An Approach for Multi-Level Visibility Scoping of IoT Services in Enterprise Environments

Qian Zhou, Omkant Pandey, and Fan Ye

Abstract—In IoT, what services from which nearby devices are available, must be discovered by a user's device (e.g., smartphone) before she can issue commands to access them. Service visibility scoping in large scale, heterogeneous enterprise environments has multiple unique features, e.g., proximity based interactions, differentiated visibility according to device natures and user attributes, frequent user churns thus revocation. They render existing solutions completely insufficient. We propose Argus, a distributed algorithm offering three-level, fine-grained visibility scoping in parallel: i) Level 1 public visibility where services are identically visible to everyone; ii) Level 2 differentiated visibility where service visibility depends on users' non-sensitive attributes; iii) Level 3 covert visibility where service visibility depends on users' sensitive attributes that are never explicitly disclosed. Extensive analysis and experiments show that: i) Argus is secure; ii) its Level 2 is 10x as scalable and computationally efficient as work using Attribute-based Encryption, Level 3 is 10x as efficient as work using Paring-based Cryptography; iii) it is fast and agile for satisfactory user experience, costing 0.25 s to discover 20 Level 1 devices, and 0.63 s for Level 2 or Level 3 devices.

Index Terms—Internet of Things, Service Discovery, Security, Privacy

1 Introduction

In Internet of Things, what services from which nearby devices are available, must be discovered by a user's device (e.g., smartphone) before she can issue commands to access them [1]. The visibility of services must be "scoped" to authorized users only, i.e., only the authorized service subset/variant should be "visible" to the user device. Service visibility scoping in large scale, heterogeneous enterprise environments has multiple unique features, which render existing solutions [2], [3], [4], [5], [6], [7] significantly ineffective or completely insufficient.

First, IoT interactions via short-range radios are largely physical proximity based. A user mostly discovers and controls nearby devices, such that desired physical changes would occur in the local environment (e.g., door unlocking, higher ambient brightness, lower room temperature, louder music volume in the current room). A centralized server does not know which devices are around the user device; accurate user location requires more complexity in localization capability. As a result, solutions [2], [3], [4] using centralized servers are unfit for such proximity based discovery.

Second, multiple services of different natures usually exist around a user, and all of them need to be discovered. Three examples (of different natures) are sketched here and will be introduced in detail in Section 4. Imagine a university or corporate campus: 1) public utilities (e.g., thermometers in aisles) may be discovered to everyone, even visitors; 2) yet most devices should be conditionally disclosed according to users' non-sensitive attributes (e.g., expensive equipment in an office should only be visible to the employees); 3) still there are services customized for specific populations of sensitive attributes that should be discovered with privacy preservation (e.g., a vending machine dispenses counseling/psychological support flyers, hidden within regular magazines to students of needs). Existing distributed schemes [5], [6], [7] are restricted to one type and unfit for enterprise IoT with numerous, heterogeneous services; the simple piling up of multiple schemes for heterogeneous service discovery is far from sufficient because those schemes can have conflict or redundancy.

Third, user churns are frequent in large scale enterprise environments. There are employment entry/termination or promotion/demotion/rotation all the time. All these may require updating messages to large numbers of devices. E.g., upon an employment termination, all the devices accessible to the employee (e.g., small numbers specifically to her but large numbers accessible to all employees) need to be informed of that churn. Such changes must be propagated and effectuated quickly and efficiently. Otherwise, new users may fail to discover and access services timely, while unauthorized users continue to see service access which they are no longer eligible for; neither is desirable.

We propose *Argus*, a distributed, proximity-based IoT service discovery algorithm that offers three levels (**public**, **differentiated**, **covert**) of visibility in parallel. Each level is built on top of its prior level with minimum extension, and together they achieve efficient, harmonious, concurrent discoveries covering IoT services of different natures. Also, it minimizes the huge overhead resulting from authorization updating (e.g., user addition/revocation), which is the scalability bottleneck in this context, and fits well with an enterprise scale. We claim our contributions as follows:

- We identify unique requirements of visibility scoping in enterprise IoT, and design a distributed 3-in-1 algorithm for scalable, concurrent, fast service discovery at three levels: **Level 1: public visibility** where services are identically visible to everyone; **Level 2: differentiated visibility** where service visibility depends on users' *non-sensitive attributes*; and **Level 3: covert visibility** where visibility depends on users' *sensitive attributes* that are never explicitly disclosed for the sake of privacy.

- We identify Level 2 as the scalability bottleneck in IoT service discovery. It is worth noting that, our Level 2 using conventional cryptography (e.g., ECDSA) is found **10x** as **scalable and computationally efficient** as more recent, fancier crypto like Attribute-based Encryption (ABE) [8].
- Argus Level 3 goes beyond existing privacy-preserving discovery work, and realizes **indistinguishability** that attackers cannot even know Level 3 is happening, let alone prying into user privacy in Level 3. Besides, our Level 3 is found **10x** as **computationally efficient** as work based on Pairing-based Cryptography (PBC) [9].
- We implement Argus's all three levels, and ABE, PBC for comparison. Extensive analysis and real experiments on a 20-node testbed are conducted. Besides the high scalability and computational efficiency mentioned above, it is found secure and fast, costing **0.25 s** to discover 20 Level 1 devices, and **0.63 s** for Level 2 or Level 3 devices.

2 Models and Assumptions

2.1 IoT Nodes in Enterprise Environments

Node Types. There are three types of nodes: the backend, subject devices and objects. Subjects (i.e. users) use subject devices (e.g., smartphones) to operate objects (i.e. IoT devices). As is widely used in practice, the backend is *not* a single server, but a hierarchy of servers run by the admin to manage registered subjects/objects; it realizes a chain of trust, and resists collapse under the load and a single point of failure. Each subject/object has multiple attributes (e.g., a user's position, department; a device's type, functions), and must register on the backend (e.g., manually out of band), which is a common requirement in real enterprises. It obtains an ID, private key, public key certificate (CERT) and attribute profile (PROF) signed by the backend's private key.

Resource & Power. Objects may have different resources. Resource-poor objects (e.g., temperature sensors, smoke detectors) have poor computing performance and are usually battery-powered. Resource-rich ones (e.g., door locks, HVAC, surveillance cameras) have fairly good computing performance and wall power; they can run public-key algorithms at reasonable speed. It is common in reality that vendors equip more important, expensive IoT devices with better hardware, so we assume Level 2 and 3 objects are resource-rich (more discussion can be found in Section 11).

Network Connectivity. Objects may have diverse communication interfaces, e.g., WiFi, Bluetooth, ZigBee. We focus on the security design above the network layer, and assume network connectivity exists among nodes (e.g., via bridging devices with multiple radios) in proximity and multi-hop routing is available [10]. The network formed by subject devices and objects is called *ground network*.

2.2 Service Visibility

Access Control Policy. The backend stores and manages access control policies about what services a subject can access on an object. Given the large numbers of subjects/objects, the policies are frequently defined on categories using attribute predicates, not just individual identities, to avoid inefficient enumeration. E.g., *[subject: position=='manager'; object: type=='door lock' && room_type=='conference'; rights: {'open'; 'close'}]* denotes that all managers have open/close access to the door locks on conference rooms.

There are two types of attributes: ***non-sensitive attributes*** can be safely included in signed credentials (e.g., PROF) and publicly propagated, e.g., a regular employee's position, a corridor light's make/model; in contrast, ***sensitive attributes*** must be kept on need-to-know basis and secret from the public or unauthorized subjects/objects, e.g., a student/employee's financial (broke), medical status (disabled, depressed).

Service Visibility. Visibility scoping policies are congruent with access control policies such that subjects and their devices should only "see" the services they are authorized to access. We call both of them *policies* for short. To this end, a subject device sends queries in the ground network, and nearby objects return their profiles (PROF), revealing their service information to the subject. Depending on the subject's access rights, an object may return differentiated variants of service information.

2.3 System Scale

We describe the typical scales of several aspects of enterprise IoT, where $10^i, i \in \mathbb{Z}$ denotes order of magnitude. E.g., 10^0: several, 10^2: hundreds. They are intended to give a rough sense, and actual systems may have larger/smaller scales.

1) huge subject/object amounts. Google search shows that the numbers of employees are: Google 98K, Apple 132K, Amazon 613K; so subject amount is $10^4 \sim 10^5$. According to the field study in [11], a typical lab/office may have $\sim$30 objects, and even a 2-story building may have $\sim$2K objects. A subject usually has access to N objects, which is around 100 (if she can access a few rooms) or up to 1K (if she can access objects in multiple buildings, e.g., on campus); thus N is usually $10^2 \sim 10^3$.

targets in one discovery operation. Though N objects are discoverable to a subject, the number of those in her proximity at any time thus the targets in one discovery operation is mostly not more than dozens. We denote it as n (10^1).

2) subject/object categories. A subject category (according to a predicate on subjects' *non-sensitive* attributes, e.g., "all students in CS Department, University X") has α (usually $10^0 \sim 10^2$, and possibly $\geq 10^3$) subjects (e.g., group: 10^0; class: 10^1; department: $\geq 10^3$; college: $\geq 10^4$). An object category (according to objects' *non-sensitive* attributes) has β objects, and β has a similar range as α (e.g., "all devices in Room Y": 10^1; "all lights in Building Z": $\geq 10^2$).

3) secret groups. If a policy allows subjects with certain *sensitive* attributes to discover objects with certain *sensitive* attributes, then they belong to one secret group. In reality, the persons with sensitive attributes (e.g., disability) in an enterprise and the objects serving them should not be too many, and a secret group has size γ ($10^0 \sim 10^1$ or 10^2).

4) frequent subject/object churns. In enterprises, employee entry/exit or promotion/demotion/rotation, and device installation/decommissioning happen all the time. All these may affect the access rights for individual/category subjects, and should be effectuated quickly and efficiently. E.g., when a subject leaves, all the N ($10^2 \sim$

10^3) objects she could access should be notified to reject her future discovery attempts.

3 Design Goals

Secrecy. We consider three subitems: ***1) service information secrecy for services behind walls (required by Level 2 and 3).*** Subjects should be kept from discovering services behind walls (i.e. services in the rooms they cannot enter) via sending queries from outside, otherwise the presence of indoor belongings is exposed to outsiders. Though services behind walls are physically invisible, they can be discovered by outsiders if no security countermeasure is taken because many radios penetrate walls. Protecting physical security and accessibility is not an IT problem, thus out of the scope.

2) sensitive attribute secrecy (required by Level 3). Sensitive attributes of a subject/object should only be disclosed to an authorized object/subject during service discovery.

3) indistinguishability (required by Level 3). Whether a subject/object has a sensitive attribute, or whether Level 3 discovery is happening, should be unknown to attackers.

Authenticity & Integrity & Freshness. Authenticity and integrity ensure that an entity is indeed the claimed one and a message is not forged or altered. Freshness means that a message is recently generated and not a replayed one.

Scalability. Upon any change in the backend database (e.g., policy addition, subject removal), the overhead of propagating and effectuating the changes to affected subjects/objects (i.e. updating overhead) should be minimized to make the system fit for an enterprise scale.

