strategies and links [18, 46, 47]. Indeed, as explained below, our solution also includes extra-group communication among nodes. Therefore, if two nodes are in proximity and, according to [18], a link can be established between them, two situations may happen, namely: *(i)* they belong to the same group and, hence, no further operation is necessary; *(ii)* they belong to different groups, in which case a *c-arc* will be created between them in such a way as to allow their (extra-group) communication.

In the next subsections, we provide a complete description of our protocol by examining node-level operations, group-level operations, and the delivery protocol in details.

4.2 Node-level operations

Node-level operations specify the tasks that a single node can perform in a MIoT. There are basically two operations, namely *join* and *leave*. We describe them in the next subsections.

4.2.1 Join of a node

A join operation is performed when a node n_i requires to join WZ or a group G_k of the MIoT.

In the former case, n_i sends a "hello message" (see Section 4.4.1) to the other nodes of WZ. These answer it by specifying the number ϵ of the nodes that already joined WZ without having

communicated their features yet. As a matter of fact, in order to preserve the k-anonymity property, it is necessary that at least k new nodes simultaneously communicate their features. To reach this objective, ϵ is increased whenever a node joins WZ. When $\epsilon \geq k$, all the nodes in WZ communicate their features and ϵ is set to 0.

In case n_i joins a group G_k, it is necessary to distinguish two further subcases, namely permanent and temporary joins. The former represents the main form of membership of a node to a group; it is a stable situation in which the node can stay in the group and can participate to all the tasks involving the members of the group without time limitation, and, therefore, until a group no longer exists or the node spontaneously decides to leave the group. The latter, instead, has been conceived to face anomalous situations in which the conditions for the formation of new groups are not satisfied for a long time interval (this generally happens when there is a lack of a sufficient number of new nodes, see Section 4.3.1). In this case, the objects waiting in WZ are temporarily joined to existing groups if the features exposed by them make it possible. In this case, nodes can join groups but with some limitations (mainly related to the features they expose) until new groups tailored to their features can be built (see Section 4.3.2 for details about this operation). Specifically, a node can temporarily join a group if the intersection between the set of its feature and that of the group is not empty. It is worth underlying that, in this case, the node would conceal the additional features it may have with respect to the ones exposed by the group it is joining.

In case of a permanent join, n_i communicates the change of its state to the nodes of WZ so that they can remove it from their lists of contacts. In case of a temporary join, n_i simultaneously belongs to G_k and WZ. Indeed, in this last case, it still interacts with the nodes of WZ in order to create new groups or to participate to the remediation or to the resize tasks involving already existing groups (see Section 4.1). As a consequence, in this case n_i acts as a c-node, as pointed out in Section 3.

4.2.2 Leave of a node

A leave operation is performed when a node n_i requires to leave WZ or a group G_k of the MIoT. In the former case, it is sufficient that n_i informs the other nodes of WZ so that they will remove the arcs linking them to n_i. In the latter case, n_i must inform the nodes of both G_k and WZ, which will remove all the arcs linking them to it.

After this task, the process terminates if n_i is an i-node. On the other hand, i.e. n_i is a c-node, it is necessary to handle the arcs between it and the nodes of the other groups of the MIoT.

For each arc between n_i and a node n_l of another group G_q, two cases might happen:

- *the arc is recent and has been rarely used*; in this case, it can be removed;
- *the arc is old and has been frequently used*; in this case, it should be "inherited" by another node of G_k.

To distinguish these two cases, it is possible to introduce a parameter ρ measuring the relevance of an arc. ρ is defined as $\rho = \frac{\nu}{\lambda}$, where ν is the number of times in which the arc was used for a communication, whereas λ is the lifetime of the arc. If ρ is less than a threshold th_ρ, the arc can be removed; otherwise, it must be "inherited" by another node of G_k.

In this latter case, it is necessary to select the node that inherits the arc. For this purpose, first the set $CSet_k$ of the candidate nodes of G_k is determined. This set comprises all the c-nodes of G_k different from n_i. Then, each node n_c of $CSet_k$ must compute a score σ_c, which takes into account both its priority π_c and the compatibility σ_c between its features and the ones of G_q. Formally speaking:

$$\sigma_c = \omega \cdot \pi_c + (1 - \omega) \cdot J(\phi_c, \Phi_q)$$

Here, ω is a weight, belonging to the real interval $[0, 1]$, used to weigh the importance of priority against compatibility.

The priority π_c of n_c is a real number that takes into account the time τ_c elapsed since n_c participated to G_k and the importance ι_c of n_c in the MIoT:

$$\pi_c = \tau_c \cdot \iota_c$$

The value of ι_c belongs to the real interval $[0, 1]$ and is determined by the human expert in a friendly fashion. For instance, a device measuring a vital parameter (e.g., the heartbeat or the blood glucose) is generally more important than one measuring the brightness. The policy above tends to assign the arcs to the nodes with a higher priority; it aims at minimizing the probability of new re-assignments of the same arc in the future. Indeed, since the priority of a node is computed as a combination of both the time elapsed from the moment it joined G_k and its importance (in terms of offered features), a node with a high priority is less probable to leave G_k.

J is the Jaccard coefficient between the features of n_c and the ones exposed by the group G_q, which n_l belongs to. We recall that the Jaccard coefficient measures the similarity between two sets and returns a value in the real interval $[0, 1]$; the higher this value the higher the similarity [48].

The competition to inherit the arc is initialized by the leaving node. After all the candidate nodes of G_k have determined their score, they anonymously communicate it by using the anonymous broadcast communication of the information delivery protocol described in Section 4.4. Hence, the node with the highest score will be selected to inherit the arc left by n_i. It will inherit this arc in an anonymous way. When this happens, the value of ν, and consequently of ρ, for this arc is reset.

As previously pointed out, when n_i leaves G_k and the MIoT, it must also inform the nodes of WZ. In fact, all the nodes belonging to WZ, or temporarily assigned to other groups, must know all the changes in every group because these changes may activate resize or remediation operations that might involve them.

4.3 Group-level operations

Group level operations indicate those operations that can be carried out by a group in a MIoT. The possible operations are three, namely *Formation*, *Remediation* and *Resize*. We describe them in the next subsections.

4.3.1 Formation of a group

A new group is formed when all the following conditions are verified:

- The number of features currently present in WZ is higher than or equal to k, in such a way as to satisfy the k-anonymity property.
- At least k of these features belong to equivalence classes that satisfy t-closeness. We recall that an equivalence class satisfies t-closeness if the distance between the distribution of a sensitive attribute in this class and the one of the same attribute in the whole data sample is lower than or equal to a threshold t.
- Each of these features is present in at least $\eta > k$ nodes.

In other words, a new group can be formed if there are at least k features with a sufficiently similar distribution in WZ. It is not necessary that each feature is present in the same number of nodes; indeed, it is sufficient that it is present in at least η nodes.

Finally, a group can also have more than k features provided that the additional ones are present in at least η nodes and the sum of their distributions is not higher than the sum of the distributions of the first k features. This condition is justified by the fact that the k features must be characterizing for the group, and this does not happen if there are other ones more present than them therein. As a consequence of the previous reasoning, the number $|NS^P|$ of the permanent nodes of a new group must be higher than or equal to $k \cdot \eta$. There is also a threshold th_{max} for the maximum number of nodes (i.e., for the maximum value of $|NS^P| + |NS^T|$) of the new group. This threshold is linked to the performance of the routing algorithm and to the fact that the graph $\mathcal{G}$ corresponding to the new group is totally connected.

A final parameter that plays a key role in the formation of a new group is the priority π_c of the candidate nodes (see Section 4.2.2). In fact, if there are two or more candidate nodes, our approach selects the one with the highest priority.

Running Example In Figure 2, we illustrate an example of the formation of a new group according to our strategy. Here, we consider a situation in which the WZ contains five nodes, namely $n_1..n_5$, whose features are reported in the legend of Figure 2. For the sake of simplicity, in this example, we set $k = 2$ and $\eta = 2$ and we assume that "energy" and "lighting" are two features belonging to an equivalence class. Therefore, because WZ contains at least 2 nodes with the features above, both the privacy requirements (i.e., $k = 2$ and $\eta = 2$) are satisfied. As a consequence, a new group, namely "$Group_x$", can be formed containing nodes n_1, n_2, n_3, and n_4. The set of features exposed by this group, and therefore by its members, will be: "energy", "ligthing", and "cooling".

It is worth noting that, because the requirement on k is already satisfied by the presence of "energy" and "lighting", the feature "cooling" can be safely exposed as it satisfies the requirement on η.

Of course, n_5 cannot be part of this new group because it does not share any feature with the other nodes. □

4.3.2 Remediation of a group

In case the rate of arrival of new nodes in the MIoT is low, the overall dynamism of the MIoT can be reduced, and some degenerative situations may arise, in which the nodes remain a long time in WZ before being able to join any group. The temporary join of a node to a group has been thought

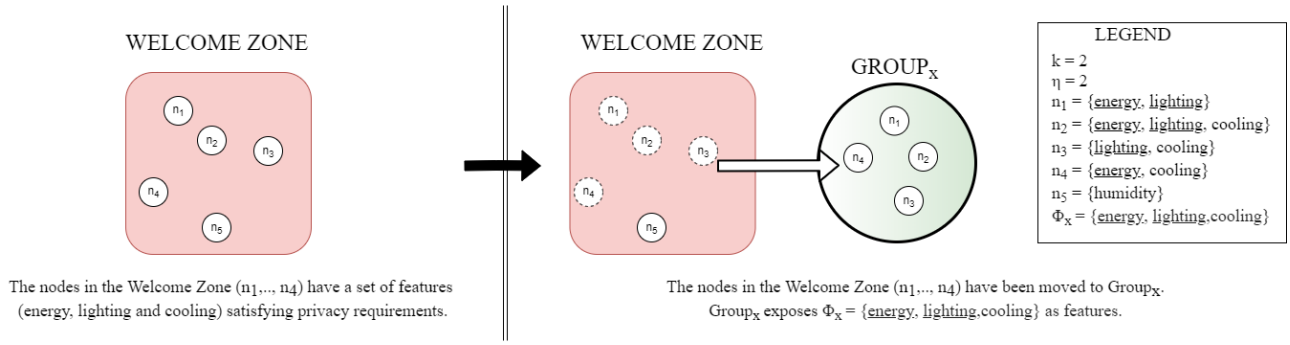

Figure 2: Tasks carried out during the formation of a new group

just to address this issue. As a matter of fact, each group can temporarily accept some nodes (if the overall number of its permanent and temporary nodes is less than th_{max}) provided that their features are already exposed by that group. In any case, WZ keeps track of temporary joins because, if the set of the nodes belonging to it or temporarily assigned to a group satisfies the conditions necessary for the formation of a new group, this last activity is started.

