Improving Unlinkability of Attribute-based Authentication through Game Theory

YEVHEN ZOLOTAVKIN and JONGKIL JAY JEONG, Deakin University, Geelong, Australia
VERONIKA KUCHTA, The University of Queensland, St Lucia, Australia
MAKSYM SLAVNENKO and ROBIN DOSS, Deakin University, Geelong, Australia

This article first formalizes the problem of unlinkable attribute-based authentication in the system where each user possesses multiple assertions and uses them interchangeably. Currently, there are no recommendations for optimal usage of assertions in such authentication systems. To mitigate this issue, we use conditional entropy to measure the uncertainty for a Relying Party who attempts to link observed assertions with user labels. Conditional entropy is the function of usage statistics for all assertions in the system. Personal *decisions* made by the users about the usage of assertions contribute to these statistics. This collective effect from all the users impacts the unlinkability of authentication and must be studied using game theory. We specify several instances of the game where context information that is provided to the users differs. Through game theory and based on conditional entropy, we demonstrate how each user optimizes usage for the personal set of assertions. In the experiment, we substantiate the advantage of the proposed rational decision-making approaches: Unlinkability that we obtain under Nash equilibrium is higher than in the system where users authenticate using their assertions at random. We finally propose an algorithm that calculates equilibrium and assists users with the selection of assertions. This manifests that described techniques can be executed in realistic settings. This does not require modification of existing authentication protocols and can be implemented in platform-independent identity agents. As a use case, we describe how our technique can be used in Digital Credential Wallets: We suggest that unlinkability of authentication can be improved for Verifiable Credentials.

CCS Concepts: • **Security and privacy → Authentication;**

Additional Key Words and Phrases: Attribute-based authentication, unlinkability, game theory, digital credential wallets, verifiable credentials

Yevhen Zolotavkin and Jongkil Jay Jeong contributed equally to this research.
Jongkil Jay Jeong, Maksym Slavnenko and Robin Doss also with Centre for Cyber Security Research & Innovation (CSRI)..
Jongkil Jay Jeong and Robin Doss also with Cyber Security Cooperative Research Centre (CSCRC), Australia.

This work was supported by the Cyber Security Research Centre Limited whose activities are partially funded by the Australian Government's Cooperative Research Centres Program.

Authors' addresses: Y. Zolotavkin, J. J. Jeong, M. Slavnenko, and R. Doss, Deakin University, Geelong, Australia; emails: author@countermail.com, {jay.jeong, Maksym.Slavnenko, robin.doss}@deakin.edu.au; V. Kuchta, The University of Queensland, St Lucia, Australia; email: v.kuchta@uq.edu.au.

2471-2566/2022/02-ART12 $15.00
https://doi.org/10.1145/3501260

ACM Reference format:
Yevhen Zolotavkin, Jongkil Jay Jeong, Veronika Kuchta, Maksym Slavnenko, and Robin Doss. 2022. Improving Unlinkability of Attribute-based Authentication through Game Theory. *ACM Trans. Priv. Secur.* 25, 2, Article 12 (February 2022), 36 pages.
https://doi.org/10.1145/3501260

1 INTRODUCTION

The ability to prove one's identity has long been an integral part of ensuring service delivery, civic participation and inclusion across private and public sectors. This is supported by a plethora of new technologies, that have emerged over the past decade, enhancing prior authentication methods.

Attribute-based authentication (ABA) is one such promising development. It provides users control over their credentials, which can then be used to access a wide variety of digital services [18, 63, 70]. Providing users with the option to select credentials is the key privacy merit of ABA. The selection of credentials can be optimized to represent the attributes that a digital service provider needs to verify and identify users [17, 51, 63]. ABA also contributes to the trustworthiness of identification and authentication—it is a means to establish a form of trust between two unfamiliar parties that share trust in a common third-party entity [50]. Assertions obtained from the personal attributes of unfamiliar parties is a medium in establishing trust between them. The means to verify these attributes and/or assertions are commonly supplied by a third-party entity.

Our inquiry into the privacy of ABA is well justified due to the following reasons. First, governmental initiatives such as **Electronic identification, authentication and trust services (eIDAS)** have recently emerged and matured across the European Union [4, 30]. This facilitates the expansion of the set of attributes available to the users [23, 29]. Second, NFC-enabled eID cards containing personal attributes with high **Level of Assurance (LoA)** are now available in many of the member states. These verifiable attributes can be securely extracted using compatible models of smartphones, which improves the usability of ABA [19, 20]. Third, a number of institutions develop solutions that support principles of Self Sovereign Identity [27, 37, 45]. This allows individuals to control the amount of the information that is revealed in ABA, which shifts the focus toward privacy-optimal usage [25, 69, 72].

Multiple initiatives have originated in the professional and academic communities intending to address privacy issues associated with the usage of attribute-rich credentials. For instance, data models such as **Verifiable Credentials (VC)** and *I Reveal my Attributes* enable principles of data minimization [1, 72]. However, any new advancements in technologies also come with a new set of challenges and issues. For example, failure to analyze the attribute selection process may result in a number of threats. One of them is *linking* of authentication events initiated by the users. This suggests that the privacy-enhancing benefits of VC may be negated if the issues pertaining to the *linkage* of user credentials are not identified.

A number of different guides exist: They provide descriptions for numerous security and privacy objectives in authentication systems. To reduce the number of inconsistencies, we follow objectives and requirements that can be found in NIST Privacy Framework, NIST SP 800-53r5, ISO/IEC 27551 [10, 32, 50]. Figure 1 explains our inquiry through a schematic diagram that defines security and privacy objectives and requirements: starting with **High Level (H/L)** concepts on the left and progressing toward **Low Level (L/L)** details on the right.

Objectives of *confidentiality* and *disassociability* are common for many organizations. This is often considered in the H/L Context of an online service provided to the users that are external to the organization [10]. H/L Threats to confidentiality and disassociability are unauthenticated usage and non-consensual profiling, respectively. These threats need to be mitigated under H/L

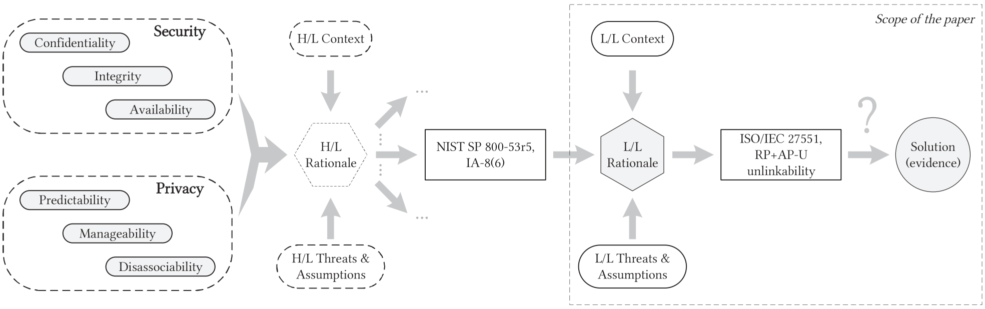


Fig. 1. Rationale and the scope of unlinkability assurance in this article.

Assumption that the organization allows users to remain pseudonymous, e.g., does not require users to identify themselves explicitly. According to H/L Rationale from NIST SP 800-53r5, this can be mitigated using control IA-8(6), which requires user authentication as well as *disassociation* of user attributes or identifier assertion relationships. Such guidance is, however, limited due to insufficient (e.g., high level) specification for disassociation. Even though a plethora of disassociability research exist [14], the problem definitions used there are not sufficient to cover all the interactions and threats that occur between various stakeholders within ABA.

To address this, we point out that ABA requires an additional layer of details that need to be considered by L/L Rationale in Figure 1. In this study, we adopt specific definitions of *unlinkability* from ISO/IEC 27551. The *scope of our study* is the process of finding evidence for the L/L requirement of unlinkability, which is based on L/L Context and L/L Threats & Assumptions for ABA. Our evidence is grounded on game-theoretical results where decisions of n users affect the measure of unlinkability in the ABA system. We demonstrate that substantial improvements of that measure are possible, which allows us to address a critical gap within the literature.

Main Contributions: we elaborate on the following main points:

- propose criterion of unlinkability C that is based on conditional entropy: It captures specifics of interchangeable usage of assertions possessed by the users in ABA;
- formalize non-cooperative coordination game where n users aim at maximizing their utilities derived from C;
- find equilibria in the system and compare resulting unlinkability with the situations where users act in a random (non-optimal) way;
- suggest how our results can be applied in practice to improve unlinkability through interchangeable usage of distinguishable assertions.