Responsiveness. The number of compute intensive operations (e.g., public-key) should be small enough such that discovering the n (defined in Section 2, usually 10^1) services in a user's proximity can be completed within 1 second to achieve positive user experience [12].

Non-Goals. Attackers may know the existence of physically visible devices (e.g., they see door locks). They may physically steal objects' service information or phish subjects' sensitive attributes; subjects may inadvertently leak service information they discovered or sensitive attributes. Besides, attackers may launch DoS attacks. We do no address those issues in this paper.

4 Backend Operations and Overview

4.1 Backend Operations

Bootstrapping. A subject or object X must first register on the backend out-of-band. This is a common requirement in real enterprises. The backend adds its information to the database, and issues an identity ID_X, private key K_X^{pri}, public key certificate (CERT) and possibly multiple attribute profiles (PROF) to X. The admin's public key K_{admin}^{pub} is also delivered. CERT and PROF are *signed by the admin* and cannot be forged or altered. A subject PROF lists the subject's *non-sensitive* attributes and can be publicly disclosed; an object PROF lists provided functions (thus service information) besides the object's *non-sensitive* attributes.

Fellows. We call subjects and objects in the same secret group *fellows*. The backend issues fellows with one symmetric group key (denoted as K_i^{grp} for secret group i). A subject or object is possibly in multiple secret groups.

Levels. An object gets its secrecy level defined (1, 2, or 3) and must keep that to itself. The level usually depends on the device type but can be configured by the admin.

1) Level 1. (no secrecy) These objects are usually resource-poor, and offer publicly accessible services identical to everyone (e.g., thermometers in aisles), or are behind walls but not worth hiding (e.g., lights in offices). Their service information needs no encryption, but is signed by the admin for integrity protection.

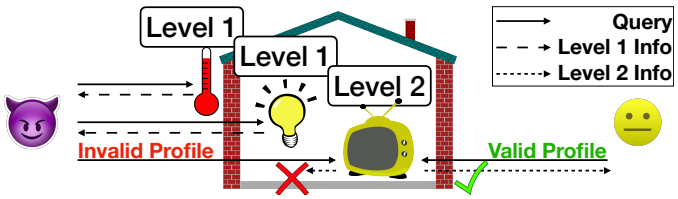


Fig. 1. Discovery in Level 1 and Level 2

2) Level 2. (service information secrecy) These objects are resource-rich for public-key operations. They are behind walls and must resist discovery by outsiders: objects return different service information (encrypted) according to subjects' non-sensitive attributes (in subjects' PROFs).

Example. In Fig. 1, a multimedia device in an office checks query senders' PROFs, and returns encrypted information only to the office employees, not outsiders.

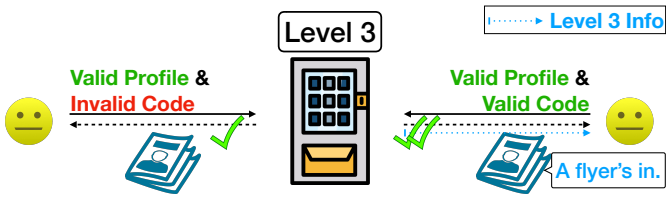


Fig. 2. Discovery in Level 3

3) Level 3. (service information secrecy+sensitive attribute secrecy+indistinguishability) These objects are also resource-rich, and they *secretly* offer customized services to special populations (i.e. subjects with sensitive attributes) while posing as Level 2 objects to other populations.

Example. In Fig. 2, Imagine student $\mathcal{S}$ with learning disability shows his diagnosis to the university, and is put in the corresponding secret group. When $\mathcal{S}$ uses a campus magazine machine $\mathcal{O}$, if $\mathcal{O}$ happens to have a sensitive attribute "machine serving students with learning disability", they will secretly confirm that they are fellows using a "code", then $\mathcal{O}$ dispenses medical/counseling/university-policy support flyers to $\mathcal{S}$, hidden within regular magazines so others will not know Level 3 discovery is happening. This is one way to offer special populations useful information when they are using daily IoT services (e.g., buy magazines/newspapers). Others will get "clean" magazines when using $\mathcal{O}$, and think $\mathcal{O}$ is Level 2. As a result, others cannot identify $\mathcal{S}$ as a sensitive attribute owner, or $\mathcal{O}$ as a Level 3 machine serving special populations.

Profile. A Level 2 object providing m different services gets from the backend an attribute-based ACL containing m PROF variants: $\{pred_i, PROF_{O,i}\}, 1 \leq i \leq m$, where $pred_i$

is a predicate on subjects' non-sensitive attributes (e.g., *[position=='manager' && department=='X']*). If a subject's PROF matches $pred_i$, $\mathcal{O}$ will send encrypted $PROF_{O,i}$ to her.

A Level 3 object in m' secret groups gets m' PROF variants: $\{K_i^{grp}, PROF_{O,i}\}, 1 \leq i \leq m'$, where K_i^{grp} is the symmetric group key of group i. After $\mathcal{O}$ confirms that a subject possesses K_i^{grp}, it will return encrypted $PROF_{O,i}$. A subject in n' secret groups gets n' group keys. Note that even a subject with no sensitive attribute still gets a (fake) group key, called *cover-up key* (details in Section 6).

Access Control Policy Update. The admin may need to update the backend database at any time, and possible operations include adding, removing, changing a subject/object individual or category and the access rights. Changes on the backend may need to be immediately propagated to the ground network and effectuated on the affected subjects/objects, such that newly authorized subjects can discover services, or de-authorized subjects stop seeing previously visible services.

4.2 Overview of Three-Level Discovery

Argus is a 3-in-1 algorithm which discovers services in three levels. We sketch the discovery process in each level.

Level 1 Discovery. The discovery of Level 1 objects is a typical 2-way data discovery/retrieval [13] process in the ground network. In Fig. 5, subject $\mathcal{S}$ broadcasts query QUE1 to objects around, and a Level 1 object $\mathcal{O}$ sends back its profile (PROF) in plaintext in response RES1. QUE1 carries random R_S for objects to detect duplicate queries. Because PROFs are signed by the admin, their integrity is ensured.

Level 2 & Level 3 Discoveries. They have a similar workflow (Fig. 5). First, $\mathcal{S}$ broadcasts query QUE1, a Level 2 or Level 3 object $\mathcal{O}$ returns response RES1 for session key establishment with $\mathcal{S}$. Second, $\mathcal{S}$ sends query QUE2 individually to each $\mathcal{O}$ found in the 1st step to complete session key establishment and deliver her PROF (for Level 2 discovery) and "code" (for Level 3 discovery). A Level 2 object checks the PROF to get $\mathcal{S}$'s non-sensitive attributes; a Level 3 object verifies the "code" to see if $\mathcal{S}$ is its fellow. Then its PROF is encrypted using the session key and returned in response RES2. We present the design details of Level 2 and 3 in the following two sections (symbols listed in Tab. 1).

TABLE 1
Symbols Used in Argus

Symbol	Description
$CERT_X$	public key certificate of entity X
$PROF_X$	profile of X
$KEXM_X$	public value for ECDH key agreement, generated by X
$[...]SIG_X$	signature of content in [ ] signed by X
MAC_X	hash-based message authentication code generated by X
$[...]ENC_K$	ciphertext of content in [ ] encrypted by key K
$\|\|$	concatenation

5 Level 2: Differentiated Discovery

This discovery targets Level 2 objects which provide differentiated services according to subjects' *non-sensitive* attributes. Argus uses conventional cryptography (e.g., ECDSA, ECDH) for mutual authentication and service information secrecy. Also, we develop an alternative using more recent cryptography Attribute-based Encryption [8] for comparison. We find that Argus precedes the alternative in many aspects (e.g., 10x high scalability and computation efficiency), and suits our service discovery context best.

5.1 Argus's Level 2 Scheme

Main Idea. Subject $\mathcal{S}$ reveals her non-sensitive attributes to object $\mathcal{O}$ by presenting her profile $PROF_S$; $\mathcal{O}$ checks her attributes and chooses the suitable $PROF_O$ variant to return. The messages between $\mathcal{S}$ and $\mathcal{O}$ must be authenticated, and $PROF_O$ must be encrypted for service information secrecy.

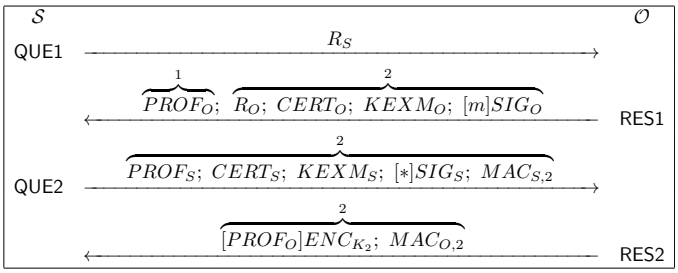


Fig. 3. v1.0, concurrent 2-in-1 algorithm for Level 1 and 2 discoveries. The content in a brace with number i is only sent by subjects performing Level i discovery, or by objects in Level i.

Algorithm. As shown in Fig. 3, $\mathcal{S}$ first broadcasts query QUE1 carrying random R_S. $\mathcal{O}$ in Level 2 sends back response RES1. Besides random R_O, RES1 carries $\mathcal{O}$'s public key certificate $CERT_O$ and key exchange material $KEXM_O$. In Fig. 3, $m = R_S||R_O||KEXM_O$ is signed by $\mathcal{O}$ for integrity protection.

Second, on receiving RES1, $\mathcal{S}$ verifies the signature, and based on $KEXM_O$ she establishes the pre-master secret $preK$. Level 2's session key $K_2 = HMAC(preK, label_K||R_S||R_O)$, $HMAC(secret, seed)$ is an HMAC-based pseudorandom function, $label_K$ is ASCII string "session key". Then $\mathcal{S}$ sends query QUE2 to $\mathcal{O}$, which contains her profile $PROF_S$, public key certificate $CERT_S$ and key exchange material $KEXM_S$. All the content sent and received so far (denoted as $*$, i.e., QUE1, QUE2 and $PROF_S$, $CERT_S$, $KEXM_S$) is signed for integrity protection. Also, Level 2's $MAC_{S,2} = HMAC(K_2, label_S||Hash(*))$, $label_S$ = "subject finished", is sent for $\mathcal{O}$ to verify the success of the handshake.

Third, on receiving QUE2, $\mathcal{O}$ verifies the signature, and based on $KEXM_S$ it establishes the same $preK$ and K_2. Based on K_2 it verifies $MAC_{S,2}$: if valid, $\mathcal{O}$ checks which subject category ($pred_i$) matches the subject's non-sensitive attributes in $PROF_S$, and encrypts the corresponding $PROF_O$ using K_2. $\mathcal{O}$ sends back response RES2 containing the $PROF_O$ ciphertext and $MAC_{O,2}$, where $MAC_{O,2} = HMAC(K_2, label_O||Hash(*))$, $label_O$ = "object finished". $\mathcal{S}$ verifies $MAC_{O,2}$ and decrypts the ciphertext using K_2, getting the service information for her securely.

Our design shares some similarities with TLS handshake [14] in realizing mutual authentication and symmetric key establishment. However, Argus efficiently embeds profile exchange in, accomplishing the whole discovery in a 4-way interaction; besides, it has minimum message overhead

by eliminating any unrelated component and fixing the key exchange algorithm at ephemeral ECDH, and authentication at ECDSA, which are significantly more efficient than other algorithms like RSA (experiments in Section 9.1).

5.2 Alternative: ABE based Strategy

Another way to realize Level 2 discovery is to use Ciphertext-Policy Attribute-based Encryption [8]. In CP-ABE, a subject is issued with a set of keys by authorities, each corresponding to one of her attributes; and a ciphertext has a built-in policy which is a predicate on subject attributes. A subject can decrypt a ciphertext if and only if she has all the attributes (thus the keys) to meet the ciphertext policy. ABE has been used for scenarios like cloud-based health record sharing [15], [16], [17], where patients upload their health record ciphertexts to the cloud and can then go offline; the ciphertexts have policies specifying consumer attributes (e.g., $[Physician \wedge HospitalX]$), and only the qualifying ones (i.e., physicians in Hospital X) can decrypt them after retrieving them from the cloud. More background introduction on ABE is in Appendix A.2.

Based on subjects' different attributes, the ciphertexts they can decrypt and the plaintexts they can view are also differentiated. This sounds a natural match for Level 2 discovery, thus we explore its application in our context.

Solution. The backend is the only key authority and data producer. It does not give objects plaintext $\{pred_i, PROF_{O,i}\}$ ($pred_i$ is a predicate on subject attributes) at bootstrapping, instead it issues them with ABE ciphertexts—$PROF_{O,i}$ encrypted using policy $pred_i$. Also, it issues subjects with ABE secret keys according to their attributes. An object performs no cryptographic operation; actually it is nothing more than a data store, and can safely hand over whatever is queried because the PROFs are encrypted. Only subjects with the secret key components corresponding to needed attributes can successfully decrypt the ciphertexts and view the service information. An object may make its ciphertexts cached by objects around to improve discovery availability and speed.

5.3 Comparison: Argus vs. ABE

Here we present the results of comprehensive comparison between Argus and ABE, which shows that Argus has many more advantages, e.g., 10x high scalability and computation efficiency. Quantitative analysis and experimental details are shown in Section 8 and Section 9.

1) Scalability. Updating overhead must be reduced to make the system scalable to enterprise environments, for the context's *huge subject/object amounts* and *frequent subject/object churns* properties (Section 2). ABE has up to **10x** as expensive updating overhead as Argus when a subject leaves the system (analysis in Section 8). Non-monotonic ABE is lighter weight but still more expensive than Argus (Appendix A.2).

2) Computation Cost. A subject/object in Argus establishes one pairwise session key for each object/subject it is interacting, thus the amount of compute intensive operations (ECDSA, ECDH) is linear to the number of interaction targets. When using ABE, a subject needs to decrypt a ciphertext for each target, while objects have no encryption work since the ciphertexts were given by the backend. ABE is at least **10x** as computationally expensive as Argus (experiments in Section 9.1), and a derivative called non-monotonic ABE [18], [19] is even more expensive.

3) Responsiveness. Argus uses 4-way messages while ABE 2-way, so Argus costs a little longer in transmission. However, its overall time cost in discovering services in proximity is much smaller than ABE due to its much smaller computation cost, achieving much higher responsiveness. Section 9.2 shows that it discovers 20 Level 2 objects in 0.63 s while ABE may cost seconds for just one object.

4) Online vs. Offline. ABE's offline property facilitates its application in scenarios like health record sharing, where data production and consumption are asynchronous, and producers can upload their data (ciphertexts), then go offline, and the data is still secured and available to consumers. Argus needs subjects and objects to be online at the same time. But service discovery is different from record sharing: objects should stay online for being discovered and accessed; actually, it is better that their service information becomes unobtainable when they go offline due to power outages or malfunctions, otherwise subjects will be misled into attempting to access unavailable services. Thus, ABE's offline property is not beneficial here.

5) Traceability. When using ABE, after a subject retrieves ciphertexts, whether she succeeds in decrypting them or not is only known by herself. An object has no way to know who has succeeded in discovering its services, thus logging is impossible. In contrast, Argus allows each object to record the result of each discovery operation.

6) Engineering Maturity. Argus uses conventional cryptography which is well optimized and has passed long-term validation. ABE has scarce or relatively preliminary implementations and might be less comfortable to engineers.

6 Level 3: Covert Discovery

This discovery targets Level 3 objects which provide covert services according to subjects' *sensitive* attributes. Like Level 2, object $\mathcal{O}$ in Level 3 ensures that subject $\mathcal{S}$ has qualifying attributes before returning $PROF_O$. However, those attributes are sensitive (e.g., learning disability) and $\mathcal{S}$ does not want to disclose them to $\mathcal{O}$ before seeing $\mathcal{O}$'s sensitive qualifying attributes (e.g., machine serving special populations). Note that using K_2 established by $\mathcal{S}$ and $\mathcal{O}$ in Section 5 to encrypt $\mathcal{S}$'s sensitive attributes does not work: it only protects $\mathcal{S}$'s privacy against third parties, but here $\mathcal{S}$ wants to guard against not only third parties, but also $\mathcal{O}$, the one she is interacting with.

Neither $\mathcal{S}$ nor $\mathcal{O}$ is willing to make the first move, thus a chicken-or-egg problem arises. We apply a symmetric-key mechanism similar to [20] for private service discovery so sensitive attribute secrecy on both sides (i.e. mutual privacy) is achieved. Besides, we beef up mutual privacy with a novel scheme which hides Level 3 discovery in Level 2, making them indistinguishable. As a result, attackers are even unaware that Level 3 discovery is happening, let alone peeking at sensitive attributes. We show that it outperforms alternatives using more recent cryptography (e.g., 10x high computation efficiency), and suits our context best.

6.1 Argus's Level 3 Scheme for Sensitive Attribute Secrecy

Main Idea. Recall that at bootstrapping subjects and objects are put into the same secret group if they have sensitive attributes which allow them to recognize each other. They are called "fellows" and share one symmetric group key (K_i^{grp} for group i). Confirming one's sensitive attributes is indirectly realized by verifying its group membership, by verifying its possession of the group key. $\mathcal{S}$ and $\mathcal{O}$ send to each other an HMAC generated from the group key to prove her/its possession. $\mathcal{O}$ verifies $\mathcal{S}$ first and then vice versa. $\mathcal{O}$ will send the suitable $PROF_O$ confidentially only if $\mathcal{S}$ is its fellow.

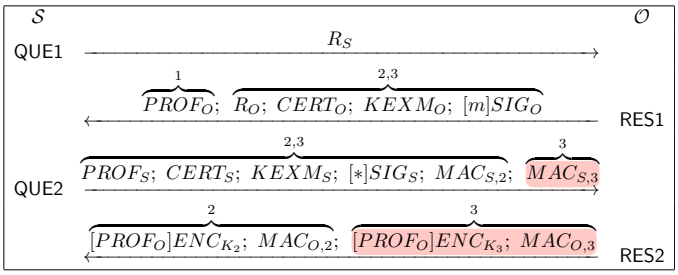


Fig. 4. v2.0, concurrent 3-in-1 algorithm for Level 1, 2, 3 discoveries. It supports sensitive attribute secrecy in Level 3. The content in red boxes are newly added on top of v1.0 in Fig. 3.

Algorithm. Fig. 4 shows Level 3 discovery which is built on top of Level 1 and 2. Level 3's session key $K_3 = HMAC(K_2||K_i^{grp}, label_K||R_S||R_O)$ (assume $\mathcal{S}$ is in group i), is computed after K_2. When $\mathcal{S}$ performs Level 3 discovery, QUE2 carries Level 3's $MAC_{S,3} = HMAC(K_3, label_S||Hash(*))$. The definitions of $label_K$, $label_S$, $label_O$ and $*$ can be found in Section 5.

$\mathcal{O}$ in Level 3 can establish the same K_3 only if it has K_i^{grp}, i.e., it is a fellow of $\mathcal{S}$. On receiving QUE2, it does all a Level 2 object does, and additionally verifies $MAC_{S,3}$: if valid, it recognizes $\mathcal{S}$ as a fellow. Then it generates $MAC_{O,3} = HMAC(K_3, label_O||Hash(*))$, encrypts the suitable $PROF_O$ variant with K_3, and adds them to RES2.

Depending on $\mathcal{O}$'s level, the HMAC of RES2 may be $MAC_{O,2}$ or $MAC_{O,3}$. $\mathcal{S}$ first tries to verify it with K_2 to see if it is a $MAC_{O,2}$: if valid, $\mathcal{S}$ knows $\mathcal{O}$ is in Level 2 and she decrypts the ciphertext using K_2. Otherwise she uses K_3 to verify if the HMAC is a $MAC_{O,3}$: if valid, $\mathcal{S}$ knows $\mathcal{O}$ is in Level 3 and she uses K_3 for decryption.

Sensitive Attribute Secrecy. Sensitive attribute secrecy is realized on both $\mathcal{S}$ and $\mathcal{O}$ due to HMAC's one-way feature: if $\mathcal{S}$ and $\mathcal{O}$ are not in the same group, $\mathcal{O}$ will find $MAC_{S,3}$ invalid. However, all it knows is $\mathcal{S}$ is not its fellow, but which secret group $\mathcal{S}$ is in stays unrevealed because the group key is not explicitly exchanged. Vice versa, $\mathcal{S}$ will not know a non-fellow $\mathcal{O}$'s group membership.

Overhead of Extensions. We see v2.0 brings in little extra discovery overhead to v1.0. Most components in RES1 and QUE2 are reused. In QUE2, one HMAC (only 32 bytes if using SHA-256) is added when performing Level 3 discovery. The length of RES2 is unchanged, because either Level 2 or Level 3 service information is sent back, not both. As for computation cost, $\mathcal{S}$ and $\mathcal{O}$ need one more HMAC generation and verification, together costing less than 1 ms.

6.2 Argus's Level 3 Scheme for Indistinguishability

So far, Level 3 discovery prevents a non-fellow from peeking at an entity's sensitive attributes (aka privacy), but performing Level 3 discovery itself implies that the entity has at least one sensitive attribute. E.g., attackers find a subject is seeking for Level 3 objects, then they guess she is a member of special crowds, though which crowd (e.g., depression or addiction) is unknown. Level 2 and 3 can be easily distinguished due to their message composition differences: i) QUE2 has one more component ($MAC_{S,3}$) when seeking for a Level 3 object; ii) RES2 from a Level 3 object carries $MAC_{O,3}$ other than $MAC_{O,2}$. Attacks in detail are presented in Section 7.

Main Idea. We make Level 2 and Level 3 indistinguishable, realizing covert visibility. QUE2s from all subjects have identical structures regardless of levels, so are RES2s from all objects. Attackers are even unaware that Level 3 discovery is happening.