However, in spite of the previous policies, it can happen that, owing to the arrival rate of new nodes in the MIoT, there exists a node n_i whose features are not exposed by any group yet, and, therefore, incapable of participating to the MIoT's life for a long time. To address this issue, our approach provides the remediation operation. It can be activated if there are at least two groups whose number of permanent and temporary nodes is less than th_{max}. Let G_h and G_l be two of these groups. Remediation starts by recalling the nodes of G_h and G_l in WZ. This task aims at constructing two new groups G'_h and G'_l starting from the nodes of G_h and G_l in such a way that one of the new groups can contain n_i[2].

The approach followed by the remediation plan leverages the fact that each node knows only the nodes of its group and, in case it is a c-node, some other ones of different groups.

Now, since in a group there are k characterizing features and each feature is exposed by η nodes ($\eta > k$), our remediation operation can guarantee that a feature exposed by the new node is "hidden" among the ones exposed by at least $(k \cdot \eta) - 1$ existing nodes in the corresponding group. As a consequence, the probability that a node of this group detects the node providing the new feature is less than $\frac{1}{(k \cdot \eta)-1}$ that, in turn, is less than $\frac{1}{k}$. This implies that our remediation operation can guarantee k-anonymity.

Running Example (continued) Figure 3 shows a possible evolution of the previous example. Now, a new node, say n_7 is entering the WZ already containing nodes n_5 and n_6. Once again, the features of all nodes are reported in the legend of Figure 3. In this situation, a new group cannot be created as features of nodes in the WZ do not satisfy privacy requirements. However, while n_5 and n_6 do not share any feature with existing groups, n_7 has the feature "lighting" already exposed by "$Group_x$". Therefore, according to the remediation operation, n_7 could safely joins "$Group_x$" provided that it conceals the feature "alarm" not exposed by this group. □

[2]Clearly, it is not sure that the features of n_i allow it to be a member of G'_h or G'_l. If this does not happen, n_i will

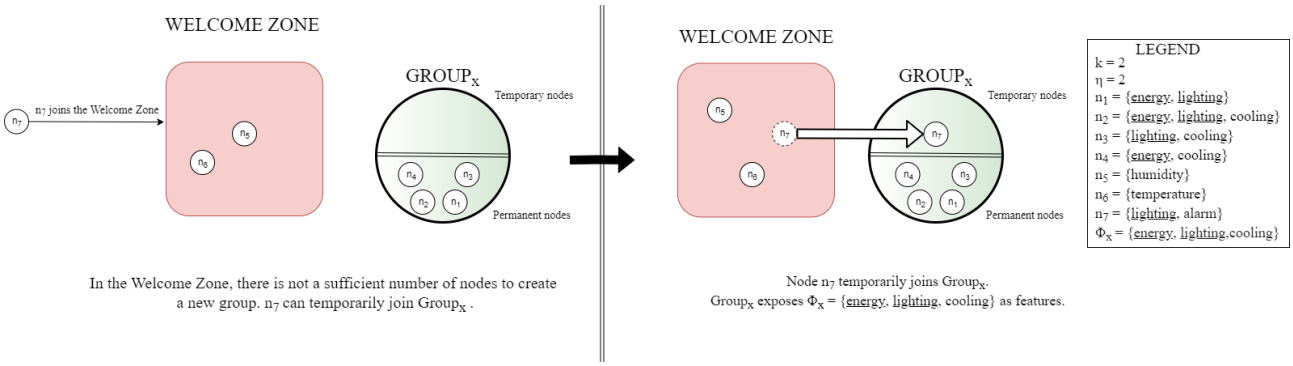

Figure 3: Tasks performed during the remediation of a group

4.3.3 Resize of a group

A group resize operation is activated after that k permanent nodes performed a leave operation in a group. Waiting for k leave operations before carrying out this task is necessary to guarantee k-anonymity. In fact, we can reconstruct one or more features of a node leaving the group if we verify the corresponding impact on the set of features after each node leaves, at least in some cases. By contrast, waiting for k leave operations before verifying the features of a group allows our approach to guarantee that the possible impacts can be associated with k different nodes and, then, that k-anonymity is preserved.

When the resize of a group G_k starts, two different cases are possible, namely:

- all the features previously exposed by G_k are still present, but for at least one of them k-anonymity is not guaranteed;
- at least one feature previously exposed by G_k is no longer present and t-closeness is not guaranteed; the other features may or may not guarantee k-anonymity.

If one of the previous conditions is true, it is necessary to start a group restore task. Given a feature φ that does not currently guarantee k-anonymity, the resize task tries to perform one of the following countermeasures:

- C_1: if G_k contains a temporary node that exposes φ, then it is added to G_k as a permanent node.
- C_2: if G_k contains no node that exposes φ, but a suitable node is present in WZ, then it is added to G_k as a permanent node.
- C_3: if neither a temporary node in G_k nor a node in WZ exposes φ, but at least another group contains a temporary node exposing this feature, then this node is added to G_k as a permanent node. If more than one node exposing φ exists in the MIoT, then one with the minimum priority is chosen to be added to G_k. This is justified by considering that priority depends on the time

remain in WZ.

a node elapsed in the group and on its importance. Removing from a group G_l a node with a high priority (even if it has been assigned to G_l only temporarily) could imply removing from G_l a node important for it and/or a node that spent a certain amount of time in this group. This last condition could have led this node to construct several links and relationships that are broken if it is forced to change its group.

Of course, the operations described above are carried out for all the features that are not currently guaranteeing k-anonymity in such a way as to preserve node privacy. Actually, the verification of a group G_k is performed as a challenge between nodes permanent in G_k and external nodes. Analogously to what happens for group formation, the permanent nodes of G_k start by anonymously communicating their features to WZ. The other nodes that are listening to WZ (i.e., those nodes not assigned to a group yet, or those nodes temporarily assigned to a group) participate to the challenge by adding their features (still leveraging the anonymous broadcast) until G_k satisfies the privacy requirements again.

By following the algorithm above, in the resize of G_k, its temporary nodes are preferred to the free nodes of WZ that, in turn, are preferred to the temporary nodes of other groups. Each node independently estimates its contribution to G_k; in this task, it considers the priority of its category as a key aspect. Finally, if more suitable nodes exist in the same category, a priority-based approach, similar to the one discussed in Section 4.2.2, is adopted to select the one to be added to G_k.

This task terminates when:

- G_k exposes a set of features that guarantees both k-anonymity and t-closeness;
- G_k is in one of the two cases that do not guarantee k-anonymity and/or t-closeness and there exists at least one feature of G_k for which no countermeasure can be applied.

In the former case, G_k is restored, whereas, in the latter case, it must be dissolved, and the corresponding nodes must be re-assigned to WZ. Observe that these nodes will remain in WZ only until either G_k can be fully restored or they can join (even temporarily) another group G_l, such that the set $\Phi_{int} = \Phi_k \cap \Phi_l$ contains at least k features that belong to equivalence classes satisfying t-closeness.

Running Example (continued) In Figure 4, we illustrate another possible evolution of our running example. In this case, nodes n_1 and n_4 are leaving the system so that "$Group_x$" no longer satisfies privacy constraints on feature "energy" ($\eta < 2$). Observe that, this feature also contributed to comply with the requirement on k (i.e., $k = 2$) as it belonged to an equivalence class together with the feature "lighting". In this case, the resize operation has to be executed for "$Group_x$".

According to Case A of Figure 4, the node n_8 is available in the WZ and because it has the feature "energy", it can safely join "$Group_x$". In this way, the privacy requirements for this group are restored and, hence, the group can remain alive.

In Case B, instead, no node, with the needed features, is available to join "$Group_x$". In this scenario, this group can no longer exist. Therefore, its nodes leave it to join WZ once again. □

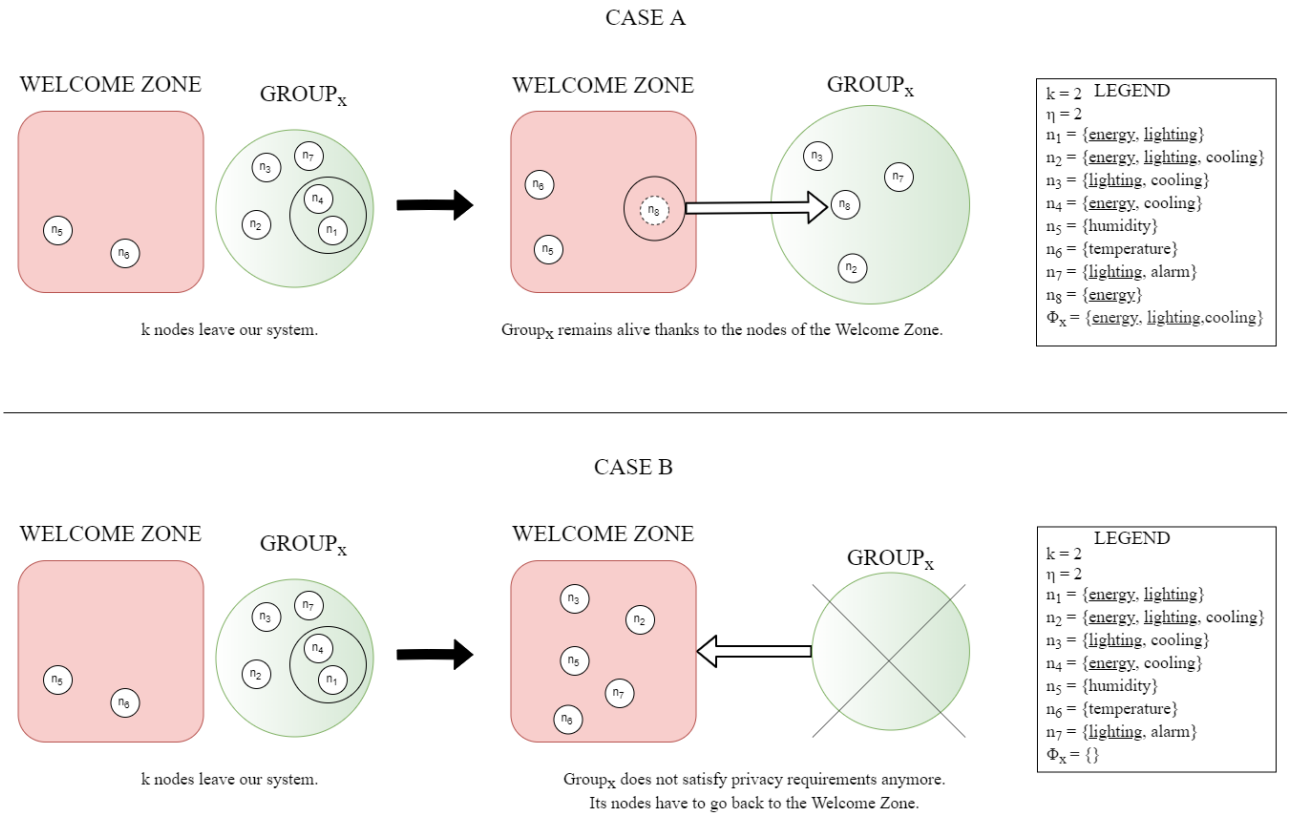

Figure 4: Tasks performed during the resize of a group

4.4 Information delivery protocol

Our information delivery protocol is based on three kinds of message, namely point-to-point, broadcast and anonymous broadcast. The first two are directly derived from the corresponding functionalities provided by the network underlying the MIoT (see Section 3). Instead, the third is based on a combination of the first two; it will be illustrated below.

The objective of anonymous broadcast is the implementation of a mechanism to anonymously deliver a message to all the nodes of a group or of WZ. Actually, anonymous broadcast can be seen as a hybrid approach consisting of a preliminary set of point-to-point exchanges of the message to deliver, handled in a way analogous to what happens in mix-net networks [49, 50], followed by a broadcast delivery of the same message.

There are several techniques to implement this strategy (see, for instance, [49, 51, 52, 53, 54, 55, 56]). A naive (but, at the same time, efficient and effective) way of proceeding is as follows. When a node n_i receives a message m, it forwards m to another node n_j with a given probability p by using the point-to-point mode. Instead, with a probability equal to $1 - p$, it forwards m in broadcast mode to all the nodes of its group (or to WZ). The value of p must be chosen to guarantee a trade-off between the need to quickly deliver m to all the nodes of the group (in such a way as to avoid that m becomes obsolete) and the need to preserve privacy. When m is received in broadcast mode by a node n_i of a

group, if n_i has arcs towards nodes of other groups that expose features characterizing m, it can use these arcs to deliver m to the corresponding groups in a point-to-point mode.

After having illustrated the three possible message modes, we now examine the possible message types provided by our information delivery protocol. They can be grouped in three categories, namely *join*, *leave* and *query*. We illustrate all of them in the following subsections.

4.4.1 Join Messages

The messages belonging to this category are the following:

- *WZ Hello*. This message has the form ⟨'`Hello`', '`WZ`'⟩. It is sent in broadcast mode by a node n to WZ when n requires to join the MIoT.
- *WZ Answer*. This message has the form ⟨'`Welcome`', $\epsilon + 1$⟩. It is sent in broadcast mode by WZ as an answer to the corresponding *WZ Hello* message previously sent by a new node n to WZ. $\epsilon + 1$ is an integer denoting the number of nodes (including n) present in WZ after the join of n.
- *Temporary Hello*. This message has the form ⟨'`Hello`', '`T`'⟩. It is sent in broadcast mode by a node n to a group G when n requires to temporarily join G.
- *Permanent Hello*. This message has the form ⟨'`Hello`', '`P`'⟩. It is sent in broadcast mode by a node n to a group G when n requires to permanently join G.
- *Feature Set*. This message has the form ⟨'`Feature Set`', ϕ⟩. It is sent in anonymous broadcast mode by a node n to the nodes of WZ. ϕ denotes the set of the features exposed by n. In order to preserve the privacy of n, this message can be sent when at least $\epsilon \geq \eta$ nodes are present in WZ. It represents the first step for the formation of a group.

4.4.2 Leave Messages

The messages belonging to this category are the following:

- *Temporary Leave*. This message has the form ⟨'`Bye`', '`T`'⟩. It is sent in broadcast mode by a node n, which has been temporarily assigned to a group G, when it decides to leave G. When this happens, n is assigned to WZ.
- *Permanent Leave*. This message has the form ⟨'`Bye`', '`P`'⟩. It is sent in broadcast mode by a node n, which has been permanently assigned to a group G, when it decides to leave G. When this happens, n is assigned to WZ.
- *WZ Leave*. This message has the form ⟨'`Bye`', '`WZ`'⟩. It is sent in broadcast mode by a node n, which is assigned to WZ, when n decides to leave WZ and, consequently, the MIoT.
- *Score Communication*. This message has the form ⟨'`Score`', '`Sc`'⟩. It is an anonymous broadcast message sent by each node during a challenge for selecting a candidate to participate to a group or to inherit an arc (see Section 4.2.2).

4.4.3 Query Messages

The messages belonging to this category are used by a node when it requires a certain feature. They are the following:

- *Intra-group Query.* This message has the form ⟨'`Intra Query`', ⟨`content`⟩, φ⟩. Here, ⟨`content`⟩[3] denotes the message payload, whereas φ represents the feature the message refers to. It is delivered in anonymous broadcast mode by a node n to the nodes of its group.

- *Extra-group Query.* This message has the form ⟨'`Extra Query`', ⟨`content`⟩, φ⟩. Here, ⟨`content`⟩ denotes the message payload, whereas φ represents the feature the message refers to. It is delivered in anonymous broadcast mode by a node n to the nodes of its group G. If G contains any c-node toward another group G', whose features match those in φ, then the c-node delivers the message to its contact in G'. However, if, in turn, G' has c-nodes, the message is not further delivered to other groups. This choice has been made to avoid the traffic overloading in the network underlying the MIoT.

5 Security Model

In this section, we describe our security model (Subsection 5.1) and analyze the corresponding properties (Subsection 5.2).

5.1 Attack Model

As a preliminary assumption, we consider a realistic situation in which a sufficient number of nodes is available so that our approach can be implemented successfully. Therefore, we will not consider anomalous situations, in which the number of the nodes available in the system is less than the minimum number necessary to guarantee, at least in principle, privacy (i.e., $k \cdot \eta$).

Furthermore, our approach focuses on the protection of node information and does not deal with attacks on the protocol, such as sinkhole or DoS attacks [57, 58]. Indeed, these threats are common for most of the communication protocols and the strategies for preventing them are orthogonal to our proposal. In our case, it is possible to adopt any of these strategies, such as the ones presented in [59, 60, 61], in such a way as to make our approach robust also to these kinds of attack.

Given this basic assumption, we now identify the security properties of our approach. They are:

- *SP1* - The definition of the groups' features ensures the privacy of nodes.

- *SP2* - Our approach is resistant to attacks exploiting group resize operation.

- *SP3* - Our approach is resistant to timing attacks exploiting cross-feature interview.

- *SP4* - The jeopardizing of the routing protocol does not have impact on the privacy of nodes.

[3] Observe that no constraint is put on the content to handle, in such a way as to guarantee data confidentiality and integrity.

- *SP5* - Our anonymous broadcast delivery protocol is resistant to classical attacks (e.g., the timing and the routing ones).
- *SP6* - Our approach is resistant to attacks based on historical data concerning join and leave operations.

In the analysis of the security properties described above we will consider the following assumptions:

- *A1* - An attacker cannot control a whole group of nodes.
- *A2* - The underlying network provider is not interested in violating node privacy.
- *A3* - The basic features delivered by the MIoT system (point-to-point communication, etc.) are robust to attacks.
- *A4* - All the features considered in our approach are not related to geographic positions.
- *A5* - At most t nodes can collude to break the security properties of our protocol.
- *A6* - The attacker has no additional knowledge derived from any direct physical access to nodes.

In the following, we will investigate the security properties mentioned above. To perform this analysis, we needed a reference scenario. To model it, and to test our approach, we constructed a prototype. Furthermore, as real MIoTs with the size and the variety handled by our model do not exist yet, we constructed a MIoT simulator.

To make "concrete" and "plausible" the simulated MIoTs, we had the necessity that our simulator was capable of returning MIoTs having the characteristics specified by the user and being as close as possible to real-world scenarios. In the simulator design, and in the next construction of the MIoTs to use for the experiments, we followed the ideas expressed in [18, 46, 47], in which the authors highlight that one of the main factors used to build links in an IoT is node proximity. In order to reproduce the creation of links among objects, we decided to leverage information about real-life paths in a city. In fact, having this information at disposal, we may associate each path with an object and link two objects if their paths have been near enough for a sufficient time period. As for a dataset containing real-life paths in a city, we selected the one reported in `http://www.geolink.pt/ecmlpkdd2015-challenge/dataset.html`. It regards taxi routes in the city of Porto from July 1^{st} 2013 to June 30^{th} 2014. Each route contains several Points of Interests corresponding to the GPS coordinates of the vehicle. As said above, our simulator associates an object with a given route recorded in the dataset. Furthermore, it creates an arc between two nodes if the distance between the corresponding routes is less than a certain threshold th_d for a predefined time interval th_t. The value of th_d and th_t can be specified through the constructor interface. Clearly, the higher this value the more connected the constructed IoT. The interested reader can find the IoTs created in this phase at the address `http://daisy.dii.univpm.it/miot/datasets/privacy`.

Regarding the MIoT construction, since group creation depends on the sequence of subscriptions of the nodes to our system (which, for the sake of simplicity, can be assumed as random) and on their features, we reproduced it by simulations, as will be clear in the following. When we defined

the distribution of the features among the nodes, we leveraged scientific literature and used the corresponding results to properly tune our simulator. In particular, we adopted the values reported in [62].

Some statistics about our dataset are reported in Table 2.

Parameter	*Value*
Number of nodes	1000
Number of relationships	6860
Mean outdegree	6.995
Mean indegree	7.002
Number of distinct features	20
Maximum number of features for an equivalence class	10
Maximum number of features for a node	3

Table 2: Parameter values for our simulator

5.2 Security Analysis

In this section, we focus on each of the security properties introduced above and analyze if and how our approach can guarantee them.

5.2.1 SP1 - The definition of the groups' features ensures the privacy of nodes

This property is fundamental in our approach because it guarantees that, inside a group, nodes are protected against attacks to their privacy. Our approach uses a combination of k-anonymity and t-closeness to ensure this property. Indeed, k-anonymity alone fails because, in real life, features are not uniformly distributed among smart objects. Therefore, an attacker, near a node, may take advantage of the probability distribution function to perform a statistical attack and to improve the guessing probability.

For this reason, our algorithm takes into account the distributions of the features that characterize a new group when it selects k features. In accordance to the t-closeness paradigm, the characterizing features of a group must belong to an equivalence class when it comes to their probability distribution. This ensures that an attacker cannot exploit the background knowledge on the popularity of features among smart objects in such a way as to exclude the least probable ones, thus increasing the probability of mapping a feature to an object.

Furthermore, as for group formation, our protocol exploits, once again, the notion of k-anonymity to allow nodes to freely exchange information about features without being identified. Indeed, each node inside WZ waits until $\epsilon > k$ nodes are available before adopting the anonymous broadcast protocol to communicate its features. Now, in absence of collusion attacks, ϵ can be set to k. In this way, an attacker can only observe that there are some features among those k nodes, without having further advantages in mapping them to the right objects. Moreover, in this case, t-closeness is not needed because the attacker is dealing with a set of k nodes each having exactly the same probability to own the specified properties. As a final observation, in accordance to Assumptions *A1* and *A5*, an attacker can only control t nodes simultaneously. Therefore, to block a collusion attack, it is possible to set $\epsilon = k + t$ in such a way as to preserve the k-anonymity property.

5.2.2 SP2 - Our approach is resistant to attacks exploiting group resize operation

The aim of this property is to protect our system from attacks based on the observations of resize operations. Indeed, during each resize operation, the structure of groups may change in terms of both the number of involved nodes and, possibly, the number of available features. An attacker can evaluate the features proposed by a group by either interacting in proximity with a node of that group or by being a member of the group itself.

Our approach adopts two countermeasures to this kind of attack. The former consists in forcing the resize algorithm in such a way that it can be activated only when k leave operations have been recorded. Due to Assumption *A1*, the attacker cannot control a group and, hence, cannot control which nodes leave the system and when it happens. Moreover, as a further security measure, we require that, for each feature, there are at least η nodes owning it. The combination of these countermeasures inhibits the attacker from detecting which feature was owned by the leaving nodes (the probability of guessing it will be the same as the one of guessing the features of any other node in the group). In this way, our approach prevents the attacker from being able to detect a reduction of the number of the available features included in the group.

5.2.3 SP3 - Our approach is resistant to timing attacks exploiting cross-feature interview

A common attack typical of scenarios similar to the one proposed in this paper is based on the statistical observation of the response time of nodes to external events. In our case, this attack can be executed by querying a node about information related to a predefined set of features and by comparing response times. Fast answers can be associated with features owned by the node, whereas slow answers (or empty ones) can be mapped to features owned by other nodes of the same group that the attacked node must contact to provide its answer.

To prevent this kind of attack, each node adopts a pattern recognition algorithm and enters a protection mode each time it recognizes a suspect querying pattern. Basically, whenever a target node receives a suspect sequence of consecutive cross-feature queries from a source node, say n_a, it starts by adding a random delay in its answers to n_a. This delay ranges from 0 to the maximum answering time detected by it in any previous communications[4]. Furthermore, if the node is not able to answer two consecutive cross-feature queries, it will stop answering any next query from n_a for a certain time interval.

These two countermeasures, when combined with Assumption *A4*, prevent the attacker from gaining advantages by maliciously interviewing any node of our system. Indeed, Assumption *A4* states that the attacker cannot leverage information about specific geographic positions (for instance, to isolate a small set of devices) when she formulates her queries. Without this assumption, an attacker can construct, and then submit, queries whose answer can be provided only by devices located in a specific geographic position. Of course, this is a local attack that, in order to have success, requires a contemporary physical attack allowing the malicious user to isolate a small set of devices to detect the features owned by them. For this reason, we have assumed that geolocalized features are out of the scope of this paper.

[4]Observe that no countermeasure is adopted in case of consecutive queries referring to the same feature. Indeed, in this case, it can be assumed as a normal interaction between two nodes.

5.2.4 SP4 - The jeopardizing of the routing protocol does not have impact on the privacy of nodes

This property guarantees that an attacker cannot gather information about the properties of nodes by tampering the communication protocol. Indeed, she can try to force any communication of a group to pass through it. Although this cannot be achieved for intra-group communications, because the corresponding path is randomly chosen by the nodes inside a group, it can be tried for inter-group communications. Indeed, an attacker may tamper the protocol during the leave of nodes and may promote itself as the node with the highest score, in such a way as to inherit all the arcs towards other groups. This is a variant of the sinkhole attack. The result is that the group will be potentially isolated and its nodes cannot use external arcs without involving the attacker.

Of course, this is an unwanted situation, which can cause issues to the communication protocol. However, no harm is done to nodes' privacy, as each node will still continue to communicate with each other leveraging the anonymous delivery protocol described in Section 4.4. Therefore, even though the attacker may force itself in the middle of all the communications towards external groups, it cannot reveal any information about the nodes being the sources of these communications.

As stated above, our approach does not directly deal with sinkhole attacks when it comes to damages to the communication protocol. Actually, the adoption of well-known countermeasures for these attacks proposed in the scientific literature (such as the ones described in [59, 60, 61]) can help preventing them.

5.2.5 SP5 - Our anonymous broadcast delivery protocol is resistant to classical attacks

This property aims at guaranteeing the robustness of the anonymous broadcast delivery protocol described in Section 4.4. First, observe that, thanks to Assumption *A3*, the basic communication functionalities, such as the private point-to-point communication mechanism among nodes, are assumed to be robust against attacks. Therefore, the anonymous delivery protocol can be built on top of these basic features by directly adapting any anonymous broadcast communication protocol proposed in the scientific literature, whose security has been already proved [49, 51, 52, 53, 54, 55, 56].

Having said these premises, let us consider the naive method to address this goal already described in Section 4.4. To achieve an anonymous broadcast delivery, this approach leverages a random sequence of private point-to-point messages among nodes to obfuscate the source of a message before broadcasting it. This strategy somehow resembles the one adopted in mix-net solutions, whose security level and possible flaws are investigated in [49]. However, because of its simplicity, this approach can be effective and efficient in low-severity scenarios, in which more advanced solutions, like the ones mentioned above, are not necessary.

Due to Assumption *A3*, an attacker cannot have access to point-to-point messages exchanged between generic pairs of nodes. To guess the original source, she can only observe broadcast messages and the point-to-point ones sent to her. As each node sends a message to another one in a point-to-point fashion with a probability p and the same message in broadcast with a probability $1 - p$, the length of the communication path will be strongly variable and unpredictable a priori. Furthermore, the next node in the communication path will be chosen randomly and there is no limit to the path length. If all these features are combined with Assumption *A1*, it is possible to conclude that our

approach prevents an attacker from being able to trace back the message source and, ultimately, from having advantages in guessing its features.

5.2.6 SP6 - Our approach is resistant to attacks based on historical data concerning join and leave operations

This property aims at guaranteeing the robustness of our approach against attacks exploiting the knowledge of historical data, which examine join and leave operations from groups to disclose the features of an object.

Although nodes can freely join and leave groups, re-join operations involving different groups are, in principle, insidious. Indeed, in this case, nodes can drastically change the exposed set of features. This would allow an attacker to intersect the previously exposed features with the currently exposed ones to determine the real subset of them owned by the attacked node.

However, in our approach, a re-join task only happens when a node leaves a group and joins another one during the resize operations (see Section 4.3.3 for all details). In any case, this issue is addressed by the condition specified in Section 4.3.3 according to which a node can re-join the same group it belonged to (even temporarily) in the past or a new one if the intersection of the features exposed by the two groups contains at least k features belonging to equivalence classes that satisfy t-closeness. This countermeasure, along with Assumptions *A1* and *A6*, contrasts this kind of attack.

Another situation to be investigated regards the case in which a node permanently leaves the MIoT (and not simply a group) and, then, re-joins it. Also in this case, historical data can lead to advantages for an attacker. Actually, this issue is not directly considered by our approach. However, a simple protection strategy can be adopted to address it. Indeed, it is sufficient to require that the nodes, which re-join a MIoT after a permanent leave, should restore information about the last group they belonged to during the previous interaction with it. In this way, it is possible to apply the countermeasures for the other re-join situation described above.

6 Solving the trade-off between privacy requirement and network performance

In this section, we aim at investigating the configuration of the privacy parameters, namely k and η, in such a way as to achieve the desired privacy level. Indeed, the more severe privacy requirements, the greater the impact on the network performances.

According to our protocol, a more demanding privacy requirement leads to an increase of the group size. The communication among nodes is influenced by both the presence of groups and the anonymous broadcast protocol, which requires the involvement of a random number of nodes inside each group before reaching the desired destination. As a consequence, both intra-group and inter-group communications are strongly dependent on the group size; specifically, the bigger the groups the higher the number of involved nodes. This has two direct implications on the network performance: *(i)* the overall load of the network increases; *(ii)* the average length of the paths among nodes grows (leading to higher average communication delays). For this reason, a first experiment is devoted to simulate the behavior of our system and to monitor the creation of groups.

The metrics we adopted for this investigation are: *(i)* the variation of the group size against different privacy settings (i.e., different configurations of k and η); *(ii)* the variation of the length of the communication paths among nodes after the application of our privacy model.

For simulation, we considered different values of both k and η. Specifically, as for k, we selected the range $[2, 8]$, with a step of 1; as for η, instead, we considered a multiple of k; in particular, its range was $[k, 2k]$.

As a first investigation, we measured the metric *(i)*. For this purpose, we simulated a random subscription to our system (i.e., a random arrival order in the Welcome Zone) of the 1000 nodes of the original IoT graph considered in this experiment. We applied our algorithm for group formation and measured the average number of nodes inside each group, as well as the average number of nodes not involved in a group and, hence, waiting in WZ. In this experiment, we did not consider temporary joins that can be adopted to minimize the number of nodes not assigned (either temporarily or permanently) to any group.

To consider different configurations of node arrivals, we repeated the experiment 250 times and averaged the corresponding results. In Figure 5, we report the average percentage of all the nodes of the MIoT that are present in a group against the increase of k and η. Instead, Figure 6 shows the average percentage of all the nodes of the MIoT that remain in WZ against the increase of k and η.

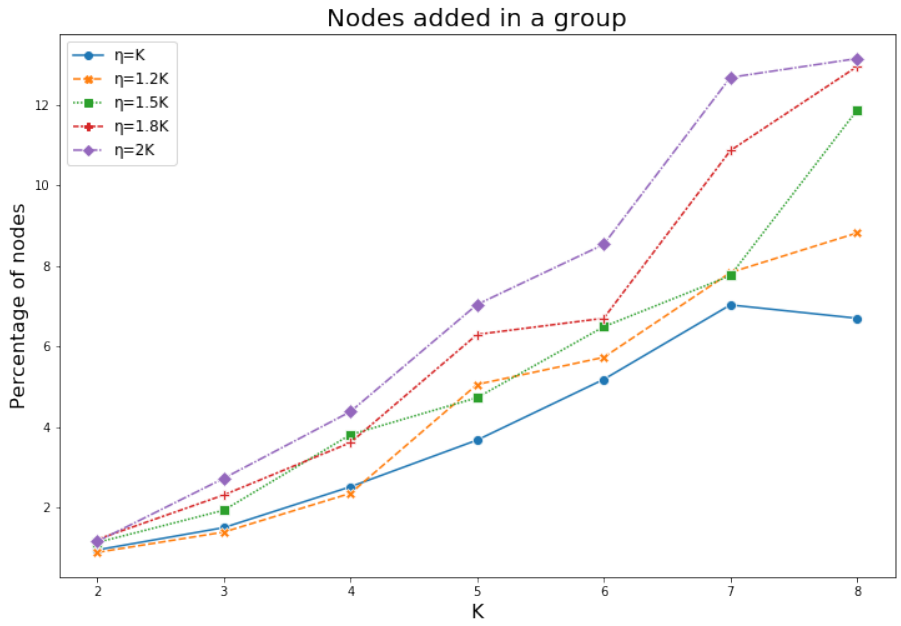

Figure 5: Percentage of nodes present in a given group against the increase of k and η

By analyzing the obtained results, we can observe that the percentage of nodes in a group grows linearly with the increase of both k and η. Interestingly, even with the most demanding privacy requirement (i.e., $k = 8$ and $\eta = 2 \cdot k$), it does not exceed 12.5% of the whole set of nodes. Of course, as proved in [63], higher values of k do not provide additional benefits, once the desired privacy requirement has been reached. With regard to this reasoning, we point out that there is no best practice in the estimation of the right value of k. Typical values adopted in the literature range from

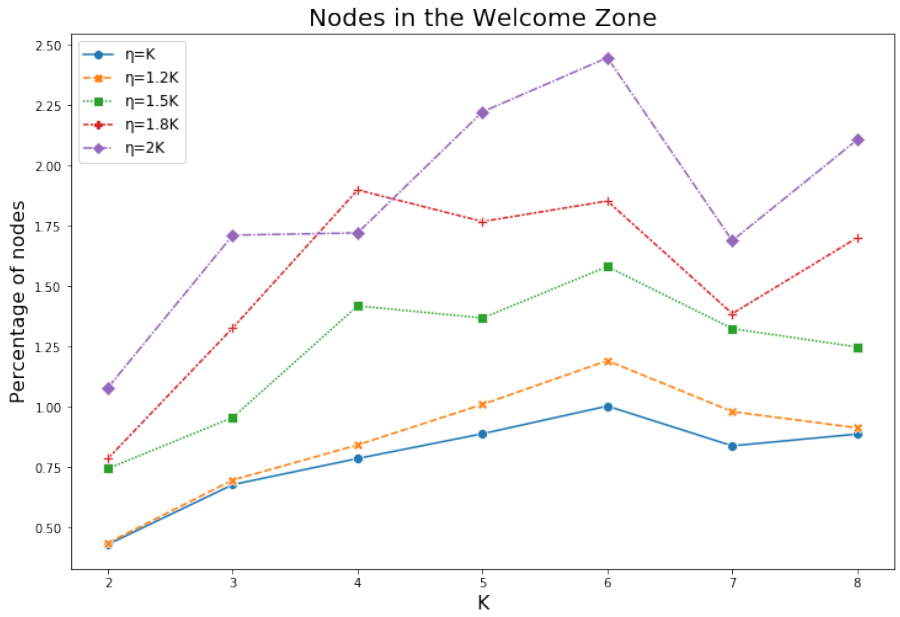

Figure 6: Percentage of nodes waiting in the Welcome Zone against the increase of k and η

2 to 5. As for η, this is a security mechanism introduced to maintain the full operation of a group also in presence of node leaves. However, since our approach for group resize is executed each time k permanent nodes leave a group, to preserve its robustness, we need to have the k-anonymity property guaranteed in the interval from the leave of the first node to the leave of the k^{th} one (after which the group size will be fixed by our approach). At a first analysis, we may affirm that, if η is equal to $2 \cdot k$, no issues will arise before the resize procedure will be executed. This setting is the most preserving one but, as a contrast, it requires a very high number of nodes for each feature. However, if we consider a limit case in which all the leave operations involve nodes owning only one of the available features without repetition, we could safely set $\eta = k + 1$ to ensure the k-anonymity property and the operability of the group during leave operations. These considerations are crucial to properly tune η. Indeed, we can conclude that its right value should range from $k + 1$ to $2 \cdot k$.

As a further observation, keeping η to the minimum values strongly reduces the number of nodes still waiting in WZ after the formation of groups. Indeed, if we set $k = 4$ and $k < \eta = 1.2 \cdot k$, the average percentage of nodes waiting in WZ after the execution of the algorithm for the formation of groups is about 0.08%. Also the number of nodes in each group is low and equals to 2.2% of the nodes of the original graph on average.

The second experiment aims at measuring the metric *(ii)*. To perform this measurement, we applied the same logic adopted in the previous experiment to simulate the formation of groups, but we preserved the original links in the graph built from our dataset for inter-group connections. Observe that this choice is compliant to what should happen in a real life scenario because inter-group connections rise in accordance to proximity events among nodes belonging to different groups, which is exactly how links have been established in the original IoT graph. Now, given a pair of nodes (n_i, n_j) such that $n_i \in G_i$, $n_j \in G_j$, $G_i \neq G_j$ and there exists a path from n_i to n_j in the original graph,

Figure 7 reports the ratio of the length of the path between n_i and n_j in our system to the length of the path between the same nodes in the original graph. We call "Cost of the Protocol" (hereafter, CoP) this parameter. The values reported in this figure are averaged on 1000 pairs of nodes satisfying the requirements above.

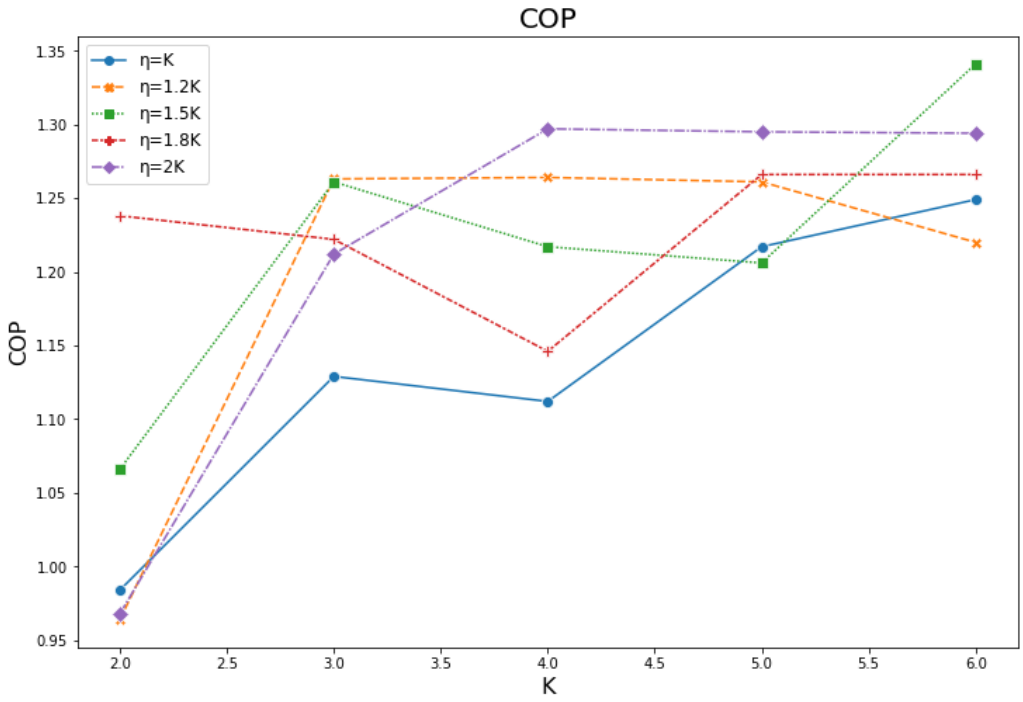

Figure 7: Value of CoP against the increase of k and η

The obtained results show that, if we keep $k \leq 4$ and $\eta = 1.2 \cdot k$, CoP reaches a maximum value of 1.263, meaning that the length of the path among the pairs of nodes obtained by applying our approach increases to a maximum of about 26% with respect to the length of the original path.

7 Discussion

In this section, first we propose a discussion on some additional properties of our approach. Then, we give further insights on its applicability and limitations, at least for an important aspect. Finally, we discuss about the possible comparison between our approach and other ones proposed in the past literature.

7.1 Privacy features

We start by analyzing the two features adopted in this paper, namely: *(i)* k-anonimity, and *(ii)* t-closeness. k-anonimity is a very old notion that, in principle, can avoid information disclosure in a database as long as sufficiently "noisy" tables (i.e., tables guaranteeing k collisions) can be generated [64]. However, it was also proved that, when dealing with value distributions of attributes, an attacker

can take advantages by comparing the distribution in the noisy dataset with the real-world attribute distribution to bypass such a privacy mechanism [65]. Therefore, even if k-anonimity can protect against identity disclosure, it cannot protect against attacks based on attribute disclosure. In this last case, an attacker can leverage the disclosure of the value of a confidential attribute associated with an external identified individual to violate k-anonimity features. In real-life scenarios, the risk of such an attack is very high and, therefore, the only application of k-anonimity appears inadequate for our privacy objectives.

t-closeness was widely studied in the scientific literature [12]. It was conceived as an evolution of k-anonimity that also protects against attribute disclosure. The scenario of interest for this paper is very close to the ones t-closeness was designed for. Indeed, our aim is concealing the features (or attributes) of an object behind a group of heterogeneous and equivalent ones (in terms of probability distributions). For this reason, in our approach, we leverage t-closeness to enhance k-anonimity with the capability of protecting against attribute disclosure, assuming that object attributes (or features, in our case) have specific and measurable distributions.

Interestingly, our solution also recalls the concept of ϵ-differential privacy [66]. This kind of privacy solution aims at limiting the knowledge gain between datasets that differ in one individual. It originally focused on the protection of the outcomes of queries performed in a database. Then, other papers extended this concept to non-interactive scenarios (i.e., cases in which it is not necessary to protect a specific query or set of queries). These solutions often deal with specific classes of generic queries (typically, count ones) [67, 68]. Interestingly, it was proved that t-closeness and ϵ-differential privacy are somehow related to each other [69]. Indeed, the authors of [70] proved that, in a dataset in which t-closeness holds, differential privacy is guaranteed on the projection over the confidential attributes.

7.2 Applicability and limitations

As for the applicability of our proposal to real-world scenarios, we highlight that our strategy is in-line with the new trend of improving the independence of nodes in an IoT. Specifically, several papers focused on the definition of approaches aiming at identifying links between objects with a reduced human intervention [41, 39]. Other papers, instead, focused on the definition of models to uniformly handle data coming from heterogeneous smart objects [40]. Our solution finds a direct application in this context because the knowledge of the features characterizing objects and the services provided by them is fundamental for improving the efficiency of links in an IoT. For this purpose, it is important to filter the contacts of an object according to the usefulness of the information that these contacts can provide. Of course, as stated throughout this paper, the knowledge of the features of an object has serious impacts on the privacy of its user.

Clearly, due to the extremely high dynamics of the considered scenario, our approach has some limitations that must be taken into account. Indeed, as stated in Assumption *A4* , our solution does not cover the protection of features related to specific geographic positions. Indeed, without this assumption, it is not possible to guarantee the security property *SP3*. To clarify this concept, consider the case in which an attacker can isolate a node in a specific location. Furthermore, assume that some of the exposed features can be related to the object position; think, for instance, of the temperature of a room. In this case, the attacker can evaluate whether the node is capable of correctly answering

a query about the temperature of the zone controlled by it. Either a positive or a negative answer results in a privacy leakage, as the attacker is able to identify one of the features of the object for reducing the admissible set. In addition to Assumption *A4* , this security property also requires a pattern recognition solution to detect anomalous cross-feature interviews. Of course, a naive and very conservative solution can be obtained by forcing each node to label as suspect (and, hence, to apply the countermeasure described in Section 5.2.3 to it) each direct interaction with a node that submits queries related to more than two features. A more sophisticated and refined solution can be obtained by adopting any existing approach for anomalous pattern recognition [71]; however, it requires a base knowledge to model the normal behavior of nodes.

7.3 Comparison with other approaches

As pointed out in the Introduction, to the best of our knowledge, our approach is the first one conceived to prevent feature disclosure in a multiple IoT scenario. As a consequence, a direct comparison between our approach and a strictly related one is not possible. Nevertheless, it is possible to perform an "indirect" comparison with another approach which, even if conceived for a different objective, shares some similarities with ours in both the reference scenario (i.e., smart devices and IoT) and the adopted methodology.

To carry out this task, from the scientific literature, we identified the work described in [6]. It presents an intrusion detection system aiming at protecting smart devices in vehicular networks. In this approach, the main idea is to group nodes into "clusters" in order to build protected zones where nodes collaborate to improve their security. We remark, again, that the goal of the approach of [6] is different from the objective of our approach. However, both of them define a security model conceived to operate on smart devices and IoT, and their strategy is centered on the presence of groups and clusters of objects.

Interestingly, the authors of [6] measure the delay introduced by their solution to the communication time. In Section 6, we carried out a similar analysis but we evaluated another performance parameter, namely the increase of the average path length caused by our privacy preserving solution. In order to allow a comparison between our approach and the one of [6], we decided to measure the communication delay introduced by our approach. We defined it as the average difference, in terms of time to delivery, between a scenario in which our approach is used and another in which it is not adopted. As done in [6], we measured such a variation against the size of groups. To estimate communication time, we leveraged a global ping service available at the address `https://wondernetwork.com/pings`. In Figure 8, we report both our results and the ones of the approach described in [6].

From the analysis of this figure we observe that the average delay introduced by our approach ranges from 22 ms to 130 ms, whereas the average delay of the approach of [6] ranges from 24 ms to 170 ms. The outcome of this experiment shows that the performance of our approach is comparable with the one of other solutions, already present in the scientific literature, addressing security issues in the context of smart devices and IoT. This encourages us to state that our approach achieves pretty satisfactory results, still preserving the overall IoT usability to values considered acceptable by the scientific community in this application scenario.

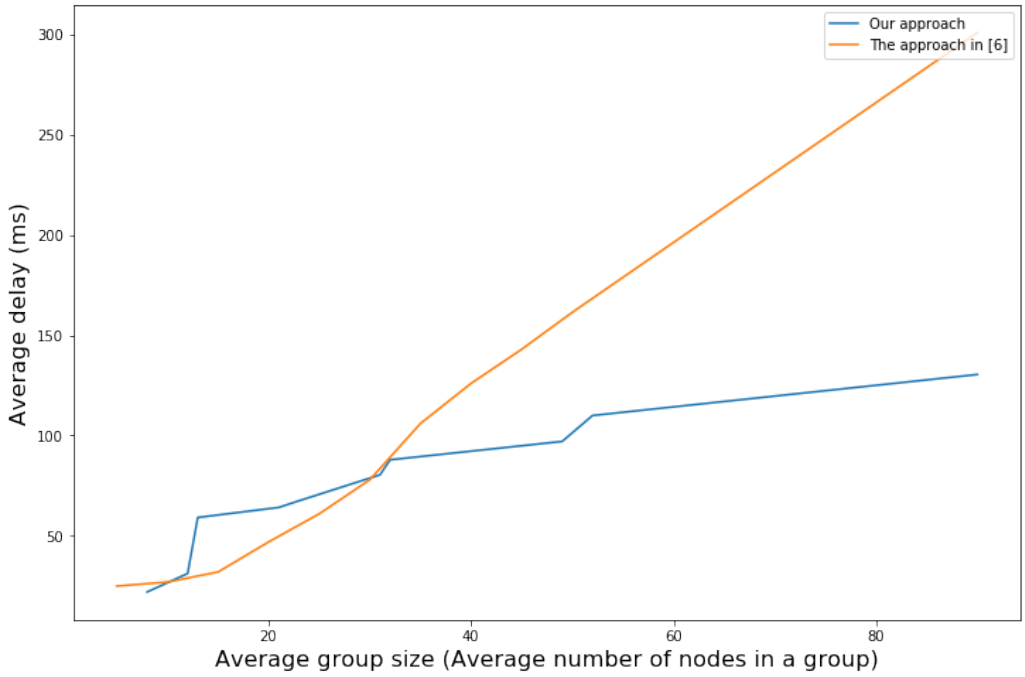

Figure 8: Average delay in the objects' communication introduced by our approach against the group size

8 Conclusion

In this paper, we have proposed a privacy-preserving approach capable of preventing feature (and, consequently, information) disclosure in an IoT scenario. Our approach is capable of guaranteeing privacy also in complex contexts where many smart objects are strongly interconnected to form a set of (partially overlapped) networks interacting with each other. After the study of related literature, we have illustrated the proposed model. Then, we have described the privacy-preserving object grouping scheme, which represents the core of our approach. Finally, we have illustrated the corresponding security model and we have analyzed the associated properties.

Our approach provides important advantages on user privacy protection in all those scenarios where the knowledge of the objects' features may help an attacker to derive information about user habits and behavior. For instance, in the Introduction, we presented a use case in the medical domain in which an attacker correlates the features provided by three devices (glucose meter, hemoglobin meter, and oxygen meter) applied to the same patient to the implications of some forms of lung cancer. If we apply our approach to this use case, each of the three objects is mapped to a set of (possibly) more generic features. Therefore, the attacker cannot know the exact scope of the objects and, hence, cannot guess the reason why the patient uses them. As a consequence, she is no longer capable of identifying the patient's health conditions.

This paper must not be considered as an ending point. Actually, it could be the starting point of several research efforts in this setting. As a first research direction, we plan to improve our approach by enhancing group formation using the probability of the associated nodes to be good contacts for

each other (i.e., to share common interests and, therefore, to exchange valuable information with the other nodes of their group). Indeed, currently, we consider only available features and the arrival time in WZ as a triggering factor for group creation. It would be useful to understand whether an improved algorithm can be designed so that the membership to a group can also be favored when it leads to an increase of available information for its nodes.

Furthermore, we are planning to include in our approach a security mechanism that prevents malicious nodes from being able to participate to a group in order to acquire a given set of features. Although this does not have impact on the privacy of other nodes inside the attacked group, it can lead to a detriment of performance. Solutions based on trust and reputation models can be adopted to prevent this kind of attack.

Empowering our solution with a reputation model would allow another future development. Indeed, currently, our privacy model includes some static countermeasures for node protection, based on the features involved in the queries received (cross-feature interview, see Section 5.2.3). The basic idea states that a node under attack changes its normal behavior by reducing its answer rate; however, no action is taken against the suspected attacker. It could be useful to exploit information about the suspected attacker to train a reputation model, so that whenever this kind of attack occurs, nodes can exchange information about the attacking node, thus updating their trust on it and its overall reputation.

Finally, we are currently studying another possible extension of our work that leverages the Blockchain capability of improving the security of transactions between nodes. Indeed, in our scenario, a possible attack could involve a malicious node that distributes false information inside a group (i.e., manipulating the messages exchanged therein). The use of Blockchain as a public ledger may prevent this kind of attack because each message can be anonymously traced inside it. In this way, each modification to the generated messages can be recognized by analyzing the corresponding digest reported in the Blockchain.

Acknowledgments

This work was partially supported by the Department of Information Engineering at the Polytechnic University of Marche under the project "A network-based approach to uniformly extract knowledge and support decision making in heterogeneous application contexts" (RSAB 2018). The Authors thank Enrico Corradini, who supported them in preparing the final version of the paper.

References

[1] C. Zhang, L. Zhu, C. Xu, K. Sharif, X. Du, and M. Guizani, "LPTD: Achieving lightweight and privacy-preserving truth discovery in CIoT," *Future Generation Computer Systems*, vol. 90, pp. 175–184, 2019. Elsevier.

[2] E. Curry, W. Derguech, S. Hasan, C. Kouroupetroglou, and U. ul Hassan, "A Real-time Linked Dataspace for the Internet of Things: Enabling "Pay-As-You-Go" Data Management in Smart Environments," *Future Generation Computer Systems*, vol. 90, pp. 405–422, 2019. Elsevier.

[3] L. Atzori, A. Iera, and G. Morabito, "SIoT: Giving a social structure to the Internet of Things," *IEEE Communications Letters*, vol. 15, no. 11, pp. 1193–1195, 2011. IEEE.

[4] G. Baldassarre, P. Lo Giudice, L. Musarella, and D. Ursino, "A paradigm for the cooperation of objects belonging to different IoTs," in *Proc. of the International Database Engineering & Applications Symposium (IDEAS 2018)*, (Villa San Giovanni, Italy), pp. 157–164, 2018. ACM.

[5] G. Baldassarre, P. Lo Giudice, L. Musarella, and D. Ursino, "The MIoT paradigm: main features and an "ad-hoc" crawler," *Future Generation Computer Systems*, vol. 92, pp. 29–42, 2019. Elsevier.

[6] M. Aloqaily, S. Otoum, I. A. Ridhawi, and Y. Jararweh, "An intrusion detection system for connected vehicles in smart cities," *Ad Hoc Networks*, vol. 90, p. 101842, 2019.

[7] S. Otoum, B. Kantarci, and H. Mouftah, "On the feasibility of deep learning in sensor network intrusion detection," *IEEE Networking Letters*, vol. 1, no. 2, pp. 68–71, 2019.

[8] F. Al-Turjman and S. Alturjman, "Context-sensitive access in Industrial Internet of Things (IIoT) healthcare applications," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2736–2744, 2018.

[9] J. Liu, Y. Xiao, and C. Chen, "Authentication and access control in the Internet of Things," in *Proc. of the International Conference on Distributed Computing Systems Workshops*, (Macau, China), pp. 588–592, IEEE, 2012.

[10] F. Al-Turjman, H. Zahmatkesh, and R. Shahroze, "An overview of security and privacy in smart cities' IoT communications," *Transactions on Emerging Telecommunications Technologies*, p. e3677, 2019.

[11] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[12] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. of the International Conference on Data Engineering (ICDE'07)*, (Istanbul, Turkey), pp. 106–115, IEEE, 2007.

[13] U. M. Farooq, M. Waseem, A. Khairi, and S. Mazhar, "A critical analysis on the security concerns of Internet of Things (IoT)," *International Journal of Computer Applications*, vol. 111, no. 7, 2015.

[14] R. Roman, P. Najera, and J. Lopez, "Securing the Internet of Things," *Computer*, vol. 1, no. 9, pp. 51–58, 2011.

[15] R. Lu, K. Heung, A. Lashkari, and A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT," *IEEE Access*, vol. 5, pp. 3302–3312, 2017.

[16] A. Dorri, S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for IoT security and privacy: The case study of a smart home," in *Proc. of the International Conference on Pervasive Computing and Communications Workshops (PerCom'17 Workshops)*, (Kona, HI, USA), pp. 618–623, 2017. IEEE.

[17] Z. Wang, "A privacy-preserving and accountable authentication protocol for IoT end-devices with weaker identity," *Future Generation Computer Systems*, vol. 82, pp. 342–348, 2018. Elsevier.

[18] I. Guedalia, J. Guedalia, R. Chandhok, and S. Glickfield, "Methods to discover, configure, and leverage relationships in Internet of Things (IoT) networks," feb 20 2018. US Patent 9,900,171.

[19] F. Buccafurri, V. Foti, G. Lax, A. Nocera, and D. Ursino, "Bridge Analysis in a Social Internetworking Scenario," *Information Sciences*, vol. 224, pp. 1–18, 2013. Elsevier.

[20] F. Buccafurri, G. Lax, A. Nocera, and D. Ursino, "Discovering Missing Me Edges across Social Networks," *Information Sciences*, vol. 319, pp. 18–37, 2015. Elsevier.

[21] P. Lo Giudice, A. Nocera, D. Ursino, and L. Virgili, "Building Topic-Driven Virtual IoTs in a Multiple IoTs Scenario," *Sensors*, vol. 19, no. 13, p. 2956, 2019. MDPI.

[22] S. Babar, P. Mahalle, A. Stango, N. Prasad, and R. Prasad, "Proposed security model and threat taxonomy for the Internet of Things (IoT)," in *Proc. of the International Conference on Network Security and Applications (CNSA'10)*, (Chennai, India), pp. 420–429, 2010. Springer.

[23] M. Abomhara and G. Koien, "Security and privacy in the Internet of Things: Current status and open issues," in *Proc. of the International Conference on Privacy and Security in mMbile Systems (PRISMS'14)*, (Aalborg, Denmark), pp. 1–8, 2014. IEEE.

[24] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, I. S. Jubert, M. Mazura, M. Harrison, M. Eisenhauer, and P. Doody, "Internet of things strategic research roadmap," *Internet of things-global technological and societal trends*, vol. 1, no. 2011, pp. 9–52, 2011.

[25] S. Otoum, B. Kantarci, and H. T. Mouftah, "Hierarchical trust-based black-hole detection in wsn-based smart grid monitoring," in *Proc. of the International Conference on Communications (ICC'17)*, (Paris, France), pp. 1–6, IEEE, 2017.

[26] F. Al-Turjman and I. Baali, "Machine learning for wearable iot-based applications: A survey," *Transactions on Emerging Telecommunications Technologies*, p. e3635, 2019.

[27] J. Bernabe, J. Hernández, M. Moreno, and A. Gomez, "Privacy-preserving security framework for a social-aware Internet of Things," in *Proc. of the International Conference on Ubiquitous Computing and Ambient Intelligence (UCAMI'14)*, (Belfast, Northern Ireland, UK), pp. 408–415, 2014. Springer.

[28] L. Atzori, A. Iera, and G. Morabito, "From smart objects to social objects: The next evolutionary step of the internet of things," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 97–105, 2014.

[29] V. Sharma, I. You, D. Jayakody, and M. Atiquzzaman, "Cooperative trust relaying and privacy preservation via edge-crowdsourcing in social Internet of Things," *Future Generation Computer Systems*, vol. 92, pp. 759–776, 2019. Elsevier.

[30] A. Alchihabi, A. Dervis, E. Ever, and F. Al-Turjman, "A generic framework for optimizing performance metrics by tuning parameters of clustering protocols in WSNs," *Wireless Networks*, vol. 25, no. 3, pp. 1031–1046, 2019.

[31] S. Cha, T. Tsai, W. Peng, T. Huang, and T. Hsu, "Privacy-aware and blockchain connected gateways for users to access legacy IoT devices," in *Proc. of the Global Conference on Consumer Electronics (GCCE'17)*, (Nagoya, Japan), pp. 1–3, 2017. IEEE.

[32] Y. Rahulamathavan, R. Phan, M. Rajarajan, S. Misra, and A. Kondoz, "Privacy-preserving blockchain based IoT ecosystem using attribute-based encryption," in *Proc. of the International Conference on Advanced Networks and Telecommunications Systems (ANTS'17))*, (Bhubaneswar, India), pp. 1–6, IEEE, 2017.

[33] M. Shen, X. Tang, L. Zhu, X. Du, and M. Guizani, "Privacy-Preserving Support Vector Machine Training over Blockchain-Based Encrypted IoT Data in Smart Cities," *IEEE Internet of Things Journal*, Forthcoming.

[34] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

[35] E. Shi, H. Chan, E. Rieffel, R. Chow, and D. Song, "Privacy-preserving aggregation of time-series data," in *Proc. of the Annual Network & Distributed System Security Symposium (NDSS'11)*, (San Diego, CA, USA), 2011. Internet Society.

[36] K. Lee, D. Kim, D. Ha, U. Rajput, and H. Oh, "On security and privacy issues of fog computing supported Internet of Things environment," in *Proc. of the International Conference on the Network of the Future (NOF'15)*, (Montreal, Quebec, Canada), pp. 1–3, 2015. IEEE.

[37] A. Alrawais, A. Alhothaily, C. Hu, and X. Cheng, "Fog computing for the Internet of Things: Security and privacy issues," *IEEE Internet Computing*, vol. 21, no. 2, pp. 34–42, 2017.

[38] P. Bajpai, A. K. Sood, and R. J. Enbody, "The art of mapping IoT devices in networks," *Network Security*, vol. 2018, no. 4, pp. 8–15, 2018. Elsevier.

[39] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "Opportunistic IoT: Exploring the social side of the Internet of Things," in *Proc. of the International Conference on Computer Supported Cooperative Work in Design (CSCWD'12)*, (Wuhan, China), pp. 925–929, 2012. IEEE.

[40] M. Antunes, D. Gomes, and R. L. Aguiar, "Towards IoT data classification through semantic features," *Future Generation Computer Systems*, vol. 86, pp. 792–798, 2018. Elsevier.

[41] J. Quevedo, M. Antunes, D. Corujo, D. Gomes, and R. Aguiar, "On the application of contextual IoT service discovery in information centric networks," *Computer Communications*, vol. 89, pp. 117–127, 2016. Elsevier.

[42] D. Wu, A. Taly, A. Shankar, and D. Boneh, "Privacy discovery and authentication for the Internet of Things," in *Proc. of the European Symposium on Research in Computer Security (ESORICS 2016)*, (Heraklion, Crete, Greece), pp. 301–319, 2016. Springer.

[43] S. Datta, "Towards securing discovery services in Internet of Things," in *Proc. of the International Conference on Consumer Electronics (ICCE'16)*, (Las Vegas, NV, USA), pp. 506–507, 2016. IEEE.

[44] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed Internet of Things," *Computer Networks*, vol. 57, no. 10, pp. 2266–2279, 2013.

[45] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[46] L. Atzori, A. Iera, and G. Morabito, "Understanding the Internet of Things: definition potentials and societal role of a fast evolving paradigm," *Ad Hoc Networks*, vol. 56, pp. 122–140, 2017. Elsevier.

[47] L. Atzori, A. Iera, G. Morabito, and M. Nitti, "The Social Internet of Things (SIoT)– when social networks meet the Internet of Things: Concept architecture and network characterization," *Computer networks*, vol. 56, no. 16, pp. 3594–3608, 2012. Elsevier.

[48] C. V. Rijsbergen, *Information Retrieval*. Oxford University Press, 1979. Butterworth.

[49] D. Wikström, "A universally composable mix-net," in *Proc. of the Theory of Cryptography Conference (TCC'04)*, (Cambridge, MA, USA), pp. 317–335, 2004. Springer.

[50] J. Furukawa and K. Sako, "Mix-net system," 2010. US Patent 7672460.

[51] D. Low, R. K. Huang, P. Mishra, G. Jain, and J. Gosnell, "Group formation using anonymous broadcast information," 2013. US Patent 8,359,643.

[52] M. Zamani, J. Saia, M. Movahedi, and J. Khoury, "Towards provably-secure scalable anonymous broadcast," in *Proc. of the International Workshop on Free and Open Communications on the Internet (FOCI'13)*, (Washington, D.C., USA), 2013.

[53] M. Abe, "Mix-networks on permutation networks," in *Proc. of the International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT'99)*, (Singapore), pp. 258–273, 1999. Springer.

[54] D. Goldschlag, M. Reed, and P. Syverson, "Onion routing for anonymous and private internet connections," *Communications of the ACM*, vol. 42, no. 2, pp. 39–40, 1999.

[55] N. Aitzhan and D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures blockchain and anonymous messaging streams," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 5, pp. 840–852, 2018.

[56] R. Hsu, J. Lee, T. Quek, and J. Chen, "GRAAD: Group Anonymous and Accountable D2D Communication in Mobile Networks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 2, pp. 449–464, 2018.

[57] A. Rehman, S. U. Rehman, and H. Raheem, "Sinkhole attacks in wireless sensor networks: A survey," *Wireless Personal Communications*, pp. 1–23, 2018.

[58] V. Ďurčeková, L. Schwartz, V. Hottmar, and B. Adamec, "Detection of Attacks Causing Network Service Denial," *Advances in Military Technology*, vol. 13, no. 1, 2018.

[59] M. Bouabdellah, N. Kaabouch, F. E. Bouanani, and H. Ben-Azza, "Network layer attacks and countermeasures in cognitive radio networks: A survey," *Journal of Information Security and Applications*, vol. 38, pp. 40–49, 2018.

[60] N. Chouhan, H. Saini, and S. Jain, "Internet of things: Illuminating and study of protection and justifying potential countermeasures," in *Soft Computing and Signal Processing*, pp. 21–27, Springer, 2019.

[61] W. Yang, Y. Wang, Z. Lai, Y. Wan, and Z. Cheng, "Security Vulnerabilities and Countermeasures in the RPL-based Internet of Things," in *Proc. of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC'18)*, (ZhengZhou, China), pp. 49–495, 2018. IEEE.

[62] P. R. Group, "Sensors in Distribution: On the Cusp of New Performance Efficiencies," *https://www.logisticsmgmt.com/wp_content/honeywell_wp_sensors_022316b.pdf*, 2015.

[63] K. E. Emam and F. K. Dankar, "Protecting privacy using k-anonymity," *Journal of the American Medical Informatics Association*, vol. 15, no. 5, pp. 627–637, 2008.

[64] C. Dwork, "A firm foundation for private data analysis," *Communications of the ACM*, vol. 54, no. 1, pp. 86–95, 2011. ACM.

[65] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," in *Proc. of the International Conference on Data Engineering (ICDE'06)*, (New York, NY, USA), pp. 24–24, 2006. IEEE.

[66] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014. Now Publishers.

[67] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to noninteractive database privacy," *Journal of the ACM (JACM)*, vol. 60, no. 2, p. 12, 2013. ACM.

[68] M. Hardt, K. Ligett, and F. McSherry, "A simple and practical algorithm for differentially private data release," in *Proc. of the Advances in Neural Information Processing Systems (NIPS'12)*, (Stateline, NV, USA), pp. 2339–2347, 2012.

[69] J. Domingo-Ferrer, "On the connection between t-closeness and differential privacy for data releases," in *Proc. of the International Conference on Security and Cryptography (SECRYPT'13)*, (Reykjavk, Iceland), pp. 1–4, 2013. IEEE.

[70] J. Domingo-Ferrer and J. Soria-Comas, "From t-closeness to differential privacy and vice versa in data anonymization," *Knowledge-Based Systems*, vol. 74, pp. 151–158, 2015. Elsevier.

[71] A. Kamra, E. Terzi, and E. Bertino, "Detecting anomalous access patterns in relational databases," *The International Journal on Very Large Data Bases*, vol. 17, no. 5, pp. 1063–1077, 2008. Springer.