The remainder of this article is structured as follows. In Section 2, we define the requirement of **Attribute Provider (AP)**, **Relying Party (RP)**, and **user (U)** (RP+AP-U) unlinkability based on a typical L/L Context for the ABA system as well as corresponding L/L Assumptions and Threats. Based on this requirement we specify **Research Question (RQ)**. The methodology necessary to answer RQ is defined in Section 3, where we justify the need for a game-theoretical approach. The latter is named **Game-theoretical Approach to Privacy in Authentication Systems (GAPAS)**. It is formalized and analyzed in Section 4, where we consider two variants of the non-cooperative coordination games with incomplete information: *naïve* game and *tenable* game. This is followed by an experimental evaluation and comparison of unlinkability in these games as well as alternative scenarios of random interchangeable usage of assertions in Section 5. With the aim to emphasize the practical importance of GAPAS, we present and analyze in Section 6 the use case involving VC, which is becoming a widely used format. This section also highlights how decision-making

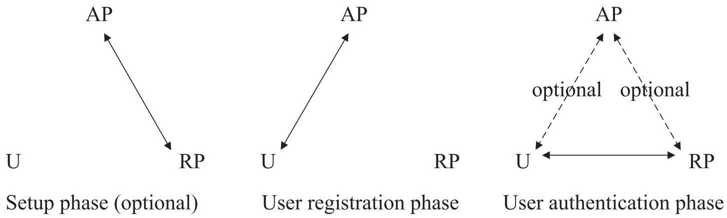


Fig. 2. Attribute-based authentication phases in the (AP, RP, U) model [50].

algorithms stemming from GAPAS can be implemented in **Digital Credential Wallets (DCW)**. In Section 7, we provide an overview of relevant literature and stipulate on existing gaps. An extended discussion about the theoretical and practical contributions of our article is provided in Section 8.

2 ESTABLISHING THE RESEARCH QUESTION

To establish the RQ, in this section, we analyze the factors that affect the *linkage* of user attributes in the ABA system. For this, we will refer to L/L factors as per Figure 1. The degree of ***generalizability*** of our work can be inferred from the details of mainly this section.

2.1 Low Level Context

For a more granular description of the ABA system, we define the (AP, RP, U) model: It consists of the AP, RP, and U (see Figure 2). Further in the text, we will use terms *attribute*, *credential*, and *assertion.*

Definition 1 (Attribute [39]). A reference of a named quality or characteristic inherent in or ascribed to someone or something.

Definition 2 (Credential [38]). An object or data structure that authoritatively binds an identity—via an identifier or identifiers—and (optionally) additional attributes, to at least one authenticator possessed and controlled by a subscriber.

Definition 3 (Assertion [39]). A statement to a relying party that contains identity attributes about a subject. Assertions may also contain authentication or other identity information about the subject.

Terminology will be used in the following context throughout this study. User registers with AP. Upon a request from the user, AP provides *credential(s).* These credential(s) usually contain *attribute(s).* A user then derives an *assertion* from this credential(s), and submits this assertion to RP to get authenticated.

2.2 Low Level Assumptions & Threats

Here we specify L/L assumptions and threats that will be used to justify L/L requirements to unlinkability in ABA.

2.2.1 Assumptions.

Assumption 1. *ABA system is a closed system with $n \geq 2$ distinct users: existing users do not leave, and new users do not join.*

The aforementioned assumption is justifiable, because there are moments in time when it holds. Representations (models) of the ABA system can be analyzed separately if at different moments n differs.

Assumption 2. *User assertions can be probabilistically linked to the users. These probabilities are known to AP.*

This assumption is justifiable, because AP identifies users and produces credentials for them. A common counterargument is that blinding techniques can be used, such as **Zero-Knowledge Proofs (ZKP)** [14, 31]. Their purpose is to make assertions indistinguishable, which creates uncertainty for AP (e.g., uniform prior). Such counterargument is defeated in practice due to the fact that *metadata* is present in the prevailing majority of assertions [39]. Specifics of AP procedures, as well as access policy requirements (e.g., veracity) set by RP, may cause such metadata to differ across assertions produced by different users [43]. This, however, does not exclude the possibility that two different users can often have indistinguishable assertions.

Assumption 3. *The frequency (or likelihood) of authentication can differ among users. These statistics are known to AP.*

The aforementioned assumption is justifiable: Publicly available sociological surveys and census data allows AP to predict how often users consume service provided by RP. This is because AP identifies users, which may provide information about their traits. For example, AP may know the gender of the users. This may be further reinforced with survey data, which tells that women are less likely to use online sports betting—service provided by RP—than men.

Assumption 4. *Each user can use multiple (distinguishable) assertions to satisfy access policy P set by RP.*

This is justifiable for flexible **Attribute-Based Access Control (ABAC)** policies P. For example, RP may require from a user a proof (assertion) of age. ABAC veracity requirement may dictate that the proofs are obtained from a credential that has substantial **Identity Assurance Level (IAL)** for identity proofing. The results may differ even if assertions are produced using selective disclosure (according to the data minimization principle) and ZKP. This is because of *metadata* that is unique for credentials issued by different issuers. RP, however, should accept any of these proofs if they are obtained from credentials that have substantially high IAL.

Assumption 5. *Neither the exact set of personal assertions of user i nor the number of different assertions that she uses is known to other users, $-i$.*

We justify Assumption 5 by referring to common settings in ABA systems where users do not normally share information about the number of their personal assertions as well as their properties. This information sharing is forfended, because RP should generally avoid revealing **Personally Identifiable Information (PII)** about the users affiliated with the service. As a result, users do not have contact details for each other. A possible counterargument here is that users may use public platforms (e.g., blogs) to make information about their assertions open to the general public. This, however, bears substantial privacy risks and will be avoided by rational users.

Assumption 6. *Authentication protocol satisfies requirements of correctness and unforgeability.*

As a matter of justification, we rely on existing research in cryptography [14]. Finally, this article considers linkability issues that arise only in the application communication layer (a similar assumption is made in ISO 27551). A possible counterargument to the latter is that malicious RP may extract network information such as Round Trip Time or the IP address of a user from the details of authentication/authorization protocol. These parameters can be used to further link users [42]. However, this line of thought can be defeated: Solutions exist that can hide network metadata [2].

2.2.2 Threats. A variety of privacy threats are described in ISO/IEC 27551. In this article, we study how to mitigate the ***threat of linking*** posed on the target entity U by malicious and colluding actors AP, RP [50]. As a result of such collusion, information available to AP (see Assumptions 2 and 3) is also available to RP. According to the standard, the objective of the threat is "...to track U across authentications while these are being controlled (read-write) and both RP and AP are subverted." This threat may arise, for example, in identity federations run by large governmental or private organizations who provide online services to the users [9, 66].

2.3 Requirement of Unlinkability

To mitigate the described threats (in the ABA context and under described assumptions), it is necessary to provide evidence that the requirement of RP+AP-U unlinkability is satisfied in ABA. Among all the unlinkability definitions in ISO/IEC 27551, "RP+AP-U unlinkability" is characterized by Unlinkability Level 5, which corresponds to the *highest degree of anonymity* [50].

Definition 4 (RP+AP-U Unlinkability). Is the unlinkability in the system where adversary can observe, actively intercept and modify exchanged messages and additionally plays the role of both RP and AP.

To better understand and satisfy that requirement, we start with a general definition of *unlinkability of operations.*

2.3.1 Unlinkability through Impossibility Proof.

Definition 5 (Unlinkability of Operations [49]). Requires that users and/or subjects are unable to determine whether the same user caused certain specific operations in the system or whether operations are related in some other manner.

One of the advantages of Definition 5 is that it does not put additional constraints on the set of "... certain specific operations." However, the limitation is due to the lack of guidance establishing whether "...subjects are unable to determine..." the users that caused authentication events in ABA. Based on this latter observation, assurance in unlinkability should be provided in the form of impossibility proof [12, 67]. Challenges associated with impossibility claims in security assurance arguments are outlined in ISO/IEC 15443-1: there is an explicit recommendation to avoid such claims [46]. This is often explained by the need to deploy an extensive set of assumptions that are used to prove impossibility claims. In most existing studies, these assumptions are not universal and cannot be easily applied in the ABA context (see Reference [7]). Unfortunately, the importance of unlinkability for privacy does not allow omitting the challenges of impossibility proofs. Hence, we analyze other definitions of unlinkability that can provide further clarifications about the approaches to supply evidence.

2.3.2 Unlinkability through Indistinguishability. Inability to distinguish the likelihoods for possible relations is a special case of *inability to determine* relations among items of interest. In this special case, the unlinkability requirement is satisfied if an observer is unable to give statistical preference to any of the distinct relations. To make such proof practical, it is often demanded the number of considered relations be finite (and quite limited). The following definition considers only two different types of relations.

Definition 6 (Unlinkability [24]). Within a particular set of information, the inability of an observer or attacker to distinguish whether two items of interest are related or not (with a high enough degree of probability to be useful to the observer or attacker) is called unlinkability.

As can be seen from Definition 6 indistinguishability plays an important role in the definition of unlinkability. To clarify the matter, we provide the following definition.

```
Output: true or false
Test:
    Adversary $\mathfrak{A}$ chooses the set of attributes for $U_0$, $U_1$ and policy P ;
    AP and RP execute the setup phase (if any) ;
    AP and $U_0$ execute the user registration phase (if any) ;
    AP and $U_1$ execute the user registration phase (if any) ;
    RP, AP and $U_0$ execute the authentication phase ;
    RP, AP and $U_b$ execute the authentication phase, $b \in \{0, 1\}$, $\Pr\big(b = 0\big) = 0.5$ ;
    $\mathfrak{A}$ returns a guess $b' \in \{0, 1\}$ on the value of b ;
    if $\Pr\big(b' = b\big) \rightarrow 0.5$ return true ;
    else return false .
```

Listing 1. RP+AP-U unlinkability, ISO/IEC DIS 27551.

Definition 7 (Indistinguishability [12]). Or qualitative identity is the property of being exactly alike or the relation that holds between such entities.

It is worth stressing the difference between indistinguishability of (a) assertions submitted to RP and (b) statistical preferences that an attacker has about relations between these assertions and each of n users. *Qualitative indistinguishability* of (a) is neither necessary nor sufficient for *quantitative indistinguishability* of (b). This is because Assumptions 2 and 3 enable substantial granularity of statistical representations. It, nevertheless, must be noted that qualitative indistinguishability can improve quantitative indistinguishability.

The prerequisite for quantitative indistinguishability of the likelihoods of possible relations among assertions is specified in the test for ABA (see Listing 1). This test defines whether Definition 4 is satisfied. As can be observed from the test, the number of possible relations is 2, because only U_0 and U_1 take part in the test. It is therefore unclear how such a test can be applied for a system with $n > 2$ users.

2.4 Justifying the Research Question

We, therefore, point out that further research on unlinkability in ABA is needed. First, we justify the usage of integral criterion C for unlinkability, which incorporates statistical characteristics for all n users. The main motivation behind this decision is to obtain a consistent measure: As opposed to the test, the measure should not be affected by the way pairs of users are selected. Second, we introduce a game-theoretical model where utilities of n users are derived from criterion C. Game-theoretical approach is justifiable, because in addition to the factors considered in Assumptions 2 and 3 statistical characteristics (that affect C) are also influenced by the ***decisions*** of the users. This latter statement is based on a flexible user-centric approach for authentication in ABA, which is reflected in Assumption 4.

Based on the factors presented above, the primary RQ of this study is as follows:

How should users use their assertions to maximize RP+AP-U unlinkability?

The RQ contributes to the evidence on Figure 1 and represents a user-centric approach to the problem of linkability in ABA. By following this line of thought, we disregard measures that can be applied to AP and RP to improve RP+AP-U unlinkability.

3 METHODOLOGY

In this section, we introduce our research methodology: It is based on *information theory* and *game theory*. First, information theory is used to quantify the degree of uncertainty (indistinguishability of preferences) that an attacker (colluding RP+AP) has about relations between observed assertions and each of n users. We demonstrate *why unlinkability* criterion C should be based on conditional entropy [16]. Second, through different examples, we demonstrate *how* user decisions impact unlinkability in ABA: Users face a difficult dilemma under uncertainty, but their decisions should not be random. To resolve this decision dilemma, we finally articulate why it is important to extend our methodology with non-cooperative coordination games with incomplete information [35]. In the literature to date, unlinkability in ABA has not been addressed using such methodology: This constitutes the academic novelty of our article.

3.1 Conditional Entropy and Unlinkability

Below, we provide an example of a situation in ABA where an attacker is establishing relations between assertions and users. We then address the question of best linking performance for such a situation. This allows us to define the criterion of unlinkability C.

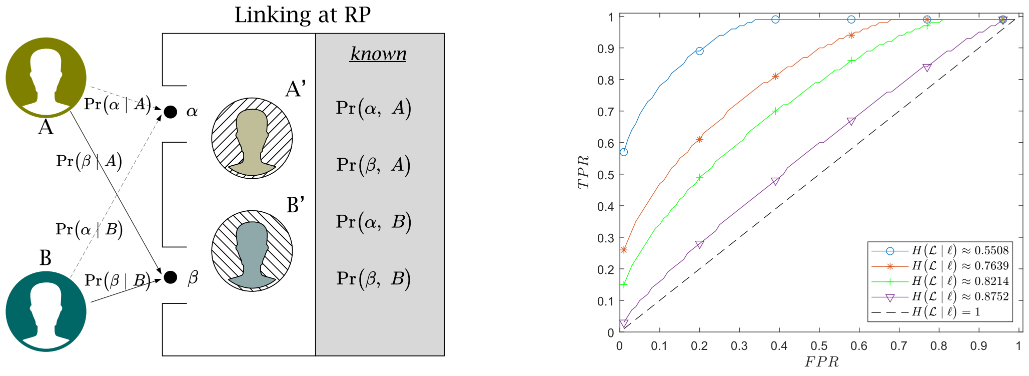
Example #1. Let us consider a simplified description of attribute-based authentication system consisting of 2 users *Alice* (A), *Bob* (B), and RP. We assume that RP establishes policy P, which allows us to perform certain actions over an object (e.g., *"to buy alcohol online"*) given that certain condition is satisfied (e.g., *"the subject is older than 18"*).

Yet another restriction of P is that RP accepts assertions of two types only: **Attribute Value Metadata (AVM)** must be either "digital driver license" (denoted as α) or "digital graduate certificate" (denoted as β). It is presumed that A and B have both types of the credentials (see Assumption 4). Both driver license and graduate certificate may contain multiple *subject attributes* within section *claim* of the credential.

These attributes may include `FirstName` and `LastName` of the subject, his/her `DOB`, `Address`, and so on. In addition, degree certificate may contain information such as `Degree`, `AwardedDate`, `Institution`, and so on. Despite the differences in their *subject attributes*, both A and B aim to reduce *qualitative distinguishability* between their assertions. For this, they apply privacy preserving techniques (such as selective disclosure and ZKP) to the *claim* section of the corresponding credential. This allows them to demonstrate that they satisfy P without revealing any additional information [14]. As a result, assertion with AVM α submitted by A cannot be distinguished by RP from assertion with AVM α submitted by B. However, RP can distinguish assertions with α from assertions with β.

Both A and B then decide on how often they use α versus β in their authentication sessions to RP. The relative frequency of usage of α by A is denoted as $\Pr(\alpha \mid A)$ and $\Pr(\beta \mid A) = 1 - \Pr(\alpha \mid A)$. Similarly, the relative frequency of usage of α by B is denoted as $\Pr(\alpha \mid B)$ and $\Pr(\beta \mid B) = 1 - \Pr(\alpha \mid B)$. The goal of A and B is to produce decisions that maximize unlinkability at RP. In contrast, RP tries to link authentication sessions. He observes realizations of random variable $l \in \ell$, $\ell = \{\alpha, \beta\}$ but does not know whether A or B has initiated authentication. From the standpoint of RP, this observed realization of l is the result of the decision that is made by the user whose label is random variable $\mathsf{L} \in \mathcal{L}$, $\mathcal{L} = \{A, B\}$ (see Figure 3(a)). For instance, it is intuitively clear that unlinkability is high if $\Pr(\alpha \mid A) = \Pr(\alpha \mid B) = 1$, e.g., both A and B always authenticate to RP with their digital driver license. This is because RP cannot distinguish their AVM. The opposite happens if $\Pr(\alpha \mid A) = \Pr(\beta \mid B) = 1$, e.g., A always uses driver license, but B always uses degree certificate when they authenticate to RP. In that case, RP can distinguish different entities behind authentication events as soon as he observes α in one event and β in the other.

As a result of "linking," for every observed assertion α or β malicious RP estimates the label L' of a user. The questions *"How does RP build his linking detector?"* and *"What are the characteristics*



(a) Scheme of linking based on attribute selection.

(b) ROC curves for optimal linking function built by RP.

Fig. 3. Linking in attribute-based authentication system with 2 users.

of it?" are out of the scope of our article. Instead, we demonstrate that the best performance of such linking at RP worsens when conditional entropy $H(\mathsf{L} \mid l)$ increases.

Lemma 1. *Best linking performance decreases with $H(\mathsf{L} \mid l)$ (for details see Appendix A).*

In addition to Lemma 1, the plots of **Receiver Operating Characteristics (ROC)** curves illustrate effect of conditional entropy on Figure 3(b): For a given **False Positive Rate (FPR),** the upper bound of **True Positive Rate (TPR)** decreases with $H(\mathsf{L} \mid l)$. Due to this property, the measure of conditional entropy can be used by A and B to produce decisions that improve unlinkability.

The proposed measure $C = H(\mathsf{L} \mid l)$ is additionally justified through the comparison with the test on Listing 1. We consider that entities U_0 and U_1 in the test are played by *Alice* and *Bob*, respectively, from the coordination game on Figure 3(a). Among the major similarities between the coordination game and the test are (1) realizations α, β in the game cohere with the set of attributes that satisfy access policy P in the test, (2) authentication events (protocol executions) repeat over time, and (3) performance is measured based on the conditional probability of correct guess (made by RP) given the value of realization/attribute. In spite of these similarities, the test in ISO/IEC DIS 27551 does not allow to evaluate the performance for the game. This is because the test (a) demands that $\Pr(A) = \Pr(B) = 0.5$, which is not always satisfied in practice and (b) details of "*guessing*" procedure (line 7, Listing 1) remain unclear. In addition, unlinkability conformance procedures remain unaddressed by the standard: It is unclear whether unlinkability can be improved if A and B can select among different assertions and *how* this selection should be done. Next, we will analyze if $C = H(\mathsf{L} \mid l)$ can be used to provide optimal unlinkability recommendations to A and B.

3.2 Optimal Selection of Assertions

As per Lemma 1, to improve unlinkability at RP users A and B may coordinate with one another to increase $H(\mathsf{L} \mid l)$. From Example #1, the best response for A and B is $\Pr(\alpha \mid B) = \Pr(\alpha \mid A)$. This also implies that $\Pr\big(\beta \mid B\big) = \Pr(\beta \mid A)$. In contrast, if $\Pr(\alpha \mid A) = 1$ and $\Pr\big(\beta \mid B\big) = 1$, then RP can link users with 100% accuracy. Coordination in the context of the described best response example is simple. This is because (a) the number of users is just 2, (b) they both use the same set of attributes, and (c) there is an implicit assumption that information for coordination is known to the users. Further, we will consider more realistic scenarios where unlinkability is measured using

Table 1. Credential Ownership by Category and AVM Realization

Category	**AVM realization**		**Total**
Drivers License	`Restricted` 771, 855(12.2%)	`Unrestricted` 5, 580, 224(87.8%)	6, 319, 611
Tertiary Qualification	`Undergraduate` 4, 030, 835(86.5%)	`Postgraduate` 631, 121(13.5%)	4, 661, 956

criterion $C = H(\mathsf{L} \mid l)$ for the system with $n \gg 2$ users whose assertions may differ and who use a specific coordinating agreement to improve unlinkability.

3.3 Unlinkability and the Lack of Communication

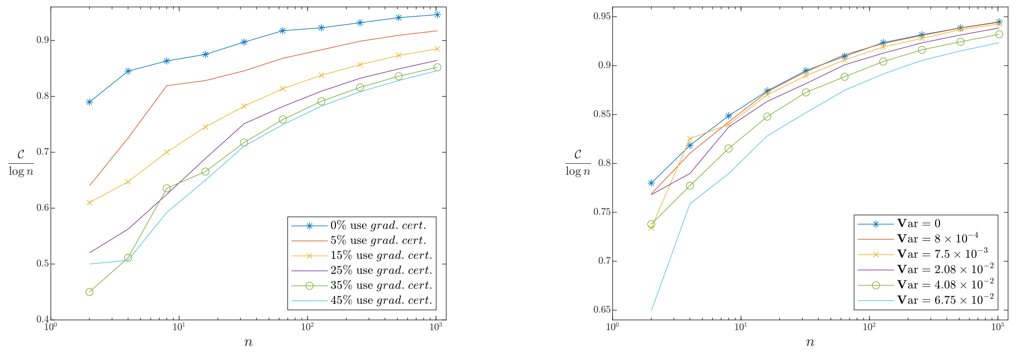
It is intuitive to suggest that in the system with $n \gg 2$ users they may coordinate by committing to some sort of de facto agreement. This does not require direct communication between the users. However, such an agreement may be sufficient to govern relative frequencies of utilization of the assertions possessed by the users. In the following example, we shall see how adherence to such an agreement (and possible deviations from it) affect unlinkability criterion $C = H(\mathsf{L} \mid l)$.

Example #2. We amend setting of Example #1 to make them more realistic. One observation is that for randomly selected 2 of n users probability that both of their assertions match must be less than 1. This, for instance, is likely to happen due to substantial granularity of AVM, which reflects variety of authoritative sources referenced in user assertions [39]. We encompass this by considering two kinds of driver licenses and two kinds of graduate certificates that are distributed among n users (see Table 1).

Drivers Licenses that are issued in the state of Victoria (Australia) follow a Graduate Licensing Scheme where a subject must go through a "`Restricted`" phase. This phase lasts until a specific level of driving experience is accumulated by the subject at which stage exams need to be passed. Successful completion of these steps would allow an individual to transition into an "`Unrestricted`" phase. The described flow implies that an individual can have a driver license assertion with AVM that refers to either a restricted or non-restricted type. However, an individual cannot have digital driver licenses of both types. The data in Table 1 stipulate that as of 2019, 12.2% of users had restricted category license, whereas 87.8% of the users had unrestricted (i.e., full) license.

Similarly, an AVM for digital graduate certificate may designate that it is either "`Undergraduate`" or "`Postgraduate`." Due to the same reasons regarding the digital drivers license for Victoria, a person cannot possess a postgraduate certificate if they have only finished their undergraduate program. However, the final qualification level cannot be displayed at undergraduate level if one has finished a postgraduate program. From Table 1, it can be observed that 86.5% and 13.5% of users have undergraduate and postgraduate certificates, respectively.

We now formalize allocation of assertions among n users. Category $\mathcal{A} = \{\alpha_1, \alpha_2\}$ represents *driver licenses* where α_1 stands for "`Restricted`" AVM and α_2 stands for "`Unrestricted`" AVM. In addition, category $\mathcal{B} = \{\beta_1, \beta_2\}$ represents *graduate certificates* where β_1 denotes '`Undergraduate`' AVM and β_2 denotes "`Postgraduate`" AVM. We use indices $i \in \{1, \dots, n\}$ to differentiate users. A pair of random variables $(\upalpha^{(i)}, \upbeta^{(i)})$, $\upalpha^{(i)} \in \mathcal{A}$, $\upbeta^{(i)} \in \mathcal{B}$ encodes *type* of a user with index i. According to Table 1, $\Pr(\upalpha^{(i)} = \alpha_1) = 0.122$, $\Pr(\upalpha^{(i)} = \alpha_2) = 0.878$, and $\Pr(\upbeta^{(i)} = \beta_1) = 0.865$, $\Pr(\upbeta^{(i)} = \beta_2) = 0.135$. Assuming that $\upalpha^{(i)}$ and $\upbeta^{(i)}$ are independent, we obtain the following joint probabilities: $\Pr(\upalpha^{(i)} = \alpha_1, \upbeta^{(i)} = \beta_1) \approx 0.106$, $\Pr(\upalpha^{(i)} = \alpha_1, \upbeta^{(i)} = \beta_2) \approx 0.017$, $\Pr(\upalpha^{(i)} = \alpha_2, \upbeta^{(i)} = \beta_1) \approx 0.759$, and $\Pr(\upalpha^{(i)} = \alpha_2, \upbeta^{(i)} = \beta_2) \approx 0.118$. Further, we will use matrix $\aleph$ to represent this



(a) Users select and use 1 attribute only. Majority selects *driver license*.

(b) Users use both of their attribute realizations interchangeably with $\mathbb{E}[s_i] = 0.5$.

Fig. 4. Unlinkability under various conditions.

distribution of types across the users. For Table 1, we obtain $\aleph = [0.122,\ 0.878]^{\mathrm{T}} \times [0.865,\ 0.135]$. Since every user i has only two alternatives, their *decision* of interchangeable usage in ABA can be represented using scalar value $s_i = \Pr(\alpha^{(i)} \mid i)$, which is a continuous strategy in general.

Next, we need to analyze how adherence to the de facto agreement affects unlinkability and how this unlinkability depends on the number n of users in the system. For this, we consider two kinds of agreements for which the results of corresponding experiments are presented in Figure 4. For simplicity, we assume that every user authenticates to RP with the probability equal to $\frac{1}{n}$. The latter information coupled with the information about $\aleph$ reflects Assumptions 2 and 3. As we shall see later in the text, this implies that $C \leq \log n$. From this stems our motivation to measure performance against normalized unlinkability rate $\frac{C}{\log n}$.

Results for the agreement to use only 1 of the 2 available assertions are presented in Figure 4(a). The agreed strategy for user i is denoted as $\dot{s}_i$ while actual strategy played by the user is s_i. According to the agreement, users must use driver license only, meaning that $\forall i(\dot{s}_i = 1)$. In this instance, we consider that strategy of each user is a discrete random variable, $s_i \in \{0, 1\}$. As per the figure, unlinkability increases with the percentage of users who adhere to the agreement. Results for different agreements are depicted in Figure 4(b). In this instance, users agree to use both of their assertions interchangeably and with equal probability, meaning that $\forall i(\dot{s}_i = 0.5)$. The strategy played by each user is a continuous random variable $s_i \in [0, 1]$. Contrary to the previous agreement, the degree of deviations is reflected with the variance. This is because the mere fact that i deviates from the agreement does not communicate "*how strong*" the deviation is. As can be seen from the figure, performance of the system is better for the cases with lower variance (e.g., better adherence to the agreement). For both experiments in Figure 4, ratio $\frac{C}{\log n}$ increases with n, meaning that importance of the de facto agreement reduces. Despite this fact, the importance of coordinating agreement for small- and-medium-sized RPs cannot be underestimated.

While the aforesaid agreements are beneficial (and departures are harmful), they are exemplar only. It is therefore premature to consider these special cases optimal, since many other agreements may be arranged for a system of n users whose types are distributed in accordance with $\aleph$.

3.4 Game Theory for Unlinkable Authentication

We argue that game theory needs to be included in the methodology to explore the RQ further. This is because game theory studies player decisions, the effect these decisions have on the utility of that player and other players in the system. This is relevant for ABA and was demonstrated in the **Examples #1 & #2**: Coordination plays paramount role in maximizing C. To coordinate, for every user i, information about decisions of other users (we denote them $-i$) must be communicated or assumed in the form of belief. Questions of information, its meaning, origins, and tools for its dissemination are often asked within game-theoretical contexts. Moreover, a deeper line of thoughts guides us through the situations where tools for information synchronization (such as agreement) are not existent or cannot be trusted by i. Some of the concepts are proven to remain consistent irrespective of the decisions made by others and are often applied in robust decision-making [71]. Finally, results of non-cooperative game theory can enhance unlinkability on the individual level: Optimal recommendations for interchangeable usage of assertions can be provided to each user. This does not require modifications of existing authentication protocols.

4 GAME-THEORETICAL APPROACH TO PRIVACY IN AUTHENTICATION SYSTEMS

In this section, we formalize game-theoretical model for the attribute-based authentication system with n users who use their assertions interchangeably. Decisions of the players are guided by the principle of *best response*. To calculate it, we find expressions for the expected utilities of the players. For this, we utilize results of Lemma 1 and make an assumption about information that is known to the players. This information is in the form of priors (or beliefs) over the set $\mathbf{M}$ of marginal probabilities at RP and can, for example, be communicated to the players through mediator M. To ensure that this information is consistent with the principle of best response, we express conditions for Nash equilibrium and use these conditions to define $\mathbf{M}$. However, for the case when information about $\mathbf{M}$ cannot be communicated (trusted M does not exist) to the players, we utilize the Wald maximin approach.

Throughout this section, we consider that all users are players in the game. We also presume that the terms "attributes," "credentials," and "assertions" have the same meaning here.

4.1 Game-theoretical Model

We define a game $\eth = \langle I, \mathcal{T}, \Pi, \mathcal{S}, u \rangle$, where $I = \{1, 2, \ldots, i, \ldots, n\}$ denotes the set of indices i for the players (we will use $-i$ to denote all other players except i); $\mathcal{T} = \mathbf{t}_1 \times \cdots \times \mathbf{t}_i \times \cdots \times \mathbf{t}_n$ is the set of all players' types where $\mathbf{t}_i = (\alpha^{(i)}, \beta^{(i)})$. Random variables $\alpha^{(i)}$ and $\beta^{(i)}$ are drawn from distinctly different categories $\mathcal{A} = \{\alpha_1, \ldots, \alpha_\iota, \ldots, \alpha_l\}$ and $\mathcal{B} = \{\beta_1, \ldots, \beta_\rho, \ldots, \beta_m\}$, respectively, $\mathcal{A} \cap \mathcal{B} = \varnothing$ and $\ell = \mathcal{A} \cup \mathcal{B}$. Discrete joint **p**robability **m**ass **f**unction (**pmf**) $\aleph$ describes distribution of $(\alpha^{(i)}, \beta^{(i)})$ over $\mathcal{A} \times \mathcal{B}$ such that $\Pr(\alpha^{(i)} = \alpha_\iota, \beta^{(i)} = \beta_\rho) = \aleph_{\iota,\rho}$. $\Pi = \boldsymbol{\pi}_1 \times \cdots \times \boldsymbol{\pi}_n$ is the set of discrete pure strategies for all players, e.g., $\pi_{i,t_i} \in \boldsymbol{\pi}_i$ denotes t_ith strategy available to player i; $\mathcal{S} = S_1 \times \cdots \times S_n$ is the set of all continuous strategies for the players, containing subsets S_i including (perhaps infinitely many) probability vectors $\mathbf{s}_i : \boldsymbol{\pi}_j \to [0,1]^{|\boldsymbol{\pi}_i|}$ defining continuous strategies, such that $\mathbf{s}_i(\pi_{i,t_i}) \geq 0$ and $\sum_{\boldsymbol{\pi}_i} \mathbf{s}_i(\pi_{i,t_i}) = 1$; $u_i : \mathcal{T} \times \mathcal{S} \to \mathbb{R}$ is a payoff function of player i over a profile of types and continuous strategies (see Table 2).

Throughout the article, the attribute selection is restricted to just 2 alternatives. As a result, pure strategies of player i will be represented by $\boldsymbol{\pi}_i = \{\pi_{i,1}, \pi_{i,2}\}$ where $\pi_{i,1}$ should be interpreted as "player i authenticates to RP with the realization of $\alpha^{(i)}$," and $\pi_{i,2}$ should be interpreted as "player i authenticates to RP with the realization of $\beta^{(i)}$." Then, continuous strategy can be expressed with a scalar $s_i \in [0,1]$ where s_i is the probability $\Pr(\alpha^{(i)} \mid i)$, while $1 - s_i$ is the probability $\Pr(\beta^{(i)} \mid i)$. This should be interpreted as "player i authenticates to RP with the realization of $\alpha^{(i)}$ in $100 s_i\%$ of authentication sessions (randomly selected by him) while in the rest $100(1 - s_i)\%$ of all sessions he uses realization of $\beta^{(i)}$."

Table 2. Important Notations

Notation	**Description**
GAPAS	Game-theoretical Approach to Privacy in Authentication Systems
VC, VP	Verifiable Credentials, Verifiable Presentation
RP, AP	Relying Party, Attribute Provider
ROC	Receiver Operating Characteristics
DCW	Digital Credentials Wallet
$n \geq 2$	Number of players in the game
$\mathcal{A} = \{\alpha_1, \ldots, \alpha_l\}$, $\mathcal{B} = \{\beta_1, \ldots, \beta_m\}$	Categories of attribute realizations.
$\ell = \mathcal{A} \cup \mathcal{B}$	Full set of user attribute realizations.
$l \in \ell$	Discrete random variable in set ℓ
$\mathcal{L} = \{A, B\}$	Set of user labels A, B in 2-player game.
$\mathsf{L} \in \mathcal{L}$	Discrete random variable in set $\mathcal{L}$.
$I = \{1, \ldots, n\}$	Set of indices of the players in n-player game, $n \gg 2$.
$i \in I$	Discrete random variable (player's index) in set I.
$\alpha^{(i)} \in \mathcal{A}$, $\beta^{(i)} \in \mathcal{B}$	Attributes of a random player i
$H(\cdot\|\cdot)$	Conditional entropy.
$\mathbf{t}_i = (\alpha^{(i)}, \beta^{(i)})$	Type of player i.
$\mathcal{T} = \{\mathbf{t}_1, \ldots \mathbf{t}_n\}$	Set of types for all n players.
$\Pi := \boldsymbol{\pi}_1 \times \cdots \times \boldsymbol{\pi}_n$	Set of pure strategies $\boldsymbol{\pi}_i$ of all players $i \in I$.
$S_i, i \in I$	Subset consisting of continuous strategies $\boldsymbol{s}_i$
$\mathcal{S} := S_1 \times \cdots \times S_n$	Set of all continuous strategies for the players.
u_i	Payoff function of player $i, i \in I$
$\Game = \langle I, \mathcal{T}, \Pi, \mathcal{S}, u \rangle$	Game over the sets $I, \mathcal{T}, \Pi, \mathcal{S}, u$.
$\Omega_{(\cdot)}$	Set of players with realization $(\cdot)$
$\boldsymbol{\aleph}$	Discrete joint probability mass function over $(\alpha^{(i)}, \beta^{(i)})$.
$\boldsymbol{\beth}$	Distribution over $\mathcal{T} \times \mathcal{S}$.
$\Pr_{\mathcal{S}}(\alpha^{(i)}), \Pr_{\mathcal{S}}(\beta^{(i)})$	Marginal probabilities at RP (for $\mathbf{t}_i$)
$\mathbf{M}$	Complete set of marginal probabilities at RP
$\mathbb{E}[\cdot]$	Expected value
$\mathrm{Var}(\cdot)$	Variance

Definition 8 (GAPAS). Game-theoretical Approach to Privacy in Authentication Systems enables user i to increase unlinkability. This is done by maximising $\mathbb{E}[u_i]$ in game $\Game$, which requires adjusting s_i.

The overview of (AP, RP, U) model where n users produce their decisions is provided on Figure 5. It describes the following interactions:

(a) The AP supplies ready-to-use attributes/assertions to n players;

(b) for every player i, the pair of received attributes $(\alpha^{(i)}, \beta^{(i)})$ specifies his/her type $\mathbf{t}_i$. Every i decides upon continuous strategy $s_i = \Pr(\alpha^{(i)} \mid i)$. All n users then authenticate to RP on multiple occasions during some prolonged time period t and use their attributes interchangeably according to s_i;

(c) RP authenticates users and registers all authentication events. As such, he knows how frequently every of attribute realization from $\mathcal{A} = \{\alpha_1, \ldots, \alpha_\iota, \ldots, \alpha_l\}$ and $\mathcal{B} = \{\beta_1, \ldots, \beta_\rho, \ldots, \beta_m\}$,

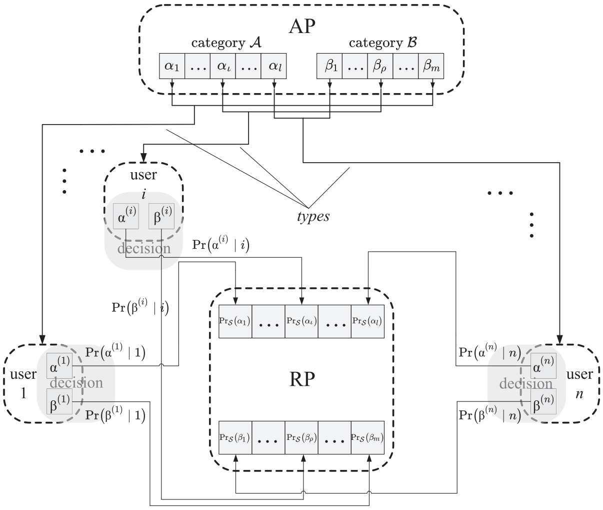

AP
category $\mathcal{A}$
α_1 ... α_ι ... α_l
category $\mathcal{B}$
β_1 ... β_ρ ... β_m
user i
$\alpha^{(i)}$
$\beta^{(i)}$
decision
types
$\Pr(\alpha^{(i)} \mid i)$
$\Pr(\beta^{(i)} \mid i)$
$\Pr_\mathcal{S}(\alpha_1)$ ... $\Pr_\mathcal{S}(\alpha_\iota)$ ... $\Pr_\mathcal{S}(\alpha_l)$
RP
$\Pr_\mathcal{S}(\beta_1)$ ... $\Pr_\mathcal{S}(\beta_\rho)$ ... $\Pr_\mathcal{S}(\beta_m)$
user 1
$\alpha^{(1)}$
$\beta^{(1)}$
decision
$\Pr(\alpha^{(1)} \mid 1)$
$\Pr(\beta^{(1)} \mid 1)$
user n
$\alpha^{(n)}$
$\beta^{(n)}$
decision
$\Pr(\alpha^{(n)} \mid n)$
$\Pr(\beta^{(n)} \mid n)$

Fig. 5. Schematic explanation of interactions among n users, AP, and RP.

is used but does not know the label i (or "*id*") of the user who authenticates to RP with specific attribute realization at specific authentication session. We call these relative frequencies $\Pr_\mathcal{S}(\alpha^{(i)})$, $\Pr_\mathcal{S}(\beta^{(i)})$ "*marginal probabilities at RP*": Every user i needs to estimate them to produce his/her *best response.*

Definition 9 (Marginal Probabilities at RP). The set $\mathbf{M}$ of marginal probabilities at RP is given as follows:
$\mathbf{M} = \bigcup_{i=1}^{n} \{\Pr_\mathcal{S}(\alpha^{(i)}), \Pr_\mathcal{S}(\beta^{(i)})\}$ where

$$\begin{aligned}\Pr_\mathcal{S}\left(\alpha^{(i)}\right) &= \sum_{j=1}^{n} \Pr\left(\alpha^{(i)} = \alpha^{(j)}\right)\Pr(j)s_j,\\ \Pr_\mathcal{S}\left(\beta^{(i)}\right) &= \sum_{j=1}^{n} \Pr\left(\beta^{(i)} = \beta^{(j)}\right)\Pr(j)(1-s_j).\end{aligned}$$

We stress the difference between (i) marginal probabilities $\Pr(\alpha^{(i)})$, $\Pr(\beta^{(i)})$ at AP, which is the probability that a randomly supplied attribute (to a randomly selected player i) takes certain realization from $\mathcal{A}$, $\mathcal{B}$, respectively—this can be obtained from $\aleph$, and (ii) marginal probabilities $\Pr_\mathcal{S}(\alpha^{(i)})$, $\Pr_\mathcal{S}(\beta^{(i)})$ at RP. In deterministic settings, this can be either directly **provided** by RP (or by independent mediator M) **to** the players or can be **calculated by** the players if the set of all decisions $\{s_1, \dots, s_i, \dots, s_n\}$ is known to the players. Also, information about elements in $\mathbf{M}$ may be non-deterministic for players, and, hence, we will talk about **priors** over $\mathbf{M}$.

4.2 Model Analysis

Here we establish the relation between $\Pr_\mathcal{S}(\alpha^{(i)})$, $\Pr_\mathcal{S}(\beta^{(i)})$ and player decisions, on the one hand, and C, on the other hand. We then use C to derive expected utilities $\mathbb{E}[u_i]$ and to calculate best responses s_i^b.

Based on Lemma 1, collective unlinkability of the entire authentication system with n users is $C = H(i \mid l)$:

$$\begin{aligned} C = -\sum_{i=1}^{n} \Pr(i) \Bigg(&\Pr\left(\alpha^{(i)} \mid i\right) \log \frac{\Pr\left(\alpha^{(i)} \mid i\right)}{\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)} \Pr(i) \\ &+ \Pr\left(\beta^{(i)} \mid i\right) \log \frac{\Pr\left(\beta^{(i)} \mid i\right)}{\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)} \Pr(i) \Bigg) = \sum_{i=1}^{n} \Pr(i) \log \frac{1}{\Pr(i)} \\ &- \sum_{i=1}^{n} \Pr(i) \Bigg(\Pr\left(\alpha^{(i)} \mid i\right) \log \frac{\Pr\left(\alpha^{(i)} \mid i\right)}{\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)} \\ &+ \Pr\left(\beta^{(i)} \mid i\right) \log \frac{\Pr\left(\beta^{(i)} \mid i\right)}{\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)} \Bigg), \end{aligned} \tag{1}$$

where $\Pr(\alpha^{(i)} \mid i) = s_i$ and $\Pr(\beta^{(i)} \mid i) = 1 - s_i$. Assuming that $\forall i \Pr(i) = \frac{1}{n}$, we can rewrite Equation (1) as

$$C = \log n - \frac{1}{n} \sum_{i=1}^{n} \left(s_i \log \frac{s_i}{\Pr_{\mathcal{S}}(\alpha^{(i)})} + (1 - s_i) \log \frac{1 - s_i}{\Pr_{\mathcal{S}}(\beta^{(i)})} \right). \tag{2}$$

Our task is then to propose utilities for the players such that every player maximizing his/her utility will maximize $\mathbb{E}\left[C\right]$.

Assumption 7. *Priors over* **M** *satisfy the following scaling constraint for all $i \in I$:*

$$\frac{\mathrm{Var}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]^2} = \frac{\mathrm{Var}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]^2} = \text{const.}$$

Lemma 2. *Expected utility for player i is (for details see Appendix A)*

$$\mathbb{E}\left[u_i\right] \approx -s_i \log \frac{s_i}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]} - \left(1 - s_i\right) \log \frac{1 - s_i}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]}. \tag{3}$$

Based on Lemma 2, we derive best response for player $i \in I$.

Corollary 1. *For* ***continuous*** *strategies,* ***best response of player*** *i is defined as*

$$s_i^{\mathrm{b}} = \frac{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right] + \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]}, \tag{4}$$

while for i playing ***discrete*** *strategies*

$$s_i^{\mathrm{b}} = \begin{cases} 1 & \text{if } \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right] > \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]; \\ 0 & \text{if } \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right] < \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right], \end{cases} \tag{5}$$

and, i is ***indifferent*** if $\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right] = \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]$.

The proof for Corollary 1 is straightforward, and we omit it here.

Remark 1. **Players of the same type** produce identical best responses and have identical best expected utilities:

$$\begin{aligned}\forall i, j(\mathbf{t}_j = \mathbf{t}_i) \implies \mathbb{E}\left[u_j^{\mathrm{b}}\right] = \mathbb{E}\left[u_i^{\mathrm{b}}\right] \\ = \log\left(\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right] + \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]\right).\end{aligned} \tag{6}$$

The result of Remark 1 follows directly from the expressions within Definition 9 and Corollary 1. Next, we need to address question of consistency of expectations [41].

Definition 10 (Consistent Expectations). Best responses $\mathcal{S}^{\mathrm{b}}$ of all n players must satisfy for all $i \in I$:

$$\Pr_{\mathcal{S}^{\mathrm{b}}}\left(\alpha^{(i)}\right) = \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right] \wedge \Pr_{\mathcal{S}^{\mathrm{b}}}\left(\beta^{(i)}\right) = \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]. \tag{7}$$

One way to enable Equation (7) is to find the conditions supporting Nash equilibrium in the form

$$s_i^{\mathrm{b}} = \frac{\Pr_{\mathcal{S}^{\mathrm{b}}}\left(\alpha^{(i)}\right)}{\Pr_{\mathcal{S}^{\mathrm{b}}}\left(\alpha^{(i)}\right) + \Pr_{\mathcal{S}^{\mathrm{b}}}\left(\beta^{(i)}\right)} \quad \text{for all } i \in I \tag{8}$$

and then to make sure that the priors on $\mathbf{M}$ are formed accordingly. This can be communicated to the players through a mediator M. We will next formulate Equation (8) using information of players' types. This is possible because the whole set of players' indices I can be represented using subsets that have direct reference to all possible realizations of $\alpha^{(i)}$, $\beta^{(i)}$. We use Ω_{α_ι} and Ω_{β_ρ} such that $\forall \phi, \xi \in I$

$$\left(\alpha^{(\xi)} = \alpha_\iota\right) \iff \left(\xi \in \Omega_{\alpha_\iota}\right), \ \left(\beta^{(\phi)} = \beta_\rho\right) \iff \left(\phi \in \Omega_{\beta_\rho}\right).$$

We denote $\Omega_{\alpha_\iota,\beta_\rho} = \Omega_{\alpha_\iota} \cap \Omega_{\beta_\rho}$ for which $\Omega_{\alpha_\iota} = \bigcup_{\rho=1}^{m} \Omega_{\alpha_\iota,\beta_\rho}$ and $\Omega_{\beta_\rho} = \bigcup_{\iota=1}^{l} \Omega_{\alpha_\iota,\beta_\rho}$ holds. For $i \in \Omega_{\alpha_\iota,\beta_\rho}$ we set $\Pr_{\mathcal{S}^{\mathrm{b}}}(\alpha^{(i)}) = \frac{1}{n}\sum_{\xi\in\Omega_{\alpha_\iota}} s_\xi^{\mathrm{b}}$ and $\Pr_{\mathcal{S}^{\mathrm{b}}}(\beta^{(i)}) = \frac{1}{n}\sum_{\phi\in\Omega_{\beta_\rho}} (1 - s_\phi^{\mathrm{b}})$.

In line with Remark 1, all players whose indices are in $\Omega_{\alpha_\iota,\beta_\rho}$ produce the same best response denoted as $\theta_{\iota,\rho}$, and, as a result $\Pr_{\mathcal{S}^{\mathrm{b}}}(\alpha^{(i)}) = \frac{1}{n}\sum_{\nu=1}^{m} |\Omega_{\alpha_\iota,\beta_\nu}|\theta_{\iota,\nu}$, while $\Pr_{\mathcal{S}^{\mathrm{b}}}(\beta^{(i)}) = \frac{1}{n}\sum_{\tau=1}^{l} |\Omega_{\alpha_\tau,\beta_\rho}|(1 - \theta_{\tau,\rho})$. Without loss of generality, for large $n >> l \times m$, we have $\aleph_{\iota,\rho} = \frac{1}{n}|\Omega_{\alpha_\iota,\beta_\rho}|$. Validity of Equation (8) is guaranteed if for all ι, ρ:

$$\theta_{\iota,\rho} = \frac{\sum_{\nu=1}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu}}{\sum_{\nu=1}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum_{\tau=1}^{l} \aleph_{\tau,\rho}(1 - \theta_{\tau,\rho})}. \tag{9}$$

Next, we discuss solutions for Equation (9) in pure continuous (e.g., authentication with 2 attributes) and pure discrete (e.g., authentication with 1 attribute) strategies. We will also consider the maximin scenario for the case when neither $\mathbf{M}$ nor priors over $\mathbf{M}$ are known. We will further separate these results by referring to "Naïve game" when $\mathbf{M}$ is provided (by mediator M, for instance) and "Tenable game" when nothing is known about $\mathbf{M}$, respectively.

4.3 Naïve Game and Its Equilibria

This game is based on the assumption that $\mathbf{M}$ is known, and some of its properties were discussed in the previous subsection.

4.3.1 Players use 2 Attributes Interchangeably. We transform Equation (9) and take into account that sums $\sum_{\nu=1}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu}$ and $\sum_{\tau=1}^{l} \aleph_{\tau,\rho}(1-\theta_{\tau,\rho})$ have terms with common $\aleph_{\iota,\rho}$. All the possible Nash equilibria in (pure) continuous strategies are represented by the following system of nonlinear equations:

$$\theta_{1,1}\left(\sum_{\nu=2}^{m} \aleph_{1,\nu}\theta_{1,\nu} + \sum_{\tau=2}^{l} \aleph_{\tau,1}(1-\theta_{\tau,1})\right) = \sum_{\nu=2}^{m} \aleph_{1,\nu}\theta_{1,\nu},$$

$$\vdots$$

$$\theta_{\iota,\rho}\left(\sum_{\substack{\nu=1\\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum_{\substack{\tau=1\\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}(1-\theta_{\tau,\rho})\right) = \sum_{\substack{\nu=1\\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu},$$

$$\vdots$$

$$\theta_{l,m}\left(\sum_{\nu=1}^{m-1} \aleph_{l,\nu}\theta_{l,\nu} + \sum_{\tau=1}^{l-1} \aleph_{\tau,m}(1-\theta_{\tau,m})\right) = \sum_{\nu=1}^{m-1} \aleph_{l,\nu}\theta_{l,\nu}. \tag{10}$$

4.3.2 Players Use 1 Attribute. As a result, each player only plays a discrete pure strategy. If all players with the same (ι,ρ) produce the same best response, then their averaged response is also discrete, e.g., $\theta_{\iota,\rho} \in \{0,1\}$. However, averaged response $\bar{\theta}_{\iota,\rho}$ of the players of the same type may be a continuous value if different players of type (ι,ρ) play different pure discrete strategies. This is only possible if players of type (ι,ρ) are indifferent as to which among two strategies to play (see Corollary 1). Taking into account all possible attribute realizations, this latter condition is represented by the following linear system:

$$\begin{cases} \sum_{\nu=1}^{m} \aleph_{1,\nu}\bar{\theta}_{1,\nu} & = \sum_{\tau=1}^{l} \aleph_{\tau,1}(1-\bar{\theta}_{\tau,1}), \\ & \vdots \\ \sum_{\nu=1}^{m} \aleph_{\iota,\nu}\bar{\theta}_{\iota,\nu} & = \sum_{\tau=1}^{l} \aleph_{\tau,\rho}(1-\bar{\theta}_{\tau,\rho}), \\ & \vdots \\ \sum_{\nu=1}^{m} \aleph_{l,\nu}\bar{\theta}_{l,\nu} & = \sum_{\tau=1}^{l} \aleph_{\tau,m}(1-\bar{\theta}_{\tau,m}). \end{cases} \tag{11}$$

4.4 Tenable Game and Its Equilibria

Here we presume that player i makes decisions under uncertainty about the priors on $\mathbf{M}$. One way to address this uncertainty is to apply Wald maxi-min principle requiring i to consider the worst-case scenario played by $-i$ [71]. The expected utility of i is then

$$\mathbb{E}_w[u_i] = \max_{s_i} \min_{\mathcal{S}_{-i}} \mathop{\mathbb{E}}_{\beth_i}[u_i], \tag{12}$$

where $\beth_i$ is the distribution over $\mathbf{t}_i \times S_i$. This can be ruminated as a special case of the "naïve" game where i calculates her best response using Corollary 1 in which instead of $\mathbb{E}[\Pr_{\mathcal{S}}(\alpha^{(i)})]$ and $\mathbb{E}[\Pr_{\mathcal{S}}(\beta^{(i)})]$ she substitutes worst possible estimates $\mathbb{E}[\Pr_{\mathcal{S}^w}(\alpha^{(i)})]$ and $\mathbb{E}[\Pr_{\mathcal{S}^w}(\beta^{(i)})]$, respectively. The following assumption explains $\beth_i$.

According to Equation (12) and Assumption 5 $-i$ minimizes the best utility $\mathbb{E}[u_i^{\mathrm{b}}]$, which is given by Equation (6) while $\beth_i$ reduces to the distribution over $\mathbf{t}_i$ only (which is $\aleph$). To calculate value of $\mathbb{E}_w[u_i^{\mathrm{b}}]$ we require $\mathcal{S}^w_{-i}$:

$$\mathcal{S}^w_{-i} = \arg\min_{\mathcal{S}_{-i}} \mathop{\mathbb{E}}_{\beth_i}\left[u_i^{\mathrm{b}}\right] \sim \arg\min_{\mathcal{S}_{-i}} \mathop{\mathbb{E}}_{\beth_i}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right) + \Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]. \tag{13}$$

To complete our calculations for the expectation over $\aleph$ we, as previously, use notation $\theta_{\iota,\rho}$:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\beth_i}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right) + \Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right] \\
&= \mathop{\mathbb{E}}_{\beth_i}\left[\aleph_{\iota,\rho} + \sum_{\substack{\nu=1\\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum_{\substack{\tau=1\\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}(1-\theta_{\tau,\rho})\right] \\
&= \sum_{\iota=1}^{l}\sum_{\rho=1}^{m}\left(\aleph^2_{\iota,\rho} + \aleph_{\iota,\rho}\sum_{\substack{\nu=1\\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} + \aleph_{\iota,\rho}\sum_{\substack{\tau=1\\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}\left(1-\theta_{\tau,\rho}\right)\right).
\end{aligned} \tag{14}$$

In the last line of Equation (14), we ignore $\aleph^2_{\iota,\rho}$ for the calculation of

$$\arg\min_{\theta} \sum_{\iota=1}^{l}\sum_{\rho=1}^{m}\aleph_{\iota,\rho}\left(\sum_{\substack{\nu=1\\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum_{\substack{\tau=1\\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}(1-\theta_{\tau,\rho})\right), \tag{15}$$

from which we conclude that for all ι, ρ:

$$\theta_{\iota,\rho} = \begin{cases} 0, & \text{if } \sum\limits_{\substack{\nu=1\\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu} \geq \sum\limits_{\substack{\tau=1\\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}; \\ 1, & \text{otherwise.} \end{cases} \tag{16}$$

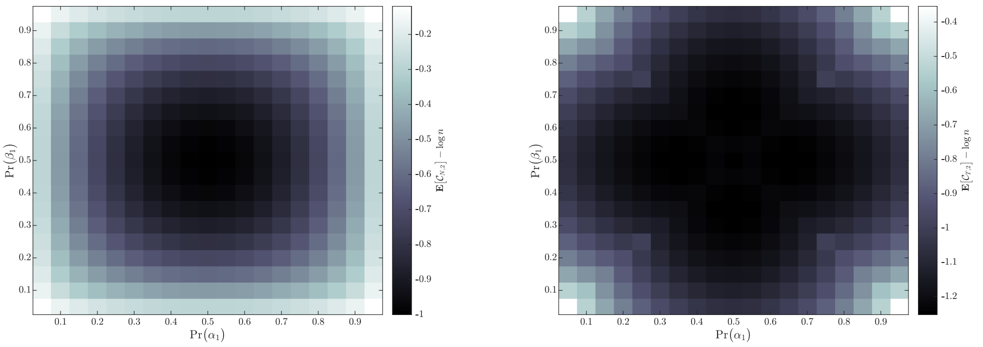
5 EXPERIMENT

To evaluate the impact of GAPAS on privacy in attribute-based authentication, we assess our game-theoretical results by conducting numerical evaluations for the system with $n \gg 2$ users.

The goal of experiment. We compare (i) unlinkability in naïve game (e.g., game with mediator) with the unlinkability in tenable game (e.g., maximin) and (ii) unlinkability in naïve and tenable games with the unlinkability in the system where users make "*alternative*" decisions. To find solutions for nonlinear systems, we run our experiment in Matlab using the trust region algorithm [22]. For the experiment, we require $\Pr(i), \aleph$. Based on Equation (9), we derive best response expressions that are identical among players i whose types $\mathbf{t}_i = \{\alpha^{(i)}, \beta^{(i)}\}$ match. As such, we further use $\theta_{\iota,\rho} = s_i$ for all i whose $\mathbf{t}_i$ realization is $(\alpha_\iota, \beta_\rho)$. We then define the systems of equations for equilibria in *naïve* as well as *tenable* game settings.

5.1 Experiment Organization

For our baseline scenarios, considerations are made for "unrestricted rationality" where two attribute realizations $\{\alpha^{(i)}, \beta^{(i)}\}$ available to player i can be used interchangeably in naïve and tenable games (see Section 4). We also analyze some of alternative scenarios with different kinds of "*irrationality*." While the discussion of many possible alternative decisions goes beyond the scope of our article we identify (a) "restricted rationality" where users play naïve or tenable game but



(a) Naïve game, players use 2 attributes interchangeably, e.g., $\theta_{\iota,\rho} \in [0,1]$.

(b) Tenable game, players use 2 attributes interchangeably, e.g., $\theta_{\iota,\rho} \in [0,1]$

Fig. 6. Comparison of expected unlinkability in naïve and tenable baseline scenarios.

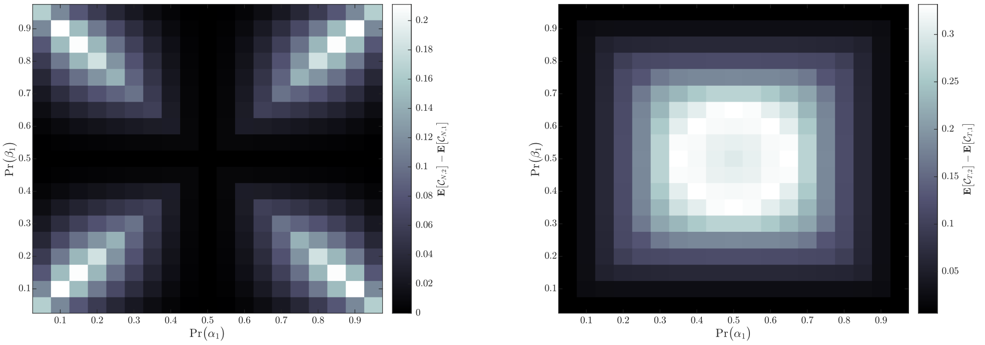
(in contrast to interchangeable usage) select and always use the same realization out of 2 realizations available to them (see Section 4.3.2) and (b) the "random move" scenario where users use both of their realizations interchangeably but in random manner, $\forall i,\ \varrho(s_i) = 1,\ s_i \in [0,1]$. We use compact notation for the unlinkability, which is obtained in different scenarios. Expected unlinkability $\mathbb{E}[C] = \log n + \sum_i \mathbb{E}[u_i]$ in rational scenarios is denoted by $\mathbb{E}[C_{\kappa,\mu}]$ where $\kappa \in \{N, T\}$ denotes either naïve (letter "N") or tenable (letter "T") game, respectively. $\mu \in \{1,2\}$ indicates the number of attribute realizations used by each player: $\mu = 1$ specifies games with restricted rationality; $\mu = 2$ specifies games with unrestricted rationality. Notation $\mathbb{E}[C_{\{\kappa,\mu\}^r}]$ is for expected unlinkability measured under random moves scenario (index "r").

To produce *outputs* in the form of expected unlinkability, our experiment requires the following *inputs*: (1) $\{\kappa,\mu\}$ or $\{\kappa,\mu\}^r$ and (2) Pr(i), for all players i and the **pmf** $\aleph$. For all the instances of experiment, we consider n users and $\Pr(i) = \frac{1}{n}$ for all i. We aim at conducting numerical evaluations for a wide range of various joint **pmf**s $\aleph$. For the purpose of convenient presentation and comparison of the outputs from the experiment, we depict corresponding unlinkability using two-dimensional heat maps (see Figures 6–8). Coordinates $(\Pr(\alpha_1), \Pr(\beta_1))$ of each point on the map define a corresponding 2×2 matrix $\aleph$: $\aleph = [\Pr(\alpha_1),\ 1 - \Pr(\alpha_1)]^{\mathrm{T}} \times [\Pr(\beta_1),\ 1 - \Pr(\beta_1)]$ where both $\Pr(\alpha_1), \Pr(\beta_1)$ were quantized with 0.05 step on interval $[0,\ 1]$. Color intensity corresponds to unlinkability.

5.2 Results

We first calculated the equilibria for our baseline scenarios on the naïve and tenable games where players can use both of their attribute realizations interchangeably (see Figure 6). For each possible $\aleph$ in naïve game, we solved complete information Nash equilibria (see Equation (10)) to find $\mathbf{M}$ that need to be communicated to the players by mediator. Among all the possible solutions, we selected those maximizing $\mathbb{E}[C_{N,2}]$. For each possible $\aleph$ in tenable game, we calculated the worst-case condition that may be created for player i by others $n-1$ players (see Equation (16)). Then the best response of i and $\mathbb{E}[C_{T,2}]$ are calculated (see Equation (4)). As can be observed from comparison of Figures 6(a) and 6(b), naïve game provides substantially better unlinkability.

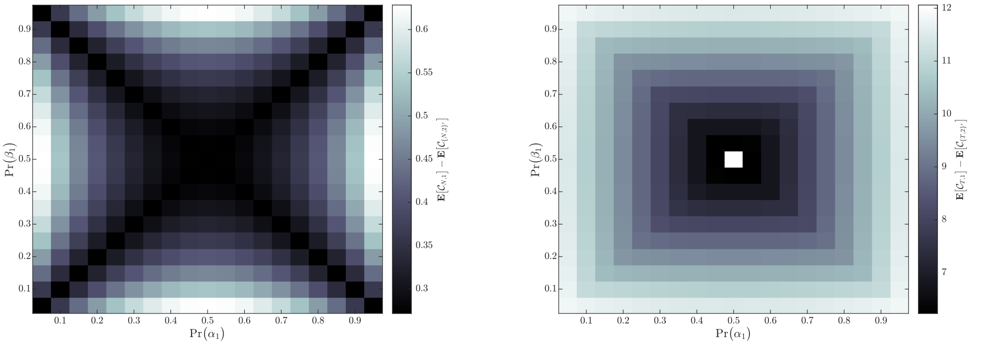
To compute equilibria for naïve games with single attribute usage (e.g., restricted rationality), we solved a linear system representing mixed and pure discrete equilibria (see Equation (11)). The benefits of using two attributes (unconstrained rationality) versus one attribute (constrained



(a) Difference between expected unlinkability $\mathbb{E}\left[C_{N,2}\right]$ and $\mathbb{E}\left[C_{N,1}\right]$.

(b) Difference between expected unlinkability $\mathbb{E}\left[C_{T,2}\right]$ and $\mathbb{E}\left[C_{T,1}\right]$.

Fig. 7. Residual expected unlinkability in "Naïve game" and "Tenable" games.



(a) Difference between expected unlinkability $\mathbb{E}\left[C_{N,1}\right]$ and $\mathbb{E}\left[C_{\{N,2\}^r}\right]$.

(b) Difference between expected unlinkability $\mathbb{E}\left[C_{T,1}\right]$ and $\mathbb{E}\left[C_{\{T,2\}^r}\right]$.

Fig. 8. Residual expected unlinkability in "Naïve" and "Tenable" games.

rationality) can be observed by comparing residual unlinkabilities on Figure 7, which are greater than 0 for both heatmaps.

We conducted a range of experiments with randomized moves, and the results are presented in Figure 8. For the two-attribute randomized game, each player i decides $0 \leq s_i \leq 1$ at random in accordance to uniform distribution on $[0, 1]$. As can be seen from the residuals of expected unlinkabilities, even constrained rationality (one attribute usage) scenario outperforms scenario where two realizations are used randomly (chaotically).

As a special case, we analyzed example described in Section 3.3 where $\aleph = \left(\begin{smallmatrix} 0.106 & 0.017 \\ 0.759 & 0.118 \end{smallmatrix}\right)$. We compared unlinkability in the system where users use one or two attributes in two variants of the game as well as randomized scenario (see Figure 9). A few important observations can be made in that regard. First, unlinkability achievable in *tenable* game is better than in randomized scenario, but it is worse than *naïve* game performance. Second, differences between relative performance $\frac{\mathbb{E}[C]}{\log n}$ for all these variants of the game change with n: For smaller n this becomes even more apparent.

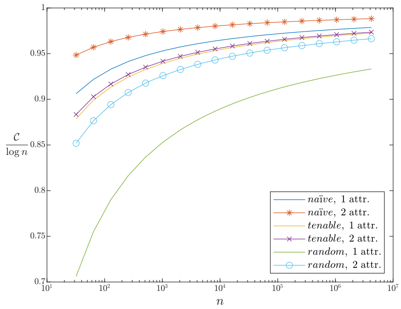


Fig. 9. Unlinkability for the example in Section 3.3 under various games.

6 APPLICATION OF RESULTS

Next, we demonstrate *why* interchangeable usage of assertions is plausible for the systems where VC format is accepted. This is achieved by further analysis of the details about practical implementation (as per Example #2) where VCs are used as a format for assertions. Due to multiple benefits (including advanced privacy preserving tools), the VC format becomes widely appreciated [18, 37, 72]. In addition, we suggest an algorithm that governs usage of these VC-compatible assertions within DCW.

6.1 VCs and Verifiable Presentations

Verifiable Credentials are a standard created by W3C. It allows a trusted issuer to issue tamper resistant statements (credentials and assertions) about attributes of the entities known to the issuer. Validity of assertions that comply with VC requirements can be unambiguously verified (e.g., RP). The primary utility of VC is that it expedites privