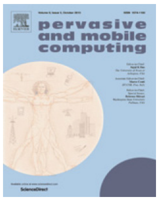
Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc



Fast track article

# Recommender system for privacy-preserving solutions in smart metering

Juan E. Rubio [*], Cristina Alcaraz, Javier Lopez

*Department of Computer Science, University of Malaga, Campus de Teatinos s/n, 29071, Malaga, Spain*

ARTICLE INFO

*Article history:*
Available online xxxx

*Keywords:*
Smart grid
Data privacy
Control
Metering
Protocols

ABSTRACT

Nowadays, Smart Grid is envisaged to provide several benefits to both customers and grid operators. However, Smart Meters introduce many privacy issues if consumption data is analysed. In this paper we analyse the main techniques that address privacy when collecting electricity readings. In addition to privacy, it is equally important to preserve efficiency to carry on with monitoring operations, so further control requirements and communication protocols are also studied. Our aim is to provide guidance to installers who intend to integrate such mechanisms on the grid, presenting an expert system to recommend an appropriate deployment strategy.

## 1. Introduction

With the introduction of the latest information, communication and operational technologies, the Smart Grid (SG) allows the utilities to accurately monitor energy consumption so they can adjust generation and delivery in near real-time. It also helps users get detailed consumption reports which are useful to save money by adapting power usage to the price fluctuation. The Advanced Metering Infrastructure (AMI) is the smart metering system that makes this possible [1]. It integrates multiple technologies that can be summarised as: Smart Meters, Communication networks disposed hierarchically, Meter Data Management System (MDMS), and Data integration systems.

Smart Meters (SM) are installed in the user end to sense and record the energy consumption data that is delivered to utilities periodically (e.g. every 15 min). Sometimes, a gateway is placed in between in order to gather all the sensed data in the Home Area Network (HAN) before forwarding it to the outsider network. HANs connects multiple devices like meters, generation and storage systems, vehicles as well as in-home displays and controllers. Then, the information is taken through different levels of aggregation points distributed over the utility network. Once it reaches the utility provider end, it is analysed for billing and control purposes, so they can handle Demand Response (DR) and react to potential changes in the grid. The MDMS is the main module responsible for these analytical operations.

In this context, AMIs with a high frequency of data collection can end up processing so many terabytes of information as to be deemed “Big Data”. Since it contains critical personal information, it is mandatory for the utilities to acquire disaster proof storage facilities and create advanced backup and contingency plans. Nevertheless, even if the information is kept secure, by subsequently analysing smart metering data of customers, it is possible to draw surprisingly accurate conclusions about their private lives: their time spent at home, working schedule, vacations, use of certain appliances, their habits, etc. which is known as “consumer profiling” and puts privacy at risk. This information is valuable for third parties such as insurance companies, entertainment agencies or government authorities. Apart from chasing a profit by analysing that data,

[*] Corresponding author.
*E-mail addresses:* rubio@lcc.uma.es (J.E. Rubio), alcaraz@lcc.uma.es (C. Alcaraz), jlm@lcc.uma.es (J. Lopez).

http://dx.doi.org/10.1016/j.pmcj.2017.03.008
1574-1192/© 2017 Elsevier B.V. All rights reserved.

there are multiple kinds of attacks against the SG infrastructure with diverse intentions: eavesdropping, denial of service, energy theft, exhibition of knowledge on behalf of the attacker or sensitive data retrieval. A collaborative effort is needed to encourage all the stakeholders involved in the SG to work together and address these issues: customers, grid operators, energy providers, billing companies, government agencies and other third party value added services. Each stakeholder has its own security requirements and sensitive objectives. In general, these objectives can be divided into the so called CIA triangle: confidentiality, integrity, and availability [2].

The purpose of this paper is to find solutions that provide privacy when processing consumption data. In this sense, there are multiple scenarios of privacy and different stakeholders whose privacy can be compromised, and hence they have different objectives in terms of security. Specifically, we consider three privacy scenarios in this paper:

- **Home**: relates to the privacy of the premise occupants, its aim being to hide life patterns when collecting energy measurements needed to calculate billing. In this context (i.e., for the customer), confidentiality is more important than the integrity of data and availability of resources.
- **Industry**: refers to the ability to keep as much information about the architecture and network resources of the organisation unknown, even if the attacker has a lot of information about energy consumed. Here, availability and integrity prevail over confidentiality, since business continuity under acceptable accuracy conditions must be ensured.
- **Vehicular**: addresses the privacy of data generated by Plug-in Hybrid Electric Vehicles in public and private charging points, comprising information like location or driver's identity. As it is a critical infrastructure, availability and integrity are crucial.

Our ultimate goal is to provide some guidelines for a hypothetical SG developer who aims to integrate an efficient privacy-preserving technique in one of these scenarios. To accomplish this, besides the primary security objective in the CIA triangle, there are other factors linked to the decision about which technique to deploy. Firstly, the architecture used in that part of the grid, and checking whether the communication model proposed by these techniques can adapt to it. Secondly, the available smart meter devices the installer has or needs to purchase, and whether they can support the cryptographic operations, bandwidth and latency imposed by these techniques. Lastly, the kind of communication protocol used by the SM to share energy data between devices within the HAN and with the gateways, also has to be studied to determine how and to what extent all the parties involved in the AMI can access the energy information.

Numerous surveys provide a brief overview of the most notable privacy-preserving techniques proposed. For instance, [3] proposes a classification of mechanisms based on the use of aggregation, and in [4] authors study the applicability of the techniques to billing and monitoring operations. However, not all take into consideration whether these mechanisms behave efficiently in terms of automation. For this reason, in addition to advising on which technique to choose in terms of privacy, we also focus on finding a trade-off between security and automation by assessing how these solutions affect the performance of the collection and supervision systems. Once we have presented all criteria and analysed each technique, we describe and show the implementation of an expert system based on a probabilistic model with a Bayesian Network. This system can compute a decision on which techniques are more suitable for deployment according to the installer's available equipment and needs.

The paper is organised as follows. In Section 2 the distinct architecture of the assessed techniques and smart meter types involved are presented. In Section 3 the performance properties analysed in the solutions are introduced. Section 4 describes and analyses the proposed privacy solutions. In Section 5 the most relevant communication protocols to use in this context are studied. Then, a recommendation system for the techniques based on the analysis is presented in Section 6. A summary, conclusions and future work are discussed in Section 7.

## 2. Smart metering architecture and devices

In this section, we introduce the communication model and architecture that the solutions analysed in Section 4 put into practice to define how SMs (producers of consumption data) connect to utilities (e.g., for monitoring or billing purposes). We then present the main types of smart meters that are capable of running these privacy techniques, according to their computational features.

Many researchers suggest tackling privacy issues in the SG by means of data aggregation. It implies gathering and computing data to preserve the anonymity of its origins [3]. On the one hand, this data can be aggregated spatially (over a set of different smart meters), so that the Energy Service Provider (ESP) can control the current status of a certain grid area without getting individual consumer's information. On the other hand, aggregation can be done temporally (collecting different readings from the same meter), in such a way that the ESP can compute the bill by receiving the overall consumption information once per billing period. After all, the objective is to prevent the ESP from checking the current energy consumption of a single consumer.

Some of the privacy preserving techniques discussed in this paper use aggregation. Since that operation can usually be computed in multiple places over the AMI depending on the approach of each solution, we can divide them into three kinds of architectures or communication models: (A) Locally centralised techniques where the SM communicates with a gateway or a third party, which aggregates the data before relaying it to the utility; (B) Distributed solutions which do not aggregate data or that process is computed in the own smart meter, and the SM communicates directly with the ESP; and (C) Other decentralised techniques whose topology, for instance, is a tree of smart meters, so each of them holds its own data and that of its children.

