The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/159578

Please be advised that this information was generated on 2019-08-19 and may be subject to change.
New Directions in IoT Privacy Using Attribute-Based Authentication

Position Paper

Gergely Alpár†, Lejla Batina‡, Lynn Batten§, Veelasha Moonsamy†
Anna Krasnova†, Antoine Guellier†, Iynkaran Natgunanathan§

1Radboud University, The Netherlands, {firstname}@cs.ru.nl
2Deakin University, Melbourne, {first.lastname}@deakin.edu.au
3CentraleSupelec/Inria, France, {first.lastname}@centralesupelec.fr

ABSTRACT
The Internet of Things (IoT) is a ubiquitous system that incorporates not only the current Internet of computers, but also smart objects and sensors. IoT technologies often rely on centralised architectures that follow the current business models. This makes efficient data collection and processing possible, which can be beneficial from a business perspective, but has many ramifications for users privacy.

As communication within the IoT happens among many devices from various contexts, they need to authenticate each other to know that they talk to the intended party. Authentication, typically including identification, is the proof of identity information. However, transactions linked to the same identifier are traceable, and ultimately make people also traceable, hence their privacy is threatened.

We propose a framework to counter this problem. We argue that applying attribute-based (AB) authentication in the context of IoT empowers users to maintain control over what data their devices disclose. At the same time AB authentication provides the possibility of data minimisation and unlinkability of user transactions. Therefore, this approach improves substantially user privacy in the IoT.

Keywords
ABC, attribute-based authentication, Internet of Things, privacy, attributes

1. INTRODUCTION
The Internet of Things is becoming a part of our daily lives. Gartner predicts [24] that within five years there will be more than 25 billion devices that take part in the IoT. Because the IoT delivers services in a fast and convenient manner to users, it is viewed as a major business opportunity [22].

Most IoT technologies rely on a centralised architecture, in which sensors collect data from the world, communicate to computers, which in turn send the data to a central service. This architecture enables efficient operation and full control over the data processing and dissemination, which makes it an appealing approach for businesses. However, privacy concerns arise since users cannot control the information that is collected about them. Furthermore, organisations may be subject to legal problems in terms of privacy regulation. In our opinion and that of the authors of [22], in order for an IoT architecture to be sustainable, it must be decentralised and user-centric.

When devices communicate in the IoT, they often need to authenticate to each other, so that they know that they talk to the intended counterpart. Typically, authentication involves identification by means of a unique number or name and a token (e.g., a password or a cryptographic proof) that proves the validity of the identifier. On the one hand, identifiers make authentication easy, provided that the verifier has access to a database of these identifiers. On the other hand, identifiers make all the transactions carried out by a particular device trackable. Moreover, in many cases devices are associated with an individual; so, indirectly people become traceable as well. Based on the increasing amount of user data, organisations, including advertisers, can build profiles about people. This is obviously harmful to people's privacy, and in many countries, the use of data for tracing and profiling is contrary to legislation governing the so-called secondary use of data [26].

With regard to user control, data collection can be effected in two ways in the IoT [25]. First, there are devices which the user can control how information is collected. Examples include wearables and smart home units. Second, there are devices for which the user has no control over the data collected; these include sensors and surveillance cameras in public places for instance. This technical distinction has to be taken into account when defining privacy and when designing new IoT technologies.

IoT applications often rely on data analysis tools which can recognise patterns and extract further useful information from already collected data. For instance, analysing
energy usage of a neighbourhood can help distribute power efficiently within a city. Such analysis tools are also suitable to discover personal information that users may want to keep private. An important way to decrease the privacy risk caused by data analysis is to reduce the amount of data collected, as proposed by the European Parliament and Council in their data minimisation principle [13].

Research in identity management [8, 23] shows that attribute-based authentication can realise data minimisation with possible user control. Attributes, characteristics or qualifications of entities, are embedded in a cryptographic container which can be used for authentication purposes. Since attributes such as the brand of a device or the nationality of a person, can be anonymous, authentication can also be anonymous. In fact, authentication should, and can, be realised by using the minimum amount of information required to successfully complete a transaction.

1.1 Our contribution

We recognise the importance of the right of the individual to have control of their own data and not to have their transactions linked and tracked within the IoT. On the other hand, the threat of mass surveillance and of linking users and their devices to transactions is very real within the IoT.

In this research, we present a framework for providing the user with the ability to avoid linkability and at the same time, maintain control of their data. Our solution is based on AB authentication instead of identification to guarantee authenticity in the communication among devices in the IoT. We demonstrate the feasibility of our framework by introducing it into a common use case, in a home environment.

We also argue that AB credentials should be considered in every scheme where authentication is needed, as it will provide the user with increased privacy and control with respect to their personal information.

2. RELATED WORK

In this section, we present some of the recent work related to authentication and privacy in IoT, and the uses of attributes.

2.1 Authentication and Privacy in IoT

Position papers on IoT such as [30] and [14] often include discussion on authentication and privacy. In fact, many such papers point out the authentication and privacy problems arising from ubiquitous presence of sensors and devices, and the subsequent analysis of massive amounts of collected data.

As observed by the authors in [2, 3, 18, 35], one of the capabilities of IoT is to allow sensors to collect information from their surroundings, record and process it. With an ongoing surge in the number of inter-connected devices [24], the tendency of collecting more information than required for the provision of a service is on the rise.

Conversely, one of the business drivers of the IoT is expressed as “improved customer retention and more targeted selling” [18]. This makes way for a strong tension between the need for massive data collection, processing and communication required for the services of IoT on the one hand, and privacy protection of individuals on the other [16, 18]. Moreover, multiple studies have shown that collection of even seemingly innocent data can lead, with very high probability, to the identification of individuals [20, 27, 28].

2.2 Identifiers in IoT

While interacting with an IoT architecture, users perform identity management [7] implicitly. They produce personal information and leave traces mostly bound to their identity. For instance, adjusting your house’s temperature using a mobile phone requires that the system knows that the instruction came from a legitimate party. Typically, a system stores a lot of information about its users. The set of all this data with respect to a particular individual is her identity in this system. In most cases at least one of these pieces of information acts as an identifier, that is, a direct link between the individual and the system.

The most common way to authenticate individuals and devices and authorise them for services in today’s Internet is by using Federated Identity Management [10]. These solutions involve an Identity Provider and one or more Service Providers. When a user needs to authenticate to a service, the identity provider intervenes to assert the user’s identity to the service provider. This allows for flexible authentication and the decoupling of identities and services. However, the Identity Provider, being involved in all transactions, can trace users connecting to services.

2.3 Use of Attributes

Authentication protocols often reveal more about the user than necessary. Indeed, in many cases, only an assertion on the user’s attributes is really needed. For instance the only information that may matter is that the user belongs to a registered service provider, or has a sufficient clearance level. In this sense, identity-based authentication does not comply with the data minimisation principle as prescribed by Directive 2006/24/EC of the European Parliament. Existing identity-based authentications solutions however already admit that attributes are what really matters, since after obtaining the user’s identity, the SAML standard [1] (in particular) allows the service provider to request specific attributes (or claims) about the user [10].

The ABC4Trust project [23] demonstrated that the attribute-based credentials (ABC) technology is an adaptable means of realising both flexible and privacy-preserving authentication. Indeed, the identity provider needs only to be online when the credential is delivered to the user; this prevents the identity provider from profiling users based on their authentication patterns, and from impersonating them. In addition, the user has control over which attributes are used thanks to the selective disclosure mechanism of the ABC constructions.

The main cryptographic schemes putting ABCs in practice are U-Prove [21] and Idemix [9]. They have been implemented on smart cards [19, 32] resulting in a U-Prove authentication processed under 1 second, while that in Idemix is between 1 and 1.5 seconds. These technologies have also been placed into light-weight infrastructures in [5, 6, 23]. However, none of the previous research projects proposes to adopt ABC technologies within the IoT.

3. NEW DIRECTIONS IN IOT PRIVACY

3.1 Privacy Threats

The usual approach in cryptography and security areas is to define an adversary by his goals and capabilities. However, in data collection the initial goals may be set with the
best intentions of a data collector, but possession of the accumulated data can be transferred rightfully (or not) to a data processor that does not share good intentions of the original one. Hence, the adversarial goals are rendered irrelevant during data collection phase but can be of impact during data processing and data dissemination phases.

Data over-collection [17] and the always-identify paradigms described in the Introduction are other existing privacy threats that become increasingly harder in IoT. Coupled with linkable transactions, that is transactions the origins of which are known to be the same [15], these threats enable extensive user profiling. For instance, service providers can create a fingerprint of an individual based on the types of devices he utilises, as presented by van Deursen [31] in the case of RFID tags. This may give a lot of information about the device owner; moreover, this fingerprint acts as a new identifier, making it impossible for the user to remain anonymous later.

3.2 Defining Privacy in IoT

We argue that the lack of control is one of the central problems in IoT in terms of privacy. Individuals become a part of a pervasive computing system, and by that, they generate a lot of personal information while having only limited oversight and control over data collection and processing. We adapt the definitions by Alan Westin [33] and Ziegeldorf et al. [35] as the former does not address the problem of data collection and data processing, and the latter places too much responsibility on the data subject.

**Definition 1.** Privacy is the right of individuals to determine for themselves when, how and to what extent information about them is collected, processed and communicated; that includes individuals having

- the right to determine these aspects within their area of control explicitly;

- trust that the right above is respected when control is not possible.

Following taxonomy of privacy by Solove [26], privacy threats are raised during data collection, data processing and data dissemination activities and intrusions. Because malicious intrusions form a whole body of work separate from privacy, we do not discuss them here. Specific to IoT, data collection is happening on a massive scale, much of it is collected by sensors without any active initiation from the user, thus often leaving users unaware of this process. The large amount of data and the high number of potential data sources increases severity of privacy threats at data processing and dissemination activities compared to the today’s situation in the Internet.

In the present, control over data processing and dissemination activities is largely in the hands of lawyers and policy makers. From the users’ perspective, technologies providing them meaningful control are virtually non-existent. Although some Privacy-Enhancing Technologies (PETs) can provide a solution to this problem, their use is left at the discretion of companies and agencies, who only employ them when regulation mandates so.

The control over data collection, on the contrary, is partially in the hands of users. Indeed, although control is impossible in the presence of sensors (such as surveillance cameras) that collect user data in a passive manner, possibility to control appears when the user or any of the devices acting on his behalf (and that are under his control) are actively engaged in a communication.

Altogether this makes privacy-friendly data collection an important stepping stone towards achieving privacy in the IoT. First of all, this is the only area in which it is feasible to implement some level of technological control from the user’s perspective. Secondly, the increased amounts of collected data leads to correspondingly increased levels of threats during data processing and dissemination. Thus, addressing this particular point – data collection – one can provide significant increase in the privacy protection level of an individual user.

4. REALISING IOT USING ATTRIBUTES

4.1 Attributes and Attributes-based Credentials

Conceptually, attributes are properties or qualifications of an entity. In practice, an attribute can be anything that can be described as a bit-string, such as a name, a date of birth or a cryptographic token. In this respect, attributes generalise the notions of identifiers and roles.

An ABC [9, 21] is a cryptographic container of attributes. Similar to a traditional X.509 certificate, it is issued by a trusted party, and is bound to a specific entity via this entity’s secret key. That is, the issuer cryptographically signs a token consisting in the concatenation of the attributes and a commitment to the entity’s secret key. For instance, a service provider may issue to each of its clients a subscription ABC containing a client number, a specified purchase level and the start date of the subscription. The issuer is trusted for verifying that the attributes in an ABC authentically belong to the entity. Having an ABC, an entity can show its attributes and prove that they are signed by the issuer. This is done using (non-interactive) zero-knowledge proofs to guarantee that no other information (e.g. some leakage about the secret key) is revealed.

In addition to the traditional unforgeability property, ABCs come with many privacy-enhancing mechanisms [4, Chapter 3]. First of all, because the proving of attributes is performed in zero-knowledge, the verifier does not learn the (commitment on) the secret key of the prover in the process. Even better, some ABCs scheme have the multi-show unlinkability property that prevents a verifier from linking two showings of the same ABC by the same entity (except if the content of attributes themselves leak information). Secondly, the selective disclosure functionality allows an entity to demonstrate only an arbitrary subset of the attributes contained in its ABC. Continuing the subscription ABC example from above, a client can choose to show only its purchase level, which may grant discounts and advantages, but not its client number.

4.2 Privacy using Attribute-based Authentication

Privacy threats can be thwarted thanks to the privacy-friendly properties of ABCs. First of all, the always identify paradigm can be avoided by making all authorisation decisions depend only on attributes, not identities. In most cases, an assertion on the user is sufficient. ABCs allow a combination of assertions to be made, by simply showing
multiple attributes (e.g. the user paid a subscription, and is an adult). It should be noted that in cases where the authentication is absolutely necessary, e.g. when an individual wants to communicate with one specific entity, ABCs can also be used. The ABC must simply contain an attribute representing the identity or public key of its holder. In summary, attributes achieve the same functionality as identifiers with the same security level and also provide unlinkability.

By design, the selective disclosure functionality of ABCs prevents data over-collection. Actually, an ABC can be shown (or proved) without disclosing any of the attributes it contains. Thus, the only leaked information is that the holder of the ABC was accredited by the trusted issuer. This information is sufficient in a scenario where resources can only be accessed by members of some institution, and that institution is the issuer of ABCs.

The multi-show unlinkability property is designed to prevent linkability of transactions. If an ABC is proved without disclosing any attributes no linking at all is possible. This is the best case scenario for the user. When showing one or more attribute(s), a user leaks some information that can be characterised in terms of $k$-anonymity [29]. For instance, consider a 30 years old male user that proves his gender and age to some service provider. From the service provider’s point of view, who did not have any a priori information, the user could be any individual in the context’s population before the showing of attributes: the anonymity set is maximal. After learning the gender and age of the user, the service provider can restrict the potential individuals to 30 years old males in the population: the anonymity set is much smaller. More generally, when a user shows a set of attributes, it is $k$-anonymous among other users that have (and show) the exact same attributes, and the service provider can not distinguish between two transactions from the same user or from two users in the anonymity set. In the worst case, if the attribute is identifying the user uniquely ($k = 1$), the service provider can link of all transactions.

To model possible information privacy harms, we use Solove’s taxonomy [26, 35] as a starting point. Solove discusses four groups of harmful activities with regard to privacy, from which three (information collection, processing and dissemination) are related to data flow while the fourth one is not (invasions).

As a starting point, to identify data flows where AB authentication is relevant, we again use Solove’s taxonomy of privacy harms [26] (data collection, processing and dissemination). This approach is similar the work of Ziegeldorf et al. [35], but the authors do not address the authentication aspects of the data flow. Therefore, we develop their model further. Besides the three groups above, we consider additional activities in the IoT. First, the user is surrounded by sensors, which can be divided in two groups. Some sensors can be controlled by the users, while others not. Second, processed data is partly used by the data subject within data dissemination. Third, some part of the whole data flow is affected by credential providers that determine information access of and related to the data subject.

In Figure 1 the data flow and possible attribute-based authentication are represented. First, the data is sensed around the data subject. Second, some of this information is sent to the service provider. Then, the data processor, after processing (e.g., aggregate, search, analyse) the collected information, disseminates it. Finally, the access to disseminated information may be restricted to a certain group of entities (devices or people).

AB authentication can take place at four points in this model – denoted by red spots:

1. Sensing. User-controlled sensors, such as wearables and smart-home devices, can communicate with each other locally. In this case, the authentication can be based on attributes.

2. Collection. Sensors, when communicating with a service (data processor) should authenticate to the service. Possibly, the authentication can be mutual (i.e. the service also authenticates to the device). In both cases, it can also be based on attributes.

3. Dissemination. After collection and processing, data processors should communicate restricted information (e.g. individuals’ personal information) only after AB authentication of the receiving party.

4. Issuing. Issuing itself is a part of the AB technology, which provides a new credential to an entity. To make sure that this entity is entitled to the particular credential, the credential provider has to authenticate the entity. This can be done with conventional identification, or by attribute-based authentication. The latter is called in [8] a dependent credential.

5. USE CASE SCENARIO

5.1 Smart Home

Consider a scenario involving a home owner, hereby called user, in possession of multiple smart devices and home appliances. The user’s devices, all under his control, are connected to the Internet via a hub (installed within the home). The devices autonomously connect to remote services in order to push data to or pull data from them. Depending on their business model, some of these services may require users to get a subscription before consuming the service.

Therefore, a form of authentication is necessary for the service to check if the devices belong to a valid user (i.e. one that paid for a subscription). In a traditional IoT setup, each user would have an account on the service’s platform, where all his devices would be registered. A service would accept requests only from devices that belong to a registered user.
5.2 Privacy Threats

This framework carries several of the security and privacy risks presented in Section 3, as described in [34]. The most notable potential breach is the tracking of user among the services ecosystem via his fingerprint (based on the set of devices he owns). This fingerprint basically acts as an user identifier of pseudonym and makes all his transactions linkable, even if he does not disclose his identity. Moreover, if a user conceals his identity to a particular service $S$ but not to another service $S'$, $S$ can re-identify the user with the help of $S'$, since both services have fingerprint for that user. Together with data over-collection, common for current practices, allows for extensive user profiling. In the discussed use case profiling is performed within such a sensitive sphere of life as one's behaviour at home.

5.3 Applying ABC

Using ABCs, one can avoid these issues. At the same time, ABCs allow the user to have full access to the services he paid for. With respect to Figure 1, ABCs are mainly relevant in the issuing and collection phases, for the given use-case. After an out-of-band authentication, the user and his devices will first be issued a set of ABCs. These credentials permit the devices to authenticate to services on behalf of the user. Issuers can be various parties, including trusted third parties and the service providers themselves. ABCs can include attributes describing the model number of the device or the subscription identifier of the user.

Then, the collection phase here consists of devices making requests to the services. Before processing any such request, services require that devices authenticate by proving possession of an AB credential. When authentication succeeds, the service is assured that the devices proving to have AB credentials are genuine and belong to a valid user. At the same time neither the user nor the device identity is revealed. Information collected by the services during authentication and transaction processing includes the nature of the request itself, some meta-data such as the time and issuer identity, and the attributes disclosed by devices. We assume that the amount of information revealed by meta-data (e.g. actual IP and MAC addresses) is reduced using other privacy enhancing technologies, like anonymous communication systems [11], [12].

5.4 Reduced Privacy Risks

The attributes disclosed by devices determine how much information is revealed to a service provider. To realise data minimisation, in some cases, no attributes should be disclosed at all: the simple fact that the device holds an ABC may show that it belongs to a valid user. However, other providers may supply different levels of a service depending on some additional information in the form of attributes; for example, a remote car diagnostic service may require the model or the manufacturer of the device. If service providers need this information to process requests, this information leakage cannot be prevented.¹

Disclosing the minimum set of attributes prevents the linking of a device’s requests. Indeed, the multi-show unlinkability property of ABCs ensures that the only information that can be collected is that by the revealed attributes. This holds even in the worst case scenario: If the issuer

¹The only solution would be not to use these services at all.