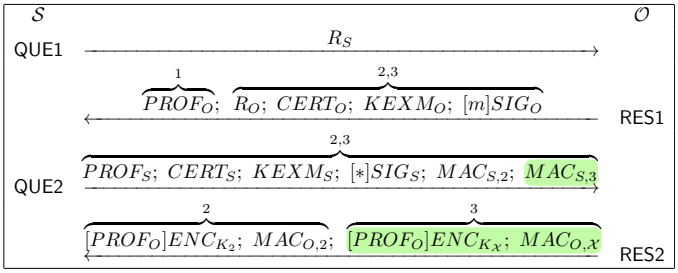


Fig. 5. v3.0, concurrent 3-in-1 algorithm for Level 1, 2, 3 discoveries. It supports both sensitive attribute secrecy and indistinguishability in Level 3. The content in green boxes are modified on top of v2.0 in Fig. 4.

Algorithm. Fig. 5 shows that Level 2 and 3 discoveries now use identical QUE2 which always carries $MAC_{O,2}$ and $MAC_{O,3}$, and are performed concurrently. Besides, a Level 3 object no longer sends $MAC_{O,3}$ constantly in RES2. Instead, its RES2 has $MAC_{O,\mathcal{X}}$, where $\mathcal{X}$ can be 2 or 3: $MAC_{O,3}$ to fellows, with $PROF_O$ encrypted by K_3; $MAC_{O,2}$ to non-fellows, with $PROF_O$ encrypted by K_2.

Indistinguishable Subjects. ***1) concurrent discoveries.*** A subject uses the same QUE2 to discover both Level 2 and 3 objects, so attackers cannot tell Level 3 discovery is happening based on QUE2's composition difference. ***2) cover-up key.*** All subjects perform concurrent Level 2 and 3 discoveries, even if some of them have no sensitive attributes. Recall that at bootstrapping even a subject $\mathcal{S}$ with no sensitive attribute still gets a (fake) secret group key (called *cover-up key*) from the backend. A cover-up key is a unique random number and there is no second entity owning it, thus the $MAC_{S,3}$ generated from it will not result in successful handshakes. But using it, $\mathcal{S}$ can send $MAC_{S,3}$ like a sensitive attribute owner; now in attackers' eyes, every subject belongs to special populations with sensitive attributes, and the real ones are concealed.

Indistinguishable Objects. ***1) double-faced role.*** Each Level 3 object plays a "double-faced" role: it returns $MAC_{O,3}$ and offers Level 3 special service information to its fellows (special populations), while returns $MAC_{O,2}$ and offers Level 2 services to non-fellows. Non-fellows always receive $MAC_{O,2}$ and they do not know the object's another role (i.e., Level 3). ***2) constant RES2 length.*** The ciphertexts

of service information ($PROF_O$) in Level 2 and 3 may have different sizes. To eliminate that, $\mathcal{O}$ appends minimum meaningless bytes to each of its $PROF_O$ variants before transmission to make them identically long. ***3) constant response time.*** In Level 3 discovery $\mathcal{O}$ verifies one more HMAC ($MAC_{S,3}$) than Level 2, costing longer. The time difference is mostly negligibly small (< 0.1 ms); if not, $\mathcal{O}$ will wait for that difference before sending Level 2 RES2 to make response time identical to Level 3.

Overhead of Extensions. We see v3.0 brings in little extra overhead to v2.0. In QUE2, now $MAC_{S,3}$ is mandatory, so QUE2 should always have these 32 bytes (if using SHA-256). RES2's length and computation cost are unchanged.

So far the entire protocol in all three levels is presented and we summarize its workflow in Fig. 6.

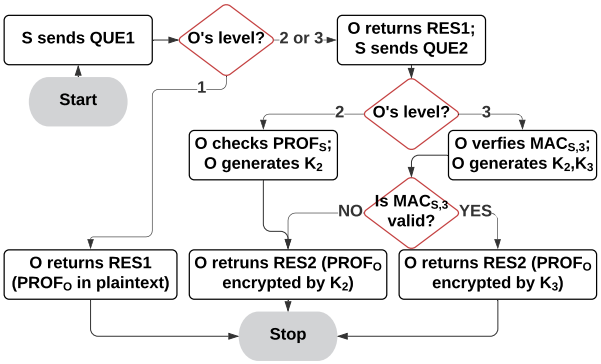


Fig. 6. Argus Workflow

6.3 Multiple Sensitive Attributes

Level 3 discovers the fellow objects in one secret group at a time, because $MAC_{S,3}$ is generated from one group key. In reality a subject may have multiple (usually no more than a few) sensitive attributes and are members in multiple secret groups. Her device can automatically use her group keys in turns (one at a time) to generate $MAC_{S,3}$ and launch discoveries, till all her authorized covert services are found.

6.4 Alternative: PBC based Strategy

An alternative for sensitive attribute secrecy in Level 3 is to use Pairing-based Cryptography [9]. When using PBC, fellows are not issued with a group symmetric key K_i^{grp} at bootstrapping; instead they each get a unique PBC key, which will be used to establish a *pairwise* symmetric key (denoted as $K_{S,O}^{pair}$ between $\mathcal{S}$ and $\mathcal{O}$) during discovery. PBC has this one extra step, and the remaining procedures are the same as Argus, i.e., $\mathcal{S}$ and $\mathcal{O}$ test each other's possession of the symmetric key to see if the other is a fellow. More backend introduction on PBC is in Appendix A.3.

Solution. PBC can use Argus's four messages as is, but with different computation operations: when receiving RES1, $\mathcal{S}$ computes $K_{S,O}^{pair}$, and replaces K_3 with $K_{S,O}^{pair}$ in generating $MAC_{S,3}$; similarly, when receiving QUE2, $\mathcal{O}$ computes $K_{S,O}^{pair}$, and uses $K_{S,O}^{pair}$ to generate $MAC_{O,3}$.

Comparison: Argus vs. PBC. We find that Argus has smaller computation cost (PBC is **10x** as expensive) while achieving the same secrecy and scalability.

1) Computation Cost. The symmetric-key mechanism in Argus Level 3 brings in no compute intensive operations (HMAC generation and verification cost < 0.1 ms), and the overall computation cost is almost identical to that in Level 2. PBC, however, has **10x** as expensive cost as Level 2 (experiments in Section 9.1).

2) Secrecy. PBC uses a pairwise key $K_{S,O}^{pair}$ to generate $MAC_{S,3}$ and $MAC_{O,3}$, so the conversation between $\mathcal{S}$ and $\mathcal{O}$ is kept from any third party. If the two HMACs are purely generated from a group key K_i^{grp}, an eavesdropper which is a fellow (thus also has K_i^{grp}) can understand the conversation. However, note that Argus Level 3 is not a pure symmetric-key mechanism which uses K_i^{grp} directly; instead, it uses K_3 (generated using K_2 and K_i^{grp}). Leveraging K_2 from Level 2 which is a pairwise key between $\mathcal{S}$ and $\mathcal{O}$, K_3 is also pairwise. Thus, Argus performs as well as PBC in resisting eavesdropping.

3) Scalability. To remove a subject from a secret group, PBC just needs to notify the fellow objects to revoke her ID. A symmetric-key mechanism needs all the fellows (both subjects and objects) to update the group key, which is more expensive. But again, Argus Level 3 is not purely symmetric-key; it is built on top of Level 2. Thus, to prevent a subject from further Level 3 discovery, we can revoke it in Level 2 by adding her ID into the fellow objects' revocation lists, which has the same overhead as PBC.

7 Security Analysis

Threat Model. We assume the backend is trustworthy and well-protected. Also, communication between the backend and subject/object devices is secure. Subject devices and resource-rich objects are reasonably well protected, e.g., by their operating systems.

We assume breaking the cryptographic algorithms (e.g., ECDSA, ECDH) are computationally infeasible when long enough keys are used (e.g., 128-bit). Attackers can capture, inject, modify and replay messages sent over the communication channel. **Sources.** Attackers may be *external*—they are not registered on the backend thus have no backend-signed public keys, or *internal* ones that are registered but go rogue. **Roles.** Attackers may behave passively as eavesdroppers, or actively to impersonate subjects or objects and interact with benign nodes. **Targets.** Attackers may aim at service information secrecy, sensitive attribute secrecy, indistinguishability.

Like TLS and many other algorithms, the security of ours is on the premise that secret information is kept to its owner, and is computationally infeasible to compromise, and cannot be obtained from sources outside of the channel. However, in reality it can, e.g., attackers have military computing resources, or they leverage malware or social engineering to steal private keys from users/devices. Resisting those attacks is out of the scope. Our analysis below shows that attackers will fail unless they have session key K_2, K_3, or private key K_X^{pri} (X is a subject or object) and/or secret group key K_i^{grp} (for group i).

7.1 Level 2 Attacks from External Attackers

Service Information Secrecy. In Level 2, attackers may try to view a $PROF_O$ they are unauthorized to. Also, they may

spread a false $PROF_O$. Their possible roles and actions are:

Case1: Eavesdropper. Passive attacker $\mathcal{E}$ eavesdrops the conversation between subject $\mathcal{S}$ and object $\mathcal{O}$ to view $PROF_O$. She needs to compromise K_2 to decrypt the ciphertext of $PROF_O$, which is infeasible. Note that $\mathcal{E}$ cannot obtain K_2 by compromising K_S^{pri} or K_O^{pri} (cracking a long-term key might be easier than a session key), because we use ephemeral ECDH for key exchange between $\mathcal{S}$ and $\mathcal{O}$, which has *forward secrecy*.

Case2: Subject/Object Impostor. Active attacker $\mathcal{E}_S$ poses as $\mathcal{S}$ to interact with $\mathcal{O}$ and request $PROF_O$. Since the interaction is authenticated, she will fail due to the lack of K_S^{pri}. $\mathcal{E}_O$ poses as $\mathcal{O}$ to give $\mathcal{S}$ fake service information, but it will fail without K_O^{pri}. Besides, $PROF_O$ is signed by the admin for authenticity and integrity protection, breaking which is considered infeasible.

7.2 Level 3 Attacks from External Attackers

Service Information Secrecy. First, attackers may launch the same attacks in Level 2 to get $PROF_O$. We assume $\mathcal{S}$ and $\mathcal{O}$ are fellows in secret group i, and they share K_i^{grp}.

Case3: Eavesdropper. This time $\mathcal{E}$ needs K_3 to decrypt the ciphertext of $PROF_O$. She may: i) compromise K_3 directly, which is infeasible; ii) or compromise K_2 and K_i^{grp} because they together generate K_3, but this does not make things easier.

Case4: Subject/Object Impostor. $\mathcal{E}_S$ poses as $\mathcal{S}$ to interact with $\mathcal{O}$ for $PROF_O$. To succeed, she needs K_S^{pri} and K_i^{grp}. For the same reason in Case2, $\mathcal{E}_O$ will fail in giving fake service information.

Sensitive Attribute Secrecy. Second, attackers may try to find out what sensitive attributes $\mathcal{S}$ or $\mathcal{O}$ has.

Case5: Eavesdropper. Attacker $\mathcal{E}$ eavesdrops the conversation of $\mathcal{S}$ and $\mathcal{O}$, but she will know $\mathcal{S}$ or $\mathcal{O}$ is in group i only if she can confirm that $MAC_{S,3}$ or $MAC_{O,3}$ is a valid HMAC generated using K_i^{grp}, which needs K_2 and K_i^{grp}. Besides, knowing that an entity belongs to group i does not mean knowing its sensitive attribute, unless the mapping relationships between group IDs and attributes are also known. However, that knowledge is kept to the admin.

Case6: Subject/Object Impostor. $\mathcal{E}_S$ poses as $\mathcal{O}$'s subject fellow to interact with $\mathcal{O}$, and tries to find out $\mathcal{O}$'s sensitive attributes. To succeed, she needs a valid subject private key, K_i^{grp} and the mapping relationships. Similarly, $\mathcal{E}_O$ may pose as an object fellow to explore $\mathcal{S}$'s sensitive attributes, and it needs a valid object private key and K_i^{grp}.

Indistinguishability. Third, attackers may try to find out if $\mathcal{S}$ or $\mathcal{O}$ has any sensitive attribute (regardless of what sensitive attribute), and if Level 3 discovery is happening.

Case7: Eavesdropper. i) subject distinguishability. As introduced in Section 6.2, we use cover-up keys on subjects who have no sensitive attributes to make them pose as real sensitive attribute owners. $\mathcal{E}$ have no way to distinguish them. ii) object distinguishability. Since only Level 3 objects's RES2 may carry $MAC_{O,3}$ other than $MAC_{O,2}$, $\mathcal{E}$ may leverage this to recognize Level 3 objects. However, to recognize $MAC_{O,3}$, she needs K_3, which is impractical.

Case8: Subject/Object Impostor. $\mathcal{E}_S$ interacts with $\mathcal{O}$ to see if it is in Level 3. If $\mathcal{E}_S$ recognizes the HMAC in RES2 as $MAC_{O,3}$, then she knows $\mathcal{O}$ is. This needs a valid private key and K_i^{grp}. Alternatively, she may try an elimination method: tell if the HMAC is $MAC_{O,2}$, and if not, it is $MAC_{O,3}$ then. Verifying $MAC_{O,2}$ only needs a valid private key, so the security strength is degraded if this works. However, Level 3 objects play "double-faced" roles: they send $MAC_{O,2}$ to attackers, thus attackers cannot use the elimination trick.

$\mathcal{E}_O$ may impersonate $\mathcal{O}$ to interact with $\mathcal{S}$ to see if she has any sensitive attributes, but cover-up keys impede that.

Case9: Side-Channel Attacks. Beware of side-channel attacks which may compromise indistinguishability. Particularly, due to a Level 3 object's additional computation on base of Level 2, it will take longer to respond than a Level 2 one. An attacker may recognize Level 3 objects through timing measurements and analysis (i.e. *timing attacks*). However, in Argus a Level 3 object only needs to verify one more HMAC (i.e. $MAC_{S,3}$) than a Level 2 one, which costs < 0.1 ms on Raspberry Pi. It cannot be detected when buried under much larger time fluctuations from OS, network, etc.

7.3 Attacks from Internal Attackers

Assume benign entity $\mathcal{I}$ goes rogue. Unlike an external attacker, she already has a valid private key K_I^{pri}. However, note that if she eavesdrops (Case1, 3, 5, 7) or impersonates others (Case2, 4), her own private key is useless and the attacks are the same as external attacks. It becomes easier to crack Case6, 8: since $\mathcal{I}$ already has a valid private key, she just needs to compromise K_i^{grp}, which is still difficult.

7.4 Consequences of Key Compromise

If a session key is compromised, only that session's content (e.g., service information) will be exposed; if a private key is compromised, only that entity will be impersonated. If a private key and a group key are both compromised, attackers may find out members in that one secret group only, by interacting with them one by one instead of getting the entire member list. Each of these cases has a limited impact, and cannot paralyze the service discovery system.

8 Scalability Analysis

The system must be scalable to enterprises, and the most critical metric for scalability in our context is updating overhead instead of discovery overhead. This is because Argus is for discovering services *in proximity*, the number of which is usually not more than dozens (Section 2). In contrast, any change in the backend database (e.g., policy addition, subject removal) related to Level 2 or 3 should be immediately synchronized to *all* affected subjects/objects on the ground, otherwise authorized users may fail to discover and access new services timely, while unauthorized users continue to see services they are no longer eligible for. Such *updating overhead* (defined as the number of affected subjects and objects) can be huge.

Level 1 & 3 Scalability. Level 1 is little relevant to this authorization-related updating because it offers identical information to everyone. When changing a subject/object/policy in Level 3, the worst case (e.g., remove a person from a secret group) is all the other fellows in the group should get new keys, i.e., the overhead is $(\gamma - 1)$

(defined in Section 2, usually $10^0 \sim 10^1$). It is small due to a secret group's limited size.

Level 2 Scalability. Level 2 has the largest updating overhead. We find that Argus is up to **1000x** as efficient as ID-based ACL alternatives in adding a subject, and up to **10x** as efficient as Attribute-based Encryption (ABE) alternatives in removing a subject. To add/remove an object/policy, mostly just that object or the objects mentioned in that policy should be updated, thus the overhead is 1 or β (defined in Section 2, usually $10^0 \sim 10^2$). Adding/removing a subject is the bottleneck of scalability, thus we show the analysis on it in detail as below.

8.1 Detailed Scalability Analysis on Level 2

ID-based ACL. In this method, every object locally stores its access control list enumerating the identities of subjects which are allowed to access and discover it. In Section 2 we show that a subject may access N ($10^2 \sim 10^3$) objects typically. Then when subject $\mathcal{S}$ joins/leaves the system, the N objects she can access should be notified to add/remove her ID (i.e. ID_S) to/from their ACLs.

TABLE 2
Updating Overhead Comparison

	Add a subject	Rmv a subject
ID-based ACL	N	N
ABE	1	$(\xi_o N + \xi_s(\alpha - 1)) \approx 10N$
Argus	1	N

Argus. In Argus, an object stores an attribute-based ACL, which uses predicates on subject attributes (e.g., *[position=='manager' && department=='X']*) to describe its authorized subjects. To access objects, a newcomer $\mathcal{S}$ just needs to contact the backend once to get her attribute profile (overhead: 1), and present it to objects; objects do not need to update their ACLs. This significantly outperforms ID-based ACL (up to **1000x**). However, when $\mathcal{S}$ leaves, the backend should notify the N objects that she could access, to remove ID_S from their ACLs and refuse her future discovery.

ABE. Ciphertext-Policy Attribute-based Encryption [8] can be used for Level 2 discovery. At bootstrapping, the backend issues $\mathcal{S}$ with a set of keys, each corresponding to her one attribute (e.g., *department:X*); also, the backend issues $\mathcal{O}$ with ABE ciphertexts—$PROF_{O,i}$ encrypted using policy $pred_i$ (e.g., *[position=='manager' && department=='X']*). Based on ABE's principle, the $PROF_{O,i}$ ciphertext can be decrypted only if $\mathcal{S}$ has all the attributes (thus the keys) to meet $pred_i$.

About updating, a newcomer $\mathcal{S}$ just gets her secret keys from the backend, then she can discover services (overhead: 1). To revoke $\mathcal{S}$, the backend has to *globally* revoke a set of her attributes to make her no longer belong to any subject category. E.g., to revoke her attribute *department:X*, it: i) re-encrypts *all* ciphertexts whose encryption policies contain *department:X*, and delivers them to their objects (overhead: $\xi_o N, \xi_o \geq 1$); ii) re-generates those attributes' secret keys, and delivers them to *all* subjects owning the attributes except $\mathcal{S}$ (overhead: $\xi_s(\alpha - 1), \xi_s \geq 1$). α (defined in Section 2) is the number of subjects in a subject category, usually $10^0 \sim 10^2$, possibly $\geq 10^3$. Such attribute-level updating often affects more subjects than $\mathcal{S}$'s category members, and more objects than what $\mathcal{S}$ could access, that is why ξ_s, ξ_o mostly go over 1. The overall overhead is $(\xi_o N + \xi_s(\alpha - 1))$. When $\xi_o, \xi_s > 1$, or α is large (e.g., $10^3 \sim 10^4$, if $\mathcal{S}$ is in a large category like a department or college), the overhead easily goes to $10N$ or more.

9 Experimental Evaluation

We implement all the three levels of Argus. Besides, two alternatives are implemented: one is based on Attribute-based Encryption (ABE) and used for Level 2 (introduced in Section 5); the other is based on Pairing-based Cryptography (PBC) and used for Level 3 (Section 6).

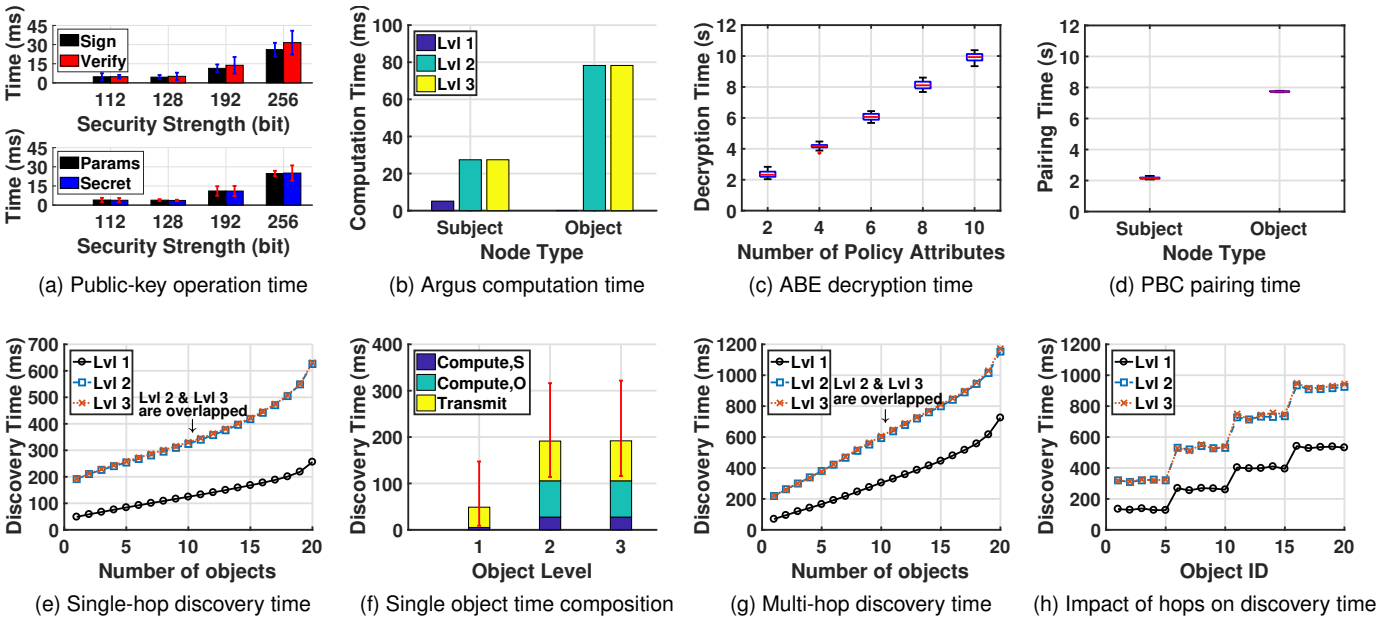
Testbed Rationality. We conduct experiments on a testbed consisting of 1 subject device (Nexus 6) and 20 objects, each emulated by a Raspberry Pi 3. 1) As mentioned in Section 2, we assume Level 2 and 3 objects, which have high security requirements, are equipped with sufficient resource and power, and can run public-key algorithms at reasonable speed. 2) Though Level 1 objects in reality have poor computing resource, they simply return profiles after receiving queries, and no computation is needed, thus emulating them using Pis or Arduinos makes no difference. 3) Our design is above the network layer and orthogonal to radios. As a result, we believe Pis communicating with WiFi well emulate objects in all three levels, and networking and power aspects. 4) One discovery operation aims to find the n (10^1, defined in Section 2) objects in a user's proximity though she has access rights to N ($10^2 \sim 10^3$, Section 2) objects. N affects updating overhead and is used in scalability analysis (Section 8); to test the time of discovering n objects we believe our tested with 20 Pis has a sufficient scale.

Settings. The subject device broadcasts QUE1s since at the beginning which objects are nearby is unknown; QUE2, RES1 and RES2 are unicast. To remedy packet losses, we use application-layer acknowledgement and retransmission for unicast messages: a packet sent for the ith time ($i \geq 1$) will wait for a random delay between $(2^{i-1} - 1)\tau$ and $(2^i - 1)\tau$ and depart unless the ACK for a previous transmission returns before this. We choose $\tau = 150$ ms because it leads to low latencies and low retransmission rates (Section 9.3), and use this setting in other experiments.

We test the computation time cost on the subject device and objects, and find that to achieve 128-bit security strength, Argus needs only 105 ms while ABE and PBC cost at least **10x** long. As for the overall discovery time cost (mainly by computation and transmission), Argus takes 0.25 s to discover 20 Level 1 objects and 0.63 s for 20 Level 2 or 3 objects (each object is 1 hop from the subject). For a multi-hop case where 20 Level 2 or Level 3 objects are from 1-hop to 4-hop away from the subject, Argus costs 1.15 s for discovery completion and is still well responsive.

9.1 Computation Time Cost

The cryptographic computation time for Argus, AES and PBC are evaluated. We use crypto libraries OpenAndroidSSL on the subject device and JCA on objects, for their efficient implementation compared with others (e.g.,



(a) Public-key operation time (b) Argus computation time (c) ABE decryption time (d) PBC pairing time

(e) Single-hop discovery time (f) Single object time composition (g) Multi-hop discovery time (h) Impact of hops on discovery time

Fig. 7. Computation Time Cost and Discovery Time Cost

Spongy Castle). ECDSA is preferred to RSA because the latter costs much longer (e.g., 18x for 128-bit strength). The experiments show that Argus is **10x** as computationally efficient as ABE and PBC.

Fig. 7 (a) shows the computation time on the subject side for ECDSA (for signature) and ECDH (for key exchange) operations, under strength 112-bit, 128-bit, 192-bit, 256-bit. As shown, computation time increases with security strength, e.g., 112-bit costs 4.7 ms in signing while 256-bit costs 26.0 ms. For each strength, ECDSA verification/ECDH secret computation costs similar or slightly longer time than signing/parameter generation. Similar results are observed on the object side. Other operations like HMAC and AES cost less than 1 ms on both sides. In the following experiments, we use 128-bit due to its fast speed while sufficient strength.

Argus. Fig. 7 (b) shows the overall computation time on subjects and objects in all levels. In Level 1 discovery, a subject only needs to verify one signature (of $PROF_O$), costing 5.1 ms; an object has no compute intensive operation. In Level 2 and 3, a subject needs 1 signing, 3 verifications (for $CERT_O$, $KEXM_O$, $PROF_O$), 2 ECDH operations (parameter generation, secret computation), costing 27.4 ms; an object needs the same, costing 78.2 ms. Note that the public-key operations in Level 2 and 3 are identical.

ABE. When using ABE, encryption is performed by the backend and decryption by subject devices. Objects do neither. Considering that the backend has superior performance and ciphertexts can be generated beforehand, here we focus more on subject decryption time cost, tested using CP-ABE library [21]. As shown in Fig. 7 (c), ABE's decryption time is well linearly to the number of attributes in the ciphertext policy. Each attribute leads to about 1 second decryption time increase.

PBC. Pairing time is the time cost for computing a pairwise symmetric key using PBC keys. We evaluate JPBC library [22] on the subject device and objects, and pairing costs 2.2 s and 7.7 s respectively, as shown in Fig. 7 (d).

Note that test results depend significantly on the crypto library implementation. The ABE and PBC libraries currently available are probably preliminary, thus the decryption time and pairing time presented should not be interpreted literally, but rather revealing the likely magnitudes. According to our experiments, ABE and PBC cost at least **10x** as long as Argus.

9.2 Overall Discovery Time Cost

We present in Fig. 7 (e) (g) the overall time cost (mainly by computation and transmission) for Argus to discover 20 objects, in all the three levels, in both single-hop and multi-hop conditions. Even in Level 2 and 3, discovering 20 single-hop objects costs only 0.63 s, while multi-hop objects only 1.15 s. Such short latencies result in positive user experience.

Single-Hop. Fig. 7 (e) shows that in each level the discovery time cost increases with the number of objects to be discovered. The completion of discovering 20 objects is quick, which costs 0.25 s for Level 1 and 0.63 s for Level 2 and 3. Level 1 needs less than half of the time of Level 2/3 because it is 2-way communication while the other two are 4-way. Also, Level 1 has less computation. Fig. 7 (f) shows the time composition for discovering 1 single-hop object: in Level 1 89% of the time is on transmission; in Level 2/3 it is 45%. The variance in (f) mainly comes from changeful wireless transmission time.

Notice that Level 2 and Level 3 have overlapped time curves, thus indistinguishable time cost. This is because Argus Level 3 only has one more HMAC generation than Level 2, which averagely costs 0.08 ms on Pi. In practice, such tiny difference will not give attackers chances to distinguish Level 3 objects from Level 2 ones via timing measurements, because it is buried in timing fluctuations (e.g., from OS, program run, network) of higher orders of magnitude.

Multi-Hop. We also test the discovery time cost in a multi-hop condition. The 20 objects are divided to 4 equal

groups: Object 1–5, 6–10, 11–15, 16–20 are 1, 2, 3, 4 hops away from the subject respectively. Fig. 7 (g) shows that the discovery costs longer than the single-hop case: here discovering 20 Level 1 objects needs 0.72 s, and Level 2 or Level 3 objects 1.15 s. But still, the latency is short. Fig. 7 (h) reveals the impact of hops on latency: discovering a 1-hop Level 1 object averagely costs 0.13 s, while 4-hop needs 0.53 s; as for a Level 2 or Level 3 object, 1-hop takes about 0.32 s (0.1 s computation + 0.22 s transmission) while 4-hop 0.92 s (0.1 s computation + 0.82 s transmission). We see transmission time increases roughly linearly with hop counts. In all cases Argus is fast enough to achieve satisfactory user experience.

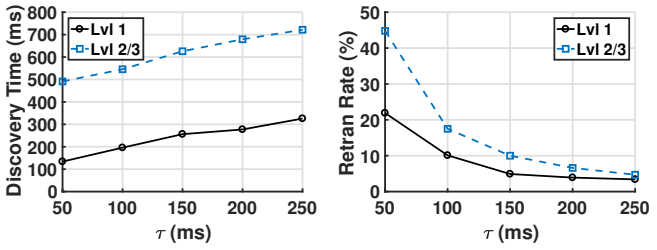
9.3 Message Overhead

We show the size (unit: byte) of every message in Tab. 3. R_X: 28 B (like TLS); MAC_X: 32 B (SHA-256). X is $\mathcal{S}$ or $\mathcal{O}$. When using 128-bit strength, $CERT_X$ is an X.509 ECDSA certificate of 552 B; $KEXM_X$ and SIG_X have 64 B. $PROF_X$ averagely has 200. $[PROF_O]ENK_K$ is assumed to use AES in CBC mode, with 16-byte IV, 32-byte MAC, (padding ignored), thus a ciphertext has size 248. Notice that Argus has light message overhead, and actually in Level 2 and 3 about half of the bytes are from the two X.509 certificates.

TABLE 3
Message Size

MessageSize (byte)	QUE1	RES1	QUE2	RES2	Total
Lvl 1	28	200			228
Lvl 2/3	28	772	1008	280	2088

QUE1 is a broadcast message while the others are unicast. Thus, to discover n objects which are single-hop away, if message losses are not considered: in Lvl 1, $(n+1)$ (1 QUE1, n RES1) messages are needed; in Level 2 or Level 3, $(3n+1)$ (1 QUE1, n RES1, QUE2, RES2) are needed.



(a) Impact on latencies (b) Impact on retransmission rates

Fig. 8. Retransmission Time Choosing

In real environments message losses happen, and we use application layer backoff and retransmissions for reliable delivery. Fig. 8 shows that a larger τ increases latencies while a smaller τ incurs unnecessary retransmissions. We find $\tau = 150$ ms achieves a good balance between low latency (0.6 s for discovering 20 Level 2 or Level 3 objects) and low retransmission rates (10%).

10 Related Work

10.1 Centralized vs. Distributed Discovery

Centralized. Many existing solutions depend on centralized infrastructures (i.e. servers, directories) which maintain the information of registered services and handle announcements to or queries from users. E.g., in Ninja SDS [3], a server stores the description of available services and services running at a specific location; MQTT [23] is a publish/subscribe protocol which realizes reliable, efficient messaging between IoT devices [24], [25]. Other centralized examples are DNS [2], Jini [26] (Sun Microsystems), Salutation [27], and SLP [4]. Such systems use servers as repositories for efficient, scalable and wide-area discovery. However, they may encounter a single point of failure or long latency. Some systems [3], [23], [28], [29] have multiple servers running concurrently to improve reliability, and the servers keep contacting each other to make the caches updated. *Eventual consistency* [30] (a type of weak consistence) is achieved, thus users may see obsolete information.

Distributed. Distributed solutions like DEAPspace [31] (IBM), UPnP [5], and SDP [6] (Bluetooth) are infrastructureless, and any service may announce itself or reply a query. Multicast DNS [32], SLP and Bonjour [7] (Apple) support both centralized and distributed discoveries. A distributed, ad-hoc discovery strategy has the advantage of discovering nearby services robustly (no single point of failure) and quickly. However, it does not have a wide-area discovery scope. E.g., UPnP provides discovery in LAN, DEAPspace and Bluetooth in a single-hop ad-hoc network, while centralized systems like DNS and Ninja have global discovery.

10.2 Secure Discovery

10.2.1 Authenticated & Encrypted Discovery.

Security issues are limitedly covered in existing work. **Authentication.** Some systems [2], [31] require neither user nor service information to be authenticated. SLP [4] authenticates services or devices but not users; Salutation [27] in contrast, authenticates users only. In Ninja [3], Jini [26] and UPnP [5], both services and uses are authenticated. **Encryption.** Ninja, UPnP and Bluetooth [6] have messages encrypted for confidentiality.

10.2.2 Private Discovery.

In an ad-hoc network, if a user and a service both have privacy concerns, neither wants to expose its sensitive information before the other does, a chicken-or-egg problem arises. In solutions like [3], [33], a trustworthy proxy is assumed and used as a bridge between users and services. Both entities simply send to the proxy encrypted messages which only the proxy can decrypt. This model avoids the chicken-or-egg problem but is infrastructure-dependent, while the solutions below are not.

Multiple cryptographies have been applied. **1) Public-key.** In Private Authentication [34], the sender encrypts its ID with the receiver's public key (e.g., RSA) to ensure only that receiver can view it. One-to-many public-key cryptographies like Prefix Encryption [35] (based on Identity-Based Encryption [36], [37]) have also been used [38]. Public-key strategies have huge computational cost. **2) Symmetric-key.** In [39], [40], [20] a user and the service she can

discover get a symmetric key at bootstrapping. In later discoveries, they do not reveal their real IDs; instead, they test each other's possession of the symmetric key, mostly by exchanging MACs generated from the key. Zhu et al. believe even disclosing a MAC might expose too much information of the sender, thus a Bloom filter [41] generated from the MAC is sent instead [42], or one piece of a MAC is sent each time [43]. **3) Pairing-based.** When Pairing-based Cryptography [9], [44] is used, fellows get PBC keys at bootstrapping which will be used to generate a pairwise symmetric key during handshake. Then two users test if the other owns the symmetric key, like symmetric-key mechanisms. MASHaBLE [45] combines this with BLE to discover secret community members. Liu et al. [46] apply it to gateways' authenticating WiFi devices.

How is Argus different? 1) Argus consciously chooses a distributed, P2P discovery strategy because IoT interactions are largely physical proximity based. 2) It achieves concurrent visibility scoping at three levels, while existing solutions are restricted to one and unfit for enterprise IoT with numerous, heterogeneous services. 3) It minimizes the updating overhead upon the frequent enterprise user churns, making the system scalable to enterprise IoT. 4) It goes beyond mutual privacy in existing work and achieves indistinguishability.

11 Discussion

Secrecy of Physically Visible Services. Argus prevents user devices from collecting service information of unauthorized objects/services, especially those behind walls, which usually have larger amounts and higher secrecy requirements than those in public areas. Humans may gain knowledge of physically visible services, but that belongs to physical access security, which is out of the scope of this paper.

Performance-Poor Objects. In this paper we assume Level 2 and Level 3 objects have resources for public-key computation at reasonable speed, because vendors naturally spend a little extra money securing important objects with better hardware, considering that nowadays hardware is already very cheap. However, it is valuable to explore a lightweight strategy fitting for even performance-constrained objects, and we leave this to future work.

Revocation. When a subject loses access rights, Argus efficiently updates the policies to the objects she could access to reject her future discoveries. Of course, if she has already gained an object's information, the revocation cannot remove the knowledge from her head. However, for proximity-based discovery in a large-scale enterprise environment, it is not uncommon that a subject has not discovered all the objects she could discover when her access rights are gone, then the revocation stops her knowing more.

KP-ABE. Besides CP-ABE which is explored, there is a Key-Policy variant [47] (KP-ABE). When using KP, a ciphertext of service information is tagged with a set of attributes (e.g., $\{Room1, Printer\}$ describing the encrypted data while a user key has a built-in policy which is a predicate on data attributes. Such tags are in plaintext to serve as search keywords and an input for decryption. We have shown that CP is not as good as Argus, and KP is even worse: like CP, it also has expensive computation cost and updating overhead; besides, its plaintext data attribute tags of ciphertexts ruin service information secrecy.

Unlinkability. Unlinkable discovery [45] is a type of private discovery which prevents attackers from identifying or tracking users. Its work usually has a city-scale context, where the visiting to certain places (e.g., hospitals, bars, clubs) involves privacy and should be kept secret. Argus does not achieve unlinkability, because we believe within an enterprise, location history (which department/building a person has been to) is a less sensitive thing. Even so, we may leave it as future work.

Message Transmission. In our implementation we use broadcast for QUE1 and unicast for QUE2, RES1 and RES2. Also, we use application-layer acknowledgement and retransmission to remedy unicast message losses. In essence, this problem is on communication reliability, which is orthogonal to and independent from Argus.

Suitable Users. Argus discovers three levels of services concurrently on an enterprise scale, and is especially suitable for enterprise IoT which has numerous, heterogeneous services. We believe IoT device manufacturers targeting enterprise customers would like to replace existing protocols in their products with Argus. In contrast, small scale IoT contexts like smart homes mostly have small amounts, only one or two levels of services, and Argus's advantages in scalability and heterogeneity are less needed, thus smart home manufacturers would not bother to update protocols.

12 Conclusion

In this paper, we describe the design, implementation and evaluation of Argus, a proximity-based service discovery algorithm which efficiently discovers three levels of visibility (public, differentiated and covert) in parallel, and is scalable to IoT in enterprises. It has much less computation cost and updating overhead than alternatives using ABE and PBC. It is very responsive, taking only 0.25 s to discover 20 nearby public services, 0.63 s for 20 differentiated or covert services, agile for satisfactory user experience.

Acknowledgment

This work is supported in part by the National Science Foundation under grant number 1652276.

References

[1] Q. Zhou, M. Elbadry, F. Ye, and Y. Yang, "Towards fine-grained access control in enterprise-scale internet-of-things," *IEEE Transactions on Mobile Computing*, 2020.

[2] P. V. Mockapetris, "Rfc1035: Domain names-implementation and specification," 1987.

[3] S. E. Czerwinski, B. Y. Zhao, T. D. Hodes, A. D. Joseph, and R. H. Katz, "An architecture for a secure service discovery service," in *MobiCom*, vol. 99, 1999, pp. 24–35.

[4] E. Guttman, C. Perkins, J. Veizades, and M. Day, "Service location protocol, version 2," 1999.

[5] U. Forum, "UPnP Device Architecture 2.0," http://upnp.org/specs/arch/UPnP-arch-DeviceArchitecture-v2.0.pdf.

[6] B. SIG, "Bluetooth Service Discovery Protocol," https://www.bluetooth.com/specifications/assigned-numbers/service-discovery.

[7] A. Inc., "Bonjour," https://developer.apple.com/bonjour/.

[8] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *2007 IEEE symposium on security and privacy (SP'07)*. IEEE, 2007, pp. 321–334.

[9] N. Koblitz and A. Menezes, "Pairing-based cryptography at high security levels," in *IMA International Conference on Cryptography and Coding*. Springer, 2005, pp. 13–36.

[10] Q. Zhou and F. Ye, "Graphiterouting: Name-based hierarchical routing for internet-of-things in enterprise environments," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–7.

[11] Q. Zhou, M. Elbadry, F. Ye, and Y. Yang, "Heracles: Scalable, fine-grained access control for internet-of-things in enterprise environments," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1772–1780.

[12] J. Nielsen, *Usability engineering*. Morgan Kaufmann, 1994.

[13] X. Song, Y. Huang, Q. Zhou, F. Ye, Y. Yang, and X. Li, "Content centric peer data sharing in pervasive edge computing environments," in *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*. IEEE, 2017, pp. 287–297.

[14] T. Dierks and E. Rescorla, "The transport layer security (tls) protocol version 1.2," 2008.

[15] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in *Infocom, 2010 proceedings IEEE*. Ieee, 2010, pp. 1–9.

[16] M. Li, S. Yu, K. Ren, and W. Lou, "Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings," in *International conference on security and privacy in communication systems*. Springer, 2010, pp. 89–106.

[17] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," *IEEE transactions on parallel and distributed systems*, vol. 24, no. 1, pp. 131–143, 2013.

[18] A. Lewko, A. Sahai, and B. Waters, "Revocation systems with very small private keys," in *2010 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2010, pp. 273–285.

[19] S. Yamada, N. Attrapadung, G. Hanaoka, and N. Kunihiro, "A framework and compact constructions for non-monotonic attribute-based encryption," in *International Workshop on Public Key Cryptography*. Springer, 2014, pp. 275–292.

[20] J. Lindqvist, T. Aura, G. Danezis, T. Koponen, A. Myllyniemi, J. Mäki, and M. Roe, "Privacy-preserving 802.11 access-point discovery," in *Proceedings of the second ACM conference on Wireless network security*. ACM, 2009, pp. 123–130.

[21] J. Wang, "Java realization for ciphertext-policy attribute-based encryption," *Computer Science College of Shandong University*, 2012.

[22] A. De Caro and V. Iovino, "jpbc: Java pairing based cryptography," in *2011 IEEE symposium on computers and communications (ISCC)*. IEEE, 2011, pp. 850–855.

[23] OASIS, "MQTT," http://mqtt.org.

[24] G. C. Hillar, *MQTT Essentials-A lightweight IoT protocol*. Packt Publishing Ltd, 2017.

[25] R. Venanzi, B. Kantarci, L. Foschini, and P. Bellavista, "Mqtt-driven sustainable node discovery for internet of things-fog environments," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[26] K. Arnold, R. Scheifler, J. Waldo, B. O'Sullivan, and A. Wollrath, *Jini specification*. Addison-Wesley Longman Publishing Co., Inc., 1999.

[27] S. Consortium *et al.*, "Salutation architecture specification version 2.0 c," *salutation specification, Salutation Consortium*, 1999.

[28] U. C. Kozat and L. Tassiulas, "Service discovery in mobile ad hoc networks: an overall perspective on architectural choices and network layer support issues," *Ad Hoc Networks*, vol. 2, no. 1, pp. 23–44, 2004.

[29] F. Sailhan and V. Issarny, "Scalable service discovery for manet," in *International Conference on Pervasive Computing and Communications: PerCom 2005*, 2005, pp. 235–244.

[30] W. Vogels, "Eventually consistent," *Communications of the ACM*, vol. 52, no. 1, pp. 40–44, 2009.

[31] M. Nidd, "Service discovery in deapspace," *IEEE personal communications*, vol. 8, no. 4, pp. 39–45, 2001.

[32] S. Cheshire and M. Krochmal, "Multicast dns," RFC 6762, February, Tech. Rep., 2013.

[33] F. Zhu, M. Mutka, and L. Ni, "Splendor: A secure, private, and location-aware service discovery protocol supporting mobile services," in *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003)*. IEEE, 2003, pp. 235–242.

[34] M. Abadi and C. Fournet, "Private authentication," *Theoretical Computer Science*, vol. 322, no. 3, pp. 427–476, 2004.

[35] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Workshop on the theory and application of cryptographic techniques*. Springer, 1984, pp. 47–53.

[36] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Annual international cryptology conference*. Springer, 2001, pp. 213–229.

[37] A. Lewko and B. Waters, "Why proving hibe systems secure is difficult," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2014, pp. 58–76.

[38] D. J. Wu, A. Taly, A. Shankar, and D. Boneh, "Privacy, discovery, and authentication for the internet of things," in *European Symposium on Research in Computer Security*. Springer, 2016, pp. 301–319.

[39] J. Pang, B. Greenstein, S. Seshan, and D. Wetherall, "Tryst: The case for confidential service discovery." in *HotNets*, vol. 2, no. 3.2, 2007, p. 1.

[40] B. Greenstein, D. McCoy, J. Pang, T. Kohno, S. Seshan, and D. Wetherall, "Improving wireless privacy with an identifier-free link layer protocol," in *Proceedings of the 6th international conference on Mobile systems, applications, and services*. ACM, 2008, pp. 40–53.

[41] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," *Communications of the ACM*, vol. 13, no. 7, pp. 422–426, 1970.

[42] F. Zhu, M. W. Mutka, and L. M. Ni, "A private, secure, and user-centric information exposure model for service discovery protocols," *IEEE Transactions on Mobile Computing*, vol. 5, no. 4, pp. 418–429, 2006.

[43] F. Zhu, W. Zhu, M. W. Mutka, and L. M. Ni, "Private and secure service discovery via progressive and probabilistic exposure," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 11, pp. 1565–1577, 2007.

[44] S. D. Galbraith, K. G. Paterson, and N. P. Smart, "Pairings for cryptographers," *Discrete Applied Mathematics*, vol. 156, no. 16, pp. 3113–3121, 2008.

[45] Y. Michalevsky, S. Nath, and J. Liu, "Mashable: mobile applications of secret handshakes over bluetooth le," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 2016, pp. 387–400.

[46] W. Liu, Z. Yan, and Y. He, "The wi-fi device authentication method based on information hiding," in *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)*. IEEE, 2018, pp. 80–86.

[47] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM conference on Computer and communications security*. Acm, 2006, pp. 89–98.

[48] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2005, pp. 457–473.

[49] J. Hur and D. K. Noh, "Attribute-based access control with efficient revocation in data outsourcing systems," *IEEE Transactions on Parallel & Distributed Systems*, no. 7, pp. 1214–1221, 2010.

[50] D. Balfanz, G. Durfee, N. Shankar, D. Smetters, J. Staddon, and H.-C. Wong, "Secret handshakes from pairing-based key agreements," in *2003 Symposium on Security and Privacy, 2003.* IEEE, 2003, pp. 180–196.

APPENDIX A
CRYPTOGRAPHIC PRELIMINARIES

A.1 Bilinear Maps

Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two multiplicative cyclic groups of prime order p. Let g be a generator of $\mathbb{G}_1$ and $\hat{e}$ be a bilinear map, $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$. The bilinear map has the following properties: i) bilinearity: for all $u, v \in \mathbb{G}_1$ and $a, b \in \mathbb{Z}_p$, $\hat{e}(u^a, v^b) = \hat{e}(u, v)^{ab}$; ii) non-degenracy: $\hat{e}(g, g) \neq 1$. We say that $\mathbb{G}_1$ is a bilinear group if the group operation in $\mathbb{G}_1$ and $\hat{e}$ are both efficiently computable. Notice that $\hat{e}$ is symmetric since $\hat{e}(g^a, g^b) = \hat{e}(g, g)^{ab} = \hat{e}(g^b, g^a)$.

A.2 Attribute-based Encryption

Attribute-based Encryption [48] is a public-key one-to-many encryption, and it has mainly two schemes: Key-Policy [47] (KP-ABE) and Ciphertext-Policy [8] (CP-ABE). In KP-ABE, a ciphertext is tagged with a set of attributes describing the data while a user key has a built-in policy which is a predicate on data attributes; CP-ABE, in contrast, issues a user with a set of keys according to her attributes, and a ciphertext is associated with a policy which is a predicate on user attributes. For either scheme, a ciphertext can be decrypted if and only if the attributes in the set match the policy. Note that the original KP- and CP-ABE are monotonic: their policies support logic AND ($\wedge$), OR ($\vee$), but not NOT. A policy example is $[(Student \wedge DeptX) \vee Admin]$. The four main algorithms of CP-ABE in [8] are:

1) Setup. This algorithm chooses a bilinear group $\mathbb{G}_1$ of prime order p, generator g, and two random exponents $\alpha, \beta \in \mathbb{Z}_p$. It outputs master key $MK = (\beta, g^\alpha)$ (kept secret) and public key $PK = (\mathbb{G}_1, g, g^\beta, \hat{e}(g,g)^\alpha)$ (publicly available).

2) KeyGen. Input: a set of user attributes $\mathcal{A}$ and MK. The algorithm outputs secret key $SK = (g^{(\alpha+r)/\beta}, \{g^r \cdot H(i)^{r_i}\}, i \in \mathcal{A})$, where $r, r_i \in \mathbb{Z}_p$ are randomly chosen and H is a hash function mapping any attribute described as a binary string to an element in $\mathbb{G}_1$. SK is securely issued to the user whose attributes constituting $\mathcal{A}$. As is seen, a user's secret key has one component for each of her attributes, but with a common r embedded in. No two users share the same r, which is for stopping collusion attacks.

3) Encrypt. Input: a plaintext M, an access policy tree $\mathcal{T}$ and PK. A leaf node in $\mathcal{T}$ specifies a user attribute and a non-leaf one represents logic AND or OR. In this way it carries a predicate on user attributes. The algorithm chooses a polynomial q_i for each node i in $\mathcal{T}$ in a top-down manner starting from the root node R. $q_R(0)$ is set to s where $s \in \mathbb{Z}p$ is randomly chosen. For any other node i, $q_i(0)$ is set to $q_{parent(i)}(index(i))$ where $parent(i)$ denotes i's parent in $\mathcal{T}$ and $index(i)$ is i's unique index given by its parent. The algorithm outputs $C = (\mathcal{T}, \tilde{C} = M \cdot \hat{e}(g,g)^{\alpha s}, g^{\beta s}, \{g^{q_i(0)}, H(i)^{q_i(0)}\}, i \in \mathcal{T}$'s leaf nodes).

4) Decrypt. Input: C and SK. The algorithm first computes $DecryptNode(CT, SK, i) = \hat{e}(g,g)^{r \cdot q_i(0)}$ for $i \in \mathcal{T}$'s leaf nodes. Then it aggregates these pairing results in a bottom-up manner using polynomial interpolation. Finally it may recover $\hat{e}(g,g)^{rs}$ and then $\hat{e}(g,g)^{\alpha s}$. $M = \tilde{C} / \hat{e}(g,g)^{\alpha s}$.

Revocation. i) Periodic revocation [8]. The original CP-ABE revokes users through a time attribute in $\mathcal{A}$ and $\mathcal{T}$. Ciphertexts are periodically re-encrypted with new time attributes (e.g., at the end of each day); users may periodically request updated keys (for the time attribute part) from the key manager and only unrevoked users will succeed. E.g., a user with SK for $\{Student, 07/30/18\}$ can decrypt the data encrypted with policy $[Student \wedge 07/30/18]$; if she is revoked, next day she will not be able to decrypt the data which is re-encrypted with policy $[Student \wedge 07/31/18]$ due to the lack of SK for $07/31/18$. ii) Immediate revocation [15], [17], [49]. Ciphertexts can also be re-encrypted using new keys immediately when a user leaves the system, and new keys are immediately, securely delivered to unrevoked users. iii) Non-monotonic revocation [18], [19]. Non-monotonic ABE additionally supports logic NOT in policies, which is useful for efficient revocation: to revoke $\mathcal{S}$, ciphertexts are re-encrypted using a policy which negates $\mathcal{S}$'s identity or unique attribute (e.g., $[Student \wedge (\textbf{NOT ID}_\mathbf{S})]$). Unrevoked users do not need to update their keys, thus updating overhead is reduced. However, as more and more subjects are revoked, a ciphertext's **NOT** list grows longer and decryption time also increases, till at some point (e.g., a month later) it is necessary to clear the **NOT** list, using periodic or immediate revocation.

A.3 Pairing-based Secret Handshake

Secret handshake [50], [45] ensures that two members in the same group will recognize each other as fellows while a non-fellow cannot identify one's membership, no matter by eavesdropping or performing a handshake. A pairing-based strategy is shown as below.

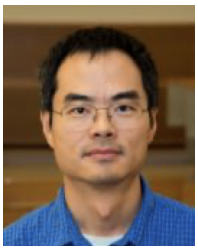
1) Setup. The admin chooses a bilinear group $\mathbb{G}_1$ of prime order p and generator g, and a group secret $s_i \in \mathbb{Z}_p$ is randomly chosen for each group i. Then, the admin issues entity X with a pseudonym ψ_X and the corresponding PBC key $K^{pbc}_{X,i} = H(\psi_X)^{s_i}$ if X is a member in group i, where H is a hash function mapping a string to an element in $\mathbb{G}_1$.

2) Handshake. Assume Alice from group i and Bob from group j attempt to shake hands secretly. First, they exchange pseudonyms, i.e. ψ_A and ψ_B. Second, Alice uses $K^{pbc}_{A,i}$ and ψ_B to compute a pairwise key $\hat{e}(H(\psi_A)^{s_i}), H(\psi_B)) = \hat{e}(H(\psi_A), H(\psi_B))^{s_i}$; similarly, Bob uses $K^{pbc}_{B,j}$ and ψ_A to compute a pairwise key $\hat{e}(H(\psi_A), H(\psi_B)^{s_j}) = \hat{e}(H(\psi_A), H(\psi_B))^{s_j}$. Apparently, the outcomes are the same if and only if $s_i = s_j$, i.e., Alice and Bob are fellows. Third, they check if the other possesses the same secret: if so, they recognize each other as fellows; otherwise non-fellows, but exactly which group she or the other is in (i.e., the membership) remains unrevealed.

Note that in the PBC alternative we develop for Level 3 discovery, the ID of X is used in place of a pseudonym, i.e., $K^{pbc}_{X,i} = H(ID_X)^{s_i}$. By exchanging profiles (containing IDs), $\mathcal{S}$ and $\mathcal{O}$ can start secret handshake. A pseudonym instead of a real ID should be used when anonymity or unlinkability is needed, which is not the case in this paper.

Qian Zhou received the BE degree from Beihang University (previously known as Beijing University of Aeronautics and Astronautics), and the PhD degree from the Department of Electrical and Computer Engineering at Stony Brook University. His research mainly focuses on enterprise-scale Internet of Things (or cyber-physical systems), particularly in security & privacy and networking aspects.

Omkant Pandey is an assistant professor at the Computer Science Department of Stony Brook University. His research interests are in cryptography and its interplay with other areas of computer science. His work on Attribute Based Encryption received the 2016 ACM CCS Test-of-Time Award. Dr. Pandey graduated from University of California Los Angeles with a Ph.D. in Computer Science.



Fan Ye received the BE and MS degrees from Tsinghua University, and the PhD degree from the Computer Science Department, UCLA. He is an associate professor with the ECE Department, Stony Brook University. He has published more than 90 peer reviewed papers that have received more than 12,000 citations according to Google Scholar. His research interests include mobile sensing systems, with applications in location based services and healthcare, Internet-of-Things, and wireless and sensor networks.