With regards to the smart meter device itself, there are diverse types subject to the proposed privacy approach. In its basic form, a smart electricity meter must provide power measurements, transmission of information and data display on the meter's front panel. With the improvements that AMI introduces, the SM includes new features: better sensing accuracy, electrical surges and outages monitoring, remote disconnection, possibility to add variable tariffs, communication with in-house devices, etc. This, together with the requirements of advanced processing algorithms like those explained below, all mean that the meter needs to have more sophisticated computational resources to cope with different communication protocols and cryptographic mechanisms. As a result, their readiness for such mechanisms depends on the family of integrated microcontroller units, the CPU and RAM requirements. In this paper, three groups of smart meters have been studied:

***Group* 1**: it encompasses low cost and simple meters used in privacy techniques for efficient cryptographic or communication mechanisms, like symmetric encryption, and they do not store consumption data locally. Regarding their benefits for the average user, they provide real time display with consumption and an approximation of the cost of energy usage. Technically, their embedded microcontroller is designed to achieve exceptional power efficiency through smaller processing cores. Their clock speed ranges from 25 to 50 MHz and the maximum RAM is 32 kB.

***Group* 2**: meters belonging to this category are expected to perform moderate operations in terms of complexity, such as asymmetric ciphering, signing or similar cryptosystems like secret sharing schemes. Additional information is held in the device to support these techniques (e.g., certificates and keys), together with temporarily aggregated data on energy usage. On the user side, these smart meters offer further functionalities, like alerts when different conditions have been reached. Their mid-range performance cores present clock speeds up to 180 MHz and a top RAM memory of 250 kB.

***Group* 3**: comprises SMs that require high performance with real time constraints and exceptional power efficiency. They are capable of running computationally expensive algorithms and embed a Trust Platform Module (TPM) in the device, as well as securely storing consumption data during the whole billing period. Furthermore, they allow the user to control all appliances on the premises and analyse energy consumption with a high level of detail. Their core speed can reach 200 MHz and it disposes from 100 to 300 kB of RAM.

That said, our objective is to recommend one of this kind of smart meter device for the grid operator who is willing to undertake privacy preserving measures on one area, depending on the technique's complexity. Whatever the type chosen, it is important to operate at low energy and include security features to protect the transmitted data and information stored in the device.

## 3. Control and automation requirements

As stated in the introduction, we offer an overview of the main privacy preserving techniques taking automation efficiency into account as one of the main priorities when choosing a solution. This is essential as a technique must not consume excessive bandwidth (which could cause delays in the data transmission) and run complex algorithms that require extra computational capabilities for the smart meter. According to the control requirements defined in [5], we can establish a taxonomy of features to be analysed in the solutions in terms of control. Each of them will be estimated for the set of privacy techniques described afterwards, according to a simple score: we will assign one point (represented with a ✓ in tables) if the mechanisms is outstanding in that feature; half a point (∽) if it mildly satisfies the respective condition; and none (×) if it does not address that feature at all. The resulting sum of points in each technique will therefore give an idea of its suitability and will ease the comparison with the rest.

***Real-time performance***: addresses the operational delays caused by the processing of information, application of techniques and the transference of the data to control utilities. These particular properties should be considered:

- **Speed**: it checks how fast the underlying algorithm is by doing an estimation of how many complex steps it takes to run (e.g., aggregating, encrypting, signing). Here, a technique has been considered as fast, denoted by ✓ in Table 1, when its cryptographic scheme is not complex and it does not imply expensive operations. An intermediate level of speed, denoted as ∽, can be conceived when the technique is somewhat competent but requires various operations. Otherwise, it is considered to be slow and represented with ×. Similar symbol criteria is applied for the rest of features.
- **Storage**: subject to the excess of operations and the massive storage, which can require extra resources to maintain aggregated meter values or additional data to implement the cryptographic schemes. In this regard, a technique is productive if the meter does not hold consumption data or it is stored in a third party for aggregation. Otherwise (i.e. the information is held locally in the SM), the technique is not considered appropriate.
- **Communication overhead**: is related to the frequency of data delivery and the number of messages transmitted between the SM, an eventual third party and the ESP, so as to not hamper the data recollection and the supervision of the area. Namely, a solution is accepted as efficient (✓) if there is only one message and it occurs between the SM and the ESP; moderately efficient (∽) when the SM contacts the ESP once but it is always transmitting data to a third party to accomplish monitoring operations; and inefficient (×) when the SM is frequently sending data to the ESP.

- **Synchronisation**: it focuses on the time when data streams are being sent from the producer (i.e., the smart meter) to the consumer (i.e., the energy service provider). Whereas certain protocols may require all data producers to send it simultaneously (increasing the complexity of the protocol, thus it is denoted as ×), in others the data producers send it independently of each other.

***Sustainability***: a technique can be considered as sustainable when it is able to meet the needs of the present without compromising the ability of future generations to meet their own needs [6] due to compatibility problems. It is necessary to consider how easy it is to configure the protocol in all the entities involved (configurability) and the ease to update and upgrade measures without reducing control tasks (maintainability). If both requirements are addressed by means of an available mechanism that enables the technique to be flexibly configured and updatable, it is marked with a ✓.

***Dependability***: can be defined as "the ability of the system to properly offer its services on time, avoiding frequent and several faults" in Al-Kuwaiti et al. [7], and includes reliability and security as main properties. However, we only address the reliability because the security is already part of the privacy solutions. In this category, we consider:

- **Fault-tolerance**: robustness of the protocols to bear unlimited software and hardware failures (✓), just a certain number of them (∽) or no failures at all (×).
- **Aggregate error**: presence of errors in the aggregated data that are a result of metering failures or a consequence of applying perturbation to preserve privacy, in which case the technique is not exact (×).

***Survivability***: capability of a system to fulfil its mission and thus address malicious, deliberate or accidental faults in a timely manner. Particularly, we consider **resilience**, which is closely related to fault-tolerance but responding to attacks and allowing the system to continue its services when part of its security is compromised. In the event that a technique defines a mechanism to recover from a cyber attack, its resilience is considered good enough (✓).

## 4. Analysis of privacy techniques

In order to ensure privacy in these scenarios, many techniques have been proposed in literature to allow the ESP to collect usage data from SMs for monitoring and billing purposes without revealing sensitive information. In this section, we present eleven techniques, giving a brief description of their implemented cryptosystem, the architecture, type of aggregation and smart meter used, as introduced in previous sections. After that, control requirements listed in Section 3 will be also estimated in each solution, resulting in Table 1. For the interest of readability, all techniques have been classified according to the Privacy Enhancing Technology (PET) that they rely upon:

### *4.1. Trusted Computation (TC)*

The SM itself or a third party is entrusted to aggregate consumption data before it is sent to the ESP, so it does not receive any sensitive data. One concern of this approach is the trust given to that party, which is responsible of the information treatment. These solutions have been considered: Bohli et al. [8] propose aggregating accurate meter readings from the SM in a Third Trusted Party (TTP) before sending them to the ESP, instead of transmitting them immediately. Specifically, SMs transfer their data through an encryption channel to the TTP, which sums up individual consumption for each smart meter at the end of the billing period and also informs the ESP about the current status of some part of the grid, by aggregating the data of various SMs in a given moment. As a result, there is no link between both communication channels, so it can be also seen as a form of anonymisation. In terms of efficiency, it is fast since it uses symmetrical encryption (usually AES) and light because it relays information to the TTP without storing data in the meter. However, it introduces some communication overheads due to the permanent data submitted by SMs to the TTP to monitor the electricity consumption of a certain area. Consequently, a Group 1 smart meter would be appropriate for this technique. On the other hand, it is also robust since it can detect the presence of fake groups (i.e., sets of SMs controlled by the ESP that emit default values in order to isolate the real customer's consumption).

LeMay et al. [9] propose a model which aggregates energy measurements in the smart meter to calculate the bill, by using a TPM. More specifically, its software architecture is composed by independent virtual machines intended to perform diverse applications like billing or DR. A hypervisor controls the access to the hardware (hence the power measurements) and integrity and confidentiality are guaranteed through remote attestation, which proves to the provider that the hardware and software are trustworthy. To achieve this, the device includes hardware-protected storage, cryptographic modules and other tamper detection components to defend against physical attacks. In terms of control requirements, this solution reduces the amount of information transmitted between SM and ESP, but all data is processed at its origin, so it is not that efficient from the perspective of storage. Despite that fact, the TPM also allows the service provider and the customer to run their own applications by creating new virtual machines, ensuring sustainability by this means. All these demands make it obligatory to choose a smart meter of Group 3.

Please cite this article in press as: J.E. Rubio, et al., Recommender system for privacy-preserving solutions in smart metering, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.03.008

### 4.2. Verifiable Computation (VC)

This privacy-enhancing technology permits the consumer to calculate the aggregated data for him/herself and send proof to the provider to ensure its correctness. Usually, a Zero-Knowledge (ZK) protocol is used, which allows a prover (the smart meter in this case) to demonstrate the knowledge of a secret (the power readings needed to compute the bill) to the verifier (the ESP) without revealing the electricity usage or permitting under reporting.

Molina-Markham et al. [10], describe a generic ZK protocol for the smart meter to compute the bill locally and prove its conformity with consumption to the utility. In addition to this, neighbourhood gateways are optionally placed between the SMs and the ESP to relay aggregated power readings corresponding to an area without disclosing any particular origin, enabling DR operations. Regarding its automation features, the protocol is computationally expensive, although the communication between the SM and the ESP takes place only once per billing cycle between SM and ESP. Also, plenty of storage is needed to hold all the measurements and then create the proof, so it is necessary to have a Group 3 smart meter to run this protocol.

Jawurek et al. [11] also specify another ZK protocol based on Pedersen commitments [12]. It introduces a plug-in Privacy Component (PC) between the SM and the ESP that intercepts consumption data and sends the provider signed commitments and the final calculation together with the random parameters used to create the Pedersen commitments from individual measurements. Taking advantage of the homomorphic property of this schema, the ESP can effectively check the bill validity computing the calculation on the received commitments, which result in a new commitment of the bill amount and random numbers presented. The PC processing is invisible to the SM and does not have to be trustworthy, since the VC protocol itself ensures a correct bill calculation, and therefore it can be implemented easily with no special hardware-protected components. With respect to its control suitability, the underlying algorithm has a high computational cost. However, it is adequate with respect to storage and communication overhead, since the PC frees the load from the smart meter. For this reason, a device from Group 2 is required.

### 4.3. Cryptographic Computation (CC)

These techniques apply secret sharing or homomorphic cryptographic schemes, so the provider can only decrypt the aggregate of consumption data and not individual data items. Rottondi et al. [13] propose an approach based on the introduction of Privacy-Preserving Nodes (PPNs) between the SMs and the ESP that aggregate data based on space (for a set of SMs spread in an area) and time (for a single SM) depending on the need and access rights managed by a central configurator. Privacy is preserved with the use of a secret sharing scheme: a secret (i.e., the energy usage information) is divided into shares that are distributed among the nodes, so that the ESP cannot reconstruct the measurements until it disposes of, at least, a defined number of them. Exploiting the homomorphic properties of the sharing scheme, these power readings can be aggregated in the PPNs and then delivered to the ESP without revealing individual measurements. Its respective smart meter is one of Group 2: it only has to calculate the secret shares and send them to the PPNs. As for configurability and maintainability, they are achieved by the central configurator. Moreover, this architecture is resilient against faulty or compromised PPNs as long as the number of healthy ones is above a certain threshold. However, it is the only technique that requires synchronising to gather measurements from a set of smart meters at the same time.

Li et al. [14] propose organising the smart meters in a certain area in a tree topology with the ESP in the root. Each node, beginning with the leaves, encrypts its individual energy values and aggregates them with those of its children using the Paillier homomorphic cryptosystem [15]. Then, the SM passes the sum to its parent, which perform the same operation until it ultimately reaches the root. Its key is used for encryption, so no inner-node can access any individual measurements and the ESP can only obtain the sum of them. The complexity derives from the creation of the tree prior to running the protocol. Its height should be small enough to reduce the hops and its nodes should not have too many children to avoid excessive computation and communication load. It offers mid-range speed (it needs to asymmetrically encrypt the data and homomorphically multiply the value with the other nodes) and its overhead and storage depend on the tree topology. It is recommendable to install a SM device that belongs to the third Group in order to satisfy these requirements.

### 4.4. Anonymisation (Anon)

Anonymisation consists of removing the smart meter identification or substituting it with pseudonyms. In this context, Efthymiou et al. [16] establish a division between two kinds of data generated by the SM. On the one hand, high-frequency measurements (e.g., collected every 15 min) transmitted to the ESP to perform monitoring operations over a set of SMs, which have to be pseudoanonymised due to the information they provide about a user's private life. On the other hand, low-frequency metering data (e.g., collected monthly) that is attributable for billing purposes. An identification is assigned to each type: HFID (High-Frequency ID) and LFID (Low-Frequency ID), respectively, with different associated certificates (the first signed by the TTP and the second issued by the ESP). Whilst high-frequency data is sent to an escrow with the HFID and remains unknown to the ESP, low-frequency data is disclosed publicly and is linked to LFID. The ESP can query the escrow to verify the connection between an HFID/LFID pair. This solution does not introduce any extra burden on the storage and it implies affordable cryptographic operations, so a Group 2 m is enough to implement such mechanisms. As a measure against power theft, the authors also propose sanctioning nodes by temporarily lifting their anonymity. One disadvantage is the complex setup phase to establish the identities for each smart meter, which includes randomly chosen waiting times to ensure unlinkability between HFID and LFID.

Petrlic et al. [17] propose an anonymisation technique that uses a trusted third party. It issues pseudonym certificates for the SMs, which are used to encrypt and sign power readings. This data is relayed by the TTP once it has verified the signature and remove any identifiable information, subsequently forwarding it to the ESP. Therefore, no aggregation is performed, and a TPM is assumed to be present in the household. This is for calculating the bill at the end of the month, while still being able to detect manipulations of the meter through remote attestation. As an outcome, it does not store additional information apart from the certificates, although readings have to be asymmetrically encrypted and signed. Furthermore, the solution presents some overheads because of the permanent data delivery between the SM and the TTP. A Group 3 device is expected to cover the specifications for this protocol.

### *4.5. Perturbation (Pert)*

Perturbation refers to deliberately adding random noise to the measurements data while keeping it valid for control purposes. Lin et al. [18] propose a semi-trusted storage system which securely stores all the data from meters in an area. On the one hand, the Load Monitoring Center (LMC) can only access a sum of meter readings from several SMs in a single time unit, enabling supervising operations. Random noise is introduced in that encrypted information, so LMC obtains an approximate aggregation that can be considered accurate with a given probability. On the other hand, the ESP can only take the sum of readings from a single SM over a time period, in order to calculate the bill. As for its control suitability, this solution is computationally efficient as it only uses modular additions to encrypt and save the usage data in the central repository. Nevertheless, it uses a TPM to compute remote attestation and generate the pseudorandom numbers needed to encrypt the measurements. One drawback that this approach has is the continuous communication that occurs between the ESP or LMC and the SMs in order to regenerate these numbers to decrypt the readings. In addition, this is the only technique surveyed in this paper that introduces noise and is not tolerant against failures, since the LMC needs the meters to reply with their blind factors used for decrypting data. In consequence, a type 3 m is required.

### *4.6. Obfuscation (Obf)*

Even though obfuscation cannot be considered as a PET as such, it strengthens privacy in smart metering systems by hiding the actual power demand. Kalogridis et al. [19] define the concept of 'load signature moderation' to shape the electricity consumption so it does not expose any sensitive data. They propose the introduction of a decentralised energy production system within the household, so power can be dynamically drawn from re-chargeable batteries and other energy storage and generation devices. Thus, actual energy usage curves can be changed, hidden, smoothed, obfuscated or emulated. This solution protects against attackers that have a physical control over the SM and does not depend on specific grid architectures or trust relationships, while being compatible with other additional mechanisms and enabling grid monitoring. With respect to control, it is a fast solution that does not require any data processing or saving, so a Group 1 m can be used. However, it requires extra computation when the battery is almost charged or empty in order to keep masking the consumption and hence preserve privacy, and it is not fault-tolerant because a failure in the power routing system leaves all the real measurements exposed.

Egarter et al. [20] propose another obfuscation approach that is not based on a controllable battery, which has limitations with respect to the maximum charging and discharging rate and is therefore costly. Whereas Kalogridis et al.'s work tries to flatten the houdehold's energy demand and hence maintain a constant metered load by charging and discharging a battery, this new model uses a variable load through a device whose consumption is adjustable, like an electric boiler. A random energy consumption is assigned to it on a daily basis, in such a way that it overlays net demand by injecting noise which impedes the detection of other appliances. In comparison with the previous approach, a system like this would not require any great changes in the households wiring, since it is not necessary for the solution to measure the actual level of net demand. Again, as for the battery model, any special meter or architecture outside of the home are needed, being an efficient, but not fault-tolerant solution. A Group 1 m can be used.

Exactly as Table 1 shows, Trusted Computation and Anonymisation are PETs that demonstrate a better behaviour when performing control operations. Particularly, [9,17] are suitable solutions that involve trust with the use of a TPM, which results in the need to introduce a three-type smart meter. Alternatively, [8,16] are similar techniques that are less complex.

## 5. Communication standards and protocols

We have reviewed the privacy-preserving techniques cataloguing each one according to the architecture, group of smart meter used and automation efficiency, which gives an accurate hint to the developer or grid operator when adopting one of these solutions. Alongside this study, we now intend to carry out a similar analysis of the communication protocol implemented in the AMI to allow the utility to retrieve energy measurements from the smart meters once the privacy technique has been integrated. Afterwards, we perform an analogous assessment with respect to the protocols in the HAN,

**Table 1**
Control requirements for surveyed privacy protocols.

| Implemented PET | | TC | | VC | | CC | | Anon | | Pert | Obf | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CR[a] | FTP[b] | [8] | [9] | [10] | [11] | [13] | [14] | [16] | [17] | [18] | [19] | [20] |
| Performance | Speed | ✓ | ∽ | × | ∽ | ∽ | ∽ | ∽ | ∽ | ✓ | ✓ | ✓ |
| | Storage | ✓ | × | × | ✓ | ✓ | × | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Comm. | ∽ | ✓ | ∽ | ✓ | ∽ | ∽ | ∽ | ∽ | × | ∽ | ∽ |
| | Sync. | ✓ | ✓ | ✓ | ✓ | × | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Sustainability | Config. maint. | | ✓ | | | ✓ | | | ✓ | | | |
| | | | ✓ | | | ✓ | | | ✓ | | | |
| Dependability | Fault-tol. | ✓ | ✓ | ✓ | ✓ | ∽ | ✓ | ✓ | ✓ | × | × | × |
| | Agg. error | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | × | ✓ | ✓ |
| Survivability | Resil. | ✓ | ✓ | | | | ✓ | ✓ | ✓ | ✓ | | |
| Architecture | | A | B | A | A | A | C | A | A | A | B | B |
| Smart meter type | | 1 | 3 | 3 | 2 | 2 | 3 | 2 | 3 | 3 | 1 | 1 |

[a] Control requirements.
[b] Features of the privacy techniques.

which regulate how the data from all the connected devices is collected. We will follow the same methodology as with the privacy techniques: firstly, we will describe each protocol and then we will study how they comply with set of reliability and security characteristics (representing this information in form of a table):

***Reliability features***: the communication protocol can be deemed as reliable if it ensures a successful delivery of data between sender and recipient, with certain mechanisms: acknowledgements (ACKs), timeouts and retransmissions to avoid the loss of messages, and error detection techniques. Even though the presence of these mechanisms is considered positive for our goals, it is important to stress that some of them introduce more overhead.

***Security features***: it is necessary that the protocol includes cryptographic functions to address this set of security services:

- **Authentication**: assurance that the communicating entity is the one that it claims to be. It can be performed in many ways, ranging from less to more robust security: with a basic identification (a password or the device ID itself), through Message Authentication Codes (MAC) or by exchanging digital certificates.
- **Confidentiality**: protection of data from unauthorised disclosure, by means of encryption algorithms. In this sense, there are stronger cryptosystems like AES or 3DS, compared to other weak algorithms such as DES or RC4.
- **Integrity**: it ensures that the data received is the same as sent by the other peer (i.e., contain no modification, insertion, deletion or replay). Hash functions can be leveraged to accomplish this (i.e. SHA).
- **Access control**: prevents against the unauthorised use of a resource. There are multiple schemes ranging from simple permissions to individual users, like the Access Control Lists (ACLs), to robust and flexible models based on roles (such as RBAC, role-based access control).
- **History logs**: used for monitoring the resource usage in case of incidents.

With this, the goal is to be able to recommend a protocol based on their security and reliability characteristics, together with the privacy scenario of those presented in the Introduction (i.e., home, industrial or vehicular).

### *5.1. Analysis of AMI communication protocols*

To start with, OSGP [21] (Open Smart Grid Protocol) is a communication protocol between smart grid devices and gateways, which support the connection between the SM and ESP and offload the utility from many networking tasks. Apart from meters and gateways, it is applicable for different devices in the Smart Grid, such as control modules or solar panels. It is a family of specifications published by the European Telecommunications Standards Institute (ETSI), and has been widely deployed in millions of smart meters. The standard offers reliable end-to-end device communications through a three tier architecture for decentralised applications: meters, concentrators following a proxy layer network, and utilities. Its layers follow this structure, according to the OSI model levels:

- At the physical level, OSGP can be used with any communication technology, but it commonly uses ETSI TS 103 908 as its power line communication standard.
- For the Networking Layer, OSGP uses ISO/IEC14908-1, a multi-purpose control network optimised for smart city applications.
- At the Application level, OSGP uses a data model based on tables, as specified in the IEEE 1377 and ANSI C 12 standards.

Concerning reliability, OSGP implements some measures to handle errors and message loss, like retransmissions and acknowledgements. It also defines a service of time-stamped records of alarms and events in the device. In terms of security, it offers message authentication and encryption by means of two associated keys: BEK (Base Encryption Key) and OMAK (Open Media Access Key). The latter is also used to control read and write permissions over data tables.

Secondly, Zigbee Smart Profile V2.0 [22] is a communication standard suitable for automatic meter reading, which also addresses load control and demand response. Its specification was published by recognised organisations (e.g., Zigbee Alliance, HomePlug Powerline Alliance, IETF) and defines an application protocol built using the four layer Internet stack model, consequently it is compatible with several link and physical protocols. It implements a REST architecture using HTTP, basing its messages on simple commands like GET or POST. Since TCP is utilised as transport protocol, it manages the session providing delivery assurance and windowing. Regarding its security aspects, transactions are secured using TLS version 1.2, which is relied on to perform authentication for both the meter and the utility through its inner handshake mechanism. Encryption is used to provide confidentiality, using the AES-CCM mode of operation, and integrity check is also ensured by TLS. Also, an access control list is integrated to allow or deny the use of certain resources basing on the authentication level.

The ANSI C12.18 [23] standard also specifies a two-way communication protocol between a smart meter device and a client, which may be a handheld reader, a portable computer or a remote station, using an ANSI Type 2 Optical Port. More precisely, it defines an OSI 7-layer model to transport data in table format, as described in ANSI C12.19. Basically, the normal communication of data takes place following these three steps: (1) establishment of communication channel and parameter negotiation; (2) transport of information; and (3) closure of the information channel.

During the first phase, both counterparts are identified prior to negotiate certain communication parameters. For this reason, the protocol contemplates an optional basic security service to set which access is provided for the client, with a password to enable reading determined data tables. However, no special confidentiality or integrity mechanisms are defined. With respect to reliability, various techniques at data link layer are specified, such as CRC (Cyclic Redundancy Check), ACKs, retransmissions or time-outs.

Modbus [24] is another application layer communication protocol, which is primarily intended for control and automation purposes but is also suitable for connecting devices on different types of buses or networks (e.g., using TCP/IP over Ethernet or with a serial communication), as is the case of the smart metering system. It was published by Modicon in 1979 and defines a request/reply protocol through function codes, being possible for the vendor to implement custom ones. More specifically, data is based on a series of tables, and it uses three kinds of protocol data units: request, response and exception response. This last message is generated, for instance, when the server (the meter in this particular case) cannot handle the client request (i.e., from the reader). Basic error detection mechanisms are provided by introducing parity and CRC checks in the PDUs, along with ACKs to notify the reception of messages. However, it leaves the responsibility to implement timeouts. As for its security, Modbus does not consider any special measure to protect against unauthorised commands or data interception. It only specifies a remote identification function for both the client and server, which allows them to present data related to their device identity: vendor name, product code, etc. It offers three types of identification, increasing the amount of data that must be provided to access the data objects.

Similar to Modbus, DNP3 [25] (Distributed Network Protocol) is a protocol principally used by water and electric companies to monitor their equipment, but is applicable for the AMI as well. It was conceived to transfer data in SCADA (Supervisory Control and Data Acquisition) applications, using serial and IP communications. Conforming to the OSI model, it defines a three layer protocol: data link, data transport and application layer. In the data link layer, it provides multiplexation and data fragmentation, and is responsible of making the physical link reliable, through error checking and duplicate frame detection. At data transport layer, it breaks long application layer messages into smaller packets called segments. Lastly, the application layer defines functions and data types to read and write values on the device.

Originally, DNP3 was designed to address reliability, but not to deal with security concerns. Since it is an open model, anyone could launch an attack on a SCADA system and potentially disrupt the control systems of these critical infrastructures. This forced DNP3 to develop a Secure Authentication service, which includes TLS, a RBAC access control and is compliant with IEC 62351-5. As a result, DNP3 although more complex is a more robust and efficient protocol compared to Modbus.

M-Bus [26] (Meter-Bus) is a protocol developed by Professor Dr. Horst Ziegler of the University of Paderborn in cooperation with Texas Instruments Deutschland GmbH and Techem GmbH. It was designed to remotely read smart meters information, such as water, gas or electricity consumption within a household. It is based on the master–slave model, so when a query is send to the meter, it responds with the requested data. In terms of the OSI model, it functions on three layers: in the physical layer, the electric specifications, topology and data representation are defined. Then, the data link layer describes the transmission parameters and format of packets (called telegrams), addressing data integrity through CRC. The application layer, based on EN 1434-3, defines a set of actions and a data structure with respect to the response from the slave to the master. Regarding its security aspects, Meter-Bus specifies multiple encryption modes at the application level, which are basically AES and DES in CBC or CTR mode and distinct initialisation of IV. However, it does not contemplate any authentication method or integrity techniques (such as MAC) beyond simple CRCs at the frame level.

IEC 60870-5-104 [27] (also known as IEC 104) is another standard for power systems' monitoring and control, also suitable to communicate devices and securely send data through a TCP/IP connection. It is based on IEC 60870-5-101 but introduces some changes, both of them being interoperable and compatible. It combines the application layer of IEC 60870-5-101 and the transport function provided by TCP/IP. In level 7 it defines message formats and information objects to classify data according to different commands. As a result, the protocol supports cyclic and polling data acquisition with a function for time synchronisation. Furthermore, at the same application level, it provides protection against frames loss and duplication in addition to similar mechanisms defined with TCP at the transport level.

Please cite this article in press as: J.E. Rubio, et al., Recommender system for privacy-preserving solutions in smart metering, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.03.008

**Table 2**
AMI protocols features comparison.

| | | OSGP | Zigbee smart profile V2.0 | ANSI C12.18 | Modbus | DNP3 | M-Bus | IEC 60870-5-104 | DLMS/IEC 62056 | OCPP |
|---|---|---|---|---|---|---|---|---|---|---|
| Reliability features | ACKs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Timeouts | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | Retransmissions | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| | CRC | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Security features | Authentication | MAC | TLS certificates | Basic identification | Basic dev ID | MAC and dev ID | Basic dev ID | | Based on pwd | Basic dev ID |
| | Confidentiality | RC4 | AES-CCM | | | AES, RC4, 3DES | AES, DES | | AES-GCM | From TLS |
| | Integrity | | From TLS | | | SHA | | | SHA | From TLS |
| | Access control | Basic permissions | ACL | pwd Access | | RBAC | | | RBAC | ACL |
| | History logs | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ |

dev = device.
pwd = password.

As far as security is concerned, it does not implement any measures but rather states that the vendor should introduce additional mechanisms to comply with the IEC 62351 standard, which is a reference in the industry to address security. It provides cyber security guidelines for communication and control protocols in the SG, defining suitable techniques that should be provided to achieve authenticity, confidentiality and integrity, as well as control access mechanisms. However, it is rarely adopted because it increases complexity and costs.

IEC 62056 [28] is a set of standards for metering data exchange and load control. Specifically, IEC 62056-21 is a standard that describes a protocol which permits reading and programming tariff devices, so it is particularly suitable for the environment of electricity metering. Its data link layer specification divides the data exchange into multiple steps: (1) identification of devices; (2) parameter negotiation (frame length, windowing, speed rate, etc.); (3) read/write operations; and (4) sign-off. During the negotiation phase, an optional authentication and control access mechanisms can be used. It specifies three levels of authentication security: no authentication, low level (identification of client based on password) and high-level (to identify both the client and the server). In the application layer, COSEM (Companion Specification for Energy Metering) is used to represent the information, defining the Transport and Application Layers of the DLMS protocol (Device Language Message Specification), a generalised concept for abstract modelling of communication entities. It has been developed and maintained by the DLMS User Association, and comprises a suite of standards that have been adopted into the IEC 62056 series of standards. As a result, DLMS/COSEM has become a commonly used language in the AMI.

Focusing on protocols related to the vehicular privacy scenario, Open Charge Point Protocol [29] (OCPP) is the standard that defines an application level protocol to establish a communication between electrical vehicles' (EVs) charging stations and a central system. Its aim is to achieve interoperability between the manufacturers of charging points, network systems and IT back-end vendors. It was published by the Open Charge Alliance (OCA), which promotes the adoption of multiple communication protocols in the EV infrastructure. The protocol relies on a client/server architecture with SOAP over HTTP, based on requests and responses initiated by either of them. The main operation is charging up a vehicle: first, the user has to be authenticated to the point and it has to inform the central system. Once charging is complete, the station has to verify the identity of the user again, after which the transaction terminates.

OCPP does not specify any communication technology, as long as it supports TCP/IP connectivity. To achieve security concerns, the specification suggests the use of TLS to ensure confidentiality and integrity of data transmissions. However, it defines an authentication and access control mechanism through an authorisation list that is cached locally and managed by the central system. Additionally, OCPP provides transmission reliability techniques, like a heartbeat service that lets the central system know that a charge point is connected by sending a control message at a configurable time interval. In this point, it is important to remark that other communication protocols commonly used for process automation (e.g., Modbus, DNP3, IEC 104) can also be applied to retrieve information from charging stations.

A summary of all the characteristics of the nine protocols and standards in terms of reliability and security is presented in Table 2. Almost all protocols offer reliability when transmitting information at the data link layer. However, when it comes to security, not all implement robust mechanisms to cover the minimum security services, and only Zigbee and DNP3 comply with the aforementioned IEC 62351 standard. IEC 62056 can also be considered suitable in this regard, as it provides several mechanisms subject to different security policies described in the DLMS/COSEM specification. This is a universal language to simplify the communication with smart meters (whose data can even be mapped to other table-based models, like OSGP). On the other hand, in terms of complexity and use of computational resources, it is worth commenting that all protocols are expected to be easily run by each of the three types of smart meters considered in Section 2. The most expensive algorithm takes place in DNP3 with the use of RSA in TLS to perform signature and key exchange, which is more efficient in Zigbee Smart Profile 2.0, using elliptic curves.

Please cite this article in press as: J.E. Rubio, et al., Recommender system for privacy-preserving solutions in smart metering, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.03.008

**Table 3**
HAN protocols features comparison.

| | | Zigbee | Z-Wave | Wi-Fi | HomePlug 1.0 (PLC) |
|---|---|---|---|---|---|
| Reliability features | ACKs | ✓ | ✓ | ✓ | ✓ |
| | Timeouts | ✓ | ✓ | ✓ | ✓ |
| | Retransmissions | ✓ | ✓ | ✓ | ✓ |
| | CRC | ✓ | ✓ | ✓ | ✓ |
| Security features | Authentication | ✓ | ✓ | ✓ | ✓ |
| | Confidentiality | ✓ (AES-128) | ✓ (AES-128) | ✓ (WPA/WPA2) | ✓ (DES) |
| | Integrity | ✓ (HMAC) | ✓ | ✓ | ✓ |
| | Access control | ✓ (ACL) | | ✓ | |
| | History logs | ✓ | | ✓ | |
| Performance features | Range (m) | 10–100 | 10–100 | 10–1000 | 10–10 000 |
| | Data rate (bps) | 250 K | 40 K | 11 M | 14 M |
| | Latency (ms) | 30 | 30 | 5 | – |
| | Battery life | years | years | hours | – |

### 5.2. *Analysis of HAN communication protocols*

A similar analysis of the available communication protocols could be addressed in the context of the HAN. The goal is to connect devices of multiple kinds with the smart meter in the household, so the home owner can monitor them using a communication protocol which collects this information before sending it to the utility. Here, wireless communication technologies are often preferred due to their flexibility to add and remove new devices, their low cost and distance they can cover. However, wired protocols are still seen as more reliable to transfer data at higher rates.

Among the most used wireless protocols are Zigbee, Wi-Fi and Z-Wave. Zigbee [30] is the most used one in this domain: the Zigbee technology itself applied to this context, compared to the standard commented before (which makes Zigbee support the IP protocol), consists of a low power and low data rate protocol at lower layers whose architecture follows the master–slave model (i.e., the appliances acting as slaves and the Zigbee coordinator is the master). It presents a short delay latency and reliability measures (e.g., against collisions and packet loss), together with a highly secured connection using 128-bit AES encryption. Z-Wave [31] is similar to Zigbee, intended to form mesh networks of appliances. However, its network size is smaller and its data rate is lower (which also makes it cheaper). Then, Wi-Fi [32] adopts IP protocol and conforms to IEEE 802.11. It is implemented in the HAN to connect with devices at high rates (e.g., computers, TVs). The main disadvantage is the high power consumption, although it provides a better encryption and lower latency.

On the other hand, Power Line Communication (PLC) is one of the most widely used wired technologies in the HAN. It enables data communication over cables that are also used for power transmission. Homeplug [33] is a set of specifications of communication over the available home electricity wiring infrastructure, mainly used for high-rate applications (e.g., multimedia appliances) with reliability. Nonetheless, its data signals can suffer electromagnetic interferences, and it utilises the less secure DES encryption scheme.

Table 3 summarises this brief analysis of the discussed communication protocols in the HAN. Other performance features have been assessed, namely their range, data rate, latency and battery life, due to the intrinsic communication technology they rely on (unlike protocols in the AMI, where most of them specify an application protocol over an unspecified physical medium, which commonly can be PLC). As it shows, Zigbee is suitable option to achieve security requirements with a low consumption and moderate latency. Homeplug is also appropriate to cover long distances at a high rate but with low security.

## 6. Recommender system

So far in the paper, several privacy-preserving techniques to keep the data in the Smart Grid secure have been illustrated, and their suitability for different architectures and smart meters has been analysed. Apart from this survey, an overview of the main communication protocols that support these solutions has been provided. With this information, the goal is to create an expert system to advise a Smart Grid developer on how to introduce these privacy measures and the associated equipment. Table 4 lists all these features already discussed, which will be used by our recommender system.

### *General functionality*

This system should work as follows: the user (i.e., a SG developer) provides the application with information related to the desired communication model (from those listed in Section 2) and the optional compliance with some control requirements of those described in Section 3. Then, the system computes a recommendation of a privacy technique that suits these conditions giving its required smart meter type, according to Table 1. As mentioned, the system also finds the most suitable communication standard to transfer data from the user domain to the utility (specifically, one to connect the appliances with the smart meter in the HAN and another to connect the meter with the gateways in the AMI). To achieve this, the user has to set the requirements of each protocol in terms of reliability, security and performance, according to Tables 2 and 3.

Please cite this article in press as: J.E. Rubio, et al., Recommender system for privacy-preserving solutions in smart metering, Pervasive and Mobile Computing (2017), http://dx.doi.org/10.1016/j.pmcj.2017.03.008

**Table 4**
Features analysed so far, used for recommendation.

| Recommendation inputs | | Recommendation outputs |
|---|---|---|
| Architecture | Locally centralised<br>Distributed<br>Other | Privacy technique and corresponding smart meter type |
| Control requirements | Real-time performance<br>Sustainability<br>Dependability<br>Survivability | |
| Privacy scenario | Home<br>Industrial<br>Vehicular | AMI and HAN communication protocol |
| Reliability requirements | ACKs<br>Timeouts<br>Retransmissions<br>CRC | |
| Security requirements | Authentication<br>Confidentiality<br>Integrity<br>Access control<br>History logs | |
| Performance requirements | Range<br>Data rate<br>Latency<br>Battery life | |

The privacy scenario also has to be established, as each one concedes different level of importance to security requirements (as explained in the Introduction). In addition, the vehicular scenario is compatible only with certain protocols in the AMI (i.e., OCPP, DNP3, Modbus and IEC104) and does not use a HAN protocol in the context of charging stations.

*Application of a probabilistic approach*

To design and develop a recommender system with these characteristics, we have to take into account that the set of requirements the user demands for both the privacy technique and the associated communication protocols cannot always be satisfied. However, there are solutions whose features are closer to these conditions and hence they suit these needs better than the rest. Therefore, as we are dealing with uncertainty and our purpose is to provide a ranking of such solutions according to the extent of suitability, a probabilistic approach has been taken to model the relationship between nodes, which is, the probability of choosing a solution based on the proposed requirements. In the following, we describe the design of a Bayesian Network (BN) for the recommendation of the privacy technique, the AMI and HAN communication protocol.

Firstly, the recommendation of a privacy-preserving technique of a given architecture depends on the nine features defined in Table 1, namely: speed, storage efficiency, low communication overhead, lack of synchronisation, configurability, maintainability, fault tolerance, lack of error when aggregating data, and resilience. Each one has been modelled with a binary variable in the BN with equal probability, in such a way that the conditional probability of choosing a particular technique comes from the satisfaction or not of these features (as shown in Fig. 1). However, this results in $2^9$ different probabilities that the system should be provided with for each technique, which makes the modelling process infeasible. To overcome this drawback, in our approach the modelling of the technique's variables is based on the Noisy-OR canonical interaction [34]. This gate reduces the growth of a Conditional Probability Table (CPT) of a variable from exponential to linear in the number of parents, modelling a non-deterministic interaction among $n$ parent cause variables $X$ (i.e., the control features) and the effect variable $Y$ (i.e., the technique itself), so this one works as a deterministic OR gate: if all the parents variables $X$ are absent (i.e., every variable is set as false), the child variable $Y$ is also absent and thus the technique has no probability to be chosen. However, if a parent variable $X_i$ is present (and hence that property is satisfied) and other parent variables are absent, it has a probability $p_i$ of causing the technique $Y$ to be chosen.

In practice, in light of our analysis of privacy techniques in Section 4, the probability of choosing a singular technique has been modelled as a score: since the probability of the affirmative state of the respective variable goes from zero to one, different probabilities of causing a "YES" have been assigned to its different parent cause variables in case the corresponding features are satisfied. The sum of all these probabilities add to one but, since our main goal is to prioritise automation efficiency over the rest of features, a weighing of the score has been introduce to the parent variables. Specifically, *performance* requirements have an overall 0.6 score, which means that if any of the variables belonging to this category (i.e. speed, storage efficiency, low communication overhead, or lack of synchronisation) is checked (the user demands this

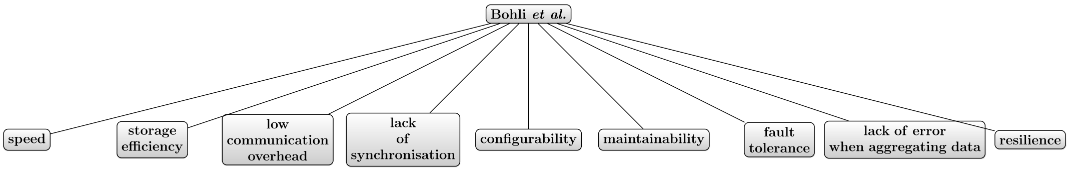


**Fig. 1.** Conditional probability for one technique variable in the BN.

characteristic), it causes a probability of 0.15 (0.6 divided into the four features) for a technique to be chosen if it satisfies that condition. Following the same rule, *dependability* and *survivability*, which group intrinsic security aspects, have been assigned with a 0.3 score. The rest (0.1) is given to *sustainability*. This way, for example, if the user seeks for a technique with low communication overhead, it results in a 0.15 probability of choosing LeMay et al. [9], but zero probability of selecting Lin et al. [18], according to Table 1. Intermediate level of compliance with this condition, as for Molina-Markham et al. [10], is resolved by giving the half of the maximum score.

Finally, the probability of selecting a particular technique (i.e., causing an effect $y$ in a variable $Y$) given a number of features required (i.e., a subset $X$ of causes which are present and whose probability of causing $y$ is $p_i$) is formally given by this formula [35]:

$$p(y|X) = 1 - \prod(1 - p_i). \tag{1}$$

Secondly, the same procedure has been applied to design the BN that enables selecting an appropriate communication standard for both the AMI and HAN context. This time, the cause variables are leveraged to model the reliability, security and performance aspects in Tables 2 and 3, so the user can pick which one is present for her desired standard. Therefore, the protocols enrol the child effect variables in the BN. However, concerning the quantitative modelling of the network, the weighing given to parent variables is slightly different: in the case of the AMI, there are protocols which are primarily designed for control operations despite they are able to take energy measurements from the smart meters (i.e., Modbus, DNP3, IEC 62056, IEC 104 and OCPP). For these, availability (which comprises the inclusion of ACKs, timeouts and retransmissions), integrity and confidentiality features prevail over the others (and in that order), which results in a cause probability of 0.35, 0.25 and 0.2, respectively. The rest of variables are given a 0.05 probability. However, for the protocols which are intended only for measurements, the weighing is the opposite: they give priority to confidentiality and then to integrity, availability and the rest of features.

With respect to the HAN, since all surveyed protocols comply with reliability features, as shown in Table 3, only security and performance features have been taken into consideration to recommend a communication standard. In this sense, confidentiality and integrity are more important than other aspects, so they are provided with a higher probability of choosing a particular protocol that complies with these features (specifically, 0.2 and 0.17 respectively, and 0.09 for the other seven properties in order to sum 1). Regarding the performance features, they can be assessed in such way that the better they can suit the property (e.g., having a longer battery life) the more score they get out of the maximum assigned (0.09 in this case). The same criteria can be applied to confidentiality, for instance, as some encryption algorithms are more robust (e.g., AES compared to DES).

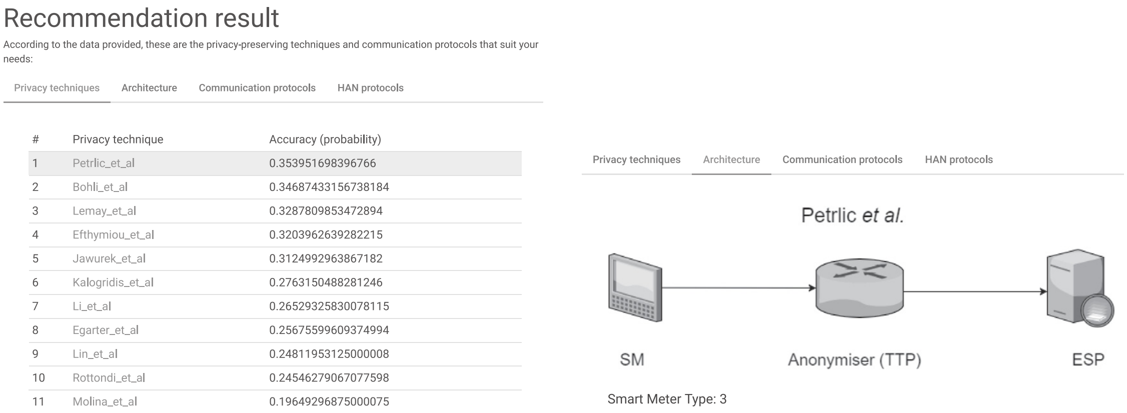
*Practical implementation*

After this process of design, we have modelled the BN in both qualitative and quantitative terms. On the one hand, we have child variables which represent privacy-preserving techniques and communication protocols; on the other hand, there are conditions to select these solutions, represented by parent variables, which have different influence on the selection. A real version of this model has been implemented for this paper using SMILE [36], a reasoning engine for graphical models, such as Bayesian networks, influence diagrams, and structural equation models. The described BN has been defined with GeNIe Modeler [36], a graphical user interface to SMILE. The satisfaction of criteria is controlled by a Java REST API which is called from a web page that offers the recommendation assistant.[1] Once the user chooses a communication model for the privacy technique and sets its requirements, the system filters the solutions according to these criteria, computes the probability of each technique and sorts them, finally giving a ranking and a visual representation of the chosen one (together with the suggested SM type), as shown in Fig. 2. In addition, the user gets advice on the communication protocols, filtering by their privacy scenario.

## 7. Conclusions and future work

The Smart Grid is a cutting-edge technology that brings with it several benefits for both operators and customers, although it presents some privacy issues when collecting and subsequently analysing consumption data in different

[1] The web is accessible in www.nics.uma.es/dev/recommendersystem/ with provisional authentication credentials, as described in the cover letter.



Recommendation result

According to the data provided, these are the privacy-preserving techniques and communication protocols that suit your needs:

| # | Privacy technique | Accuracy (probability) |
|---|---|---|
| 1 | Petrlic_et_al | 0.353951698396766 |
| 2 | Bohli_et_al | 0.34687433156738184 |
| 3 | Lemay_et_al | 0.3287809853472894 |
| 4 | Efthymiou_et_al | 0.3203962639282215 |
| 5 | Jawurek_et_al | 0.3124992963867182 |
| 6 | Kalogridis_et_al | 0.2763150488281246 |
| 7 | Li_et_al | 0.26529325830078115 |
| 8 | Egarter_et_al | 0.25675599609374994 |
| 9 | Lin_et_al | 0.24811953125000008 |
| 10 | Rottondi_et_al | 0.24546279067077598 |
| 11 | Molina_et_al | 0.19649296875000075 |

(a) Privacy technique ranking.

Smart Meter Type: 3

(b) Architecture of the suggested technique.

**Fig. 2.** Recommendation of a privacy-preserving technique and communication protocol using the web application.

installation scenarios: at home, when recharging electric vehicles and in the industrial context. New mechanisms have to be introduced to prevent against the extraction of sensitive information, taking into consideration their technical deployment constraints and the control requirements that the operators have to comply with. In this paper, we have conducted a thorough analysis on the main privacy-preserving techniques proposed in the literature that address this problem, studying their implemented PET and architecture, along with the smart meter needed to support their cryptographic algorithms. In addition, a comprehensive analysis of their performance features has been made, with the aim of providing guidance on the election of a solution. On the other hand, another studio of the communication protocols has been performed in order to ensure a set of security and reliability characteristics that support these privacy techniques in all the aforementioned scenarios. With all this information, we have described a feasible recommender system based on a probabilistic model, to systematically find a solution that suits the user needs. Future work will involve creating a larger database of other privacy mechanisms and different approaches for accurately designing a recommender system that prioritise customisable requirements defined by the user.

## Acknowledgements

The first author is supported by the Spanish Ministry of Education through the National F.P.U. Program under Grant Agreement No. FPU15/03213. In addition, this work has been partially supported by the Spanish Ministry of Economy and Competitiveness through the PERSIST (TIN2013-41739-R) and SMOG (TIN2016-79095-C2-1-R) projects.

## References

[1] R.R. Mohassel, A. Fung, F. Mohammadi, K. Raahemifar, A survey on advanced metering infrastructure, Int. J. Electr. Power Energy Syst. 63 (2014) 473–484.
[2] W. Wang, Z. Lu, Cyber security in the smart grid: Survey and challenges, Comput. Netw. 57 (5) (2013) 1344–1371.
[3] H. Souri, A. Dhraief, S. Tlili, K. Drira, A. Belghith, Smart metering privacy-preserving techniques in a nutshell, Procedia Comput. Sci. 32 (2014) 1087–1094.
[4] S. Finster, I. Baumgart, Privacy-aware smart metering: A survey, IEEE Commun. Surv. Tutor. 16 (3) (2014) 1732–1745.
[5] C. Alcaraz, J. Lopez, Analysis of requirements for critical control systems, Int. J. Crit. Infrastruct. Prot. (IJCIP) 5 (2012) 137–145.
[6] B. Commission, et al. Our common future, Chapter 2: Towards sustainable development, World Commission on Environment and Development, (WCED), Geneva, United Nation.
[7] M. Al-Kuwaiti, N. Kyriakopoulos, S. Hussein, A comparative analysis of network dependability, fault-tolerance, reliability, security, and survivability, IEEE Commun. Surv. Tutor. 11 (2) (2009) 106–124.
[8] J. Bohli, C. Sorge, O. Ugus, A privacy model for smart metering, in: 2010 IEEE International Conference on Communications Workshops, (ICC), IEEE, 2010, pp. 1–5.
[9] M. LeMay, G. Gross, C.A. Gunter, S. Garg, Unified architecture for large-scale attested metering, in: 40th Annual Hawaii International Conference on System Sciences, 2007, HICSS 2007, IEEE, 2007, 115–115.
[10] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, D. Irwin, Private memoirs of a smart meter, in: 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, ACM, 2010, pp. 61–66.
[11] M. Jawurek, M. Johns, F. Kerschbaum, Plug-in privacy for smart metering billing, in: Privacy Enhancing Technologies, Springer, 2011, pp. 192–210.
[12] T.P. Pedersen, Non-interactive and information-theoretic secure verifiable secret sharing, in: CRYPTO91, Springer, 1991, pp. 129–140.
[13] C. Rottondi, G. Verticale, A. Capone, Privacy-preserving smart metering with multiple data consumers, Comput. Netw. 57 (7) (2013) 1699–1713.
[14] F. Li, B. Luo, P. Liu, Secure information aggregation for smart grids using homomorphic encryption, in: Smart Grid Communications, SmartGridComm, 2010 First IEEE International Conference on, 2010, pp. 327–332.
[15] P. Paillier, Public-key cryptosystems based on composite degree residuosity classes, in: EUROCRYPT99, Springer, 1999, pp. 223–238.
[16] C. Efthymiou, G. Kalogridis, Smart grid privacy via anonymization of smart metering data, in: 2010 First IEEE International Conference on Smart Grid Communications, (SmartGridComm), IEEE, 2010, pp. 238–243.
[17] R. Petrlic, A privacy-preserving concept for smart grids, Sicherheit in vernetzten Sys. 18 (2010) B1–B14.
[18] H. Lin, W. Tzeng, S. Shen, B. Lin, A practical smart metering system supporting privacy preserving billing and load monitoring, in: Applied Cryptography and Network Security, Springer, 2012, pp. 544–560.

[19] G. Kalogridis, C. Efthymiou, S.Z. Denic, T.A. Lewis, R. Cepeda, Privacy for smart meters: Towards undetectable appliance load signatures, in: 2010 First IEEE International Conference on Smart Grid Communications, (SmartGridComm), IEEE, 2010, pp. 232–237.

[20] D. Egarter, C. Prokop, W. Elmenreich, Load hiding of household's power demand, in: 2014 IEEE International Conference on Smart Grid Communications, (SmartGridComm), IEEE, 2014, pp. 854–859.

[21] E.T.S. Institute, Open smart grid protocol, 2012. http://www.osgp.org/ (last retrieved in July 2016).

[22] Z. Alliance, Zigbee smart energy profile 2.0, 2013. http://www.zigbee.org/zigbee-for-developers/applicationstandards/zigbeesmartenergy/ (last retrieved in July 2016).

[23] A.N.S. Institute, ANSI c12.18, 2005. https://www.ansi.org/ (last retrieved in July 2016).

[24] T.M. Organization, Modbus, 2006. http://www.modbus.org/specs.php (last retrieved in July 2016).

[25] I.S. Association, DNP3 distributed network protocol, 2012. https://www.dnp.org/ (last retrieved in July 2016).

[26] Meter-Bus, EN 13757, 2011. http://www.m-bus.com/ (last retrieved in July 2016).

[27] I.E. Commission, IEC 60870-5-104, 2006. http://www.iec.ch/smartgrid/standards/ (last retrieved in July 2016).

[28] D.L.M.S, IEC 62056. http://www.dlms.com/ (last retrieved in July 2016).

[29] O.C. Alliance, Open charge point protocol 1.6, 2015. http://www.openchargealliance.org/ (last retrieved in July 2016).

[30] Z. Alliance, Zigbee, 2008. http://www.zigbee.org/what-is-zigbee/494-2/ (last retrieved in July 2016).

[31] Z.-W. Alliance, Z-wave. http://z-wavealliance.org/.

[32] W.-F. Alliance, Wi-fi. http://www.wi-fi.org/.

[33] H. Alliance, Homeplug 1.0, 2001. http://www.homeplug.org/.

[34] J. Pearl, Fusion, propagation, and structuring in belief networks, Artificial Intelligence 29 (3) (1986) 241–288.

[35] P. Kraaijeveld, M. Druzdzel, A. Onisko, H. Wasyluk, Genierate: An interactive generator of diagnostic Bayesian network models, in: Proc. 16th Int. Workshop Principles Diagnosis, Citeseer, 2005, pp. 175–180.

[36] L. BayesFusion, Smile engine and Genie modeler. http://www.bayesfusion.com/#!products/c1ixv (last retrieved in July 2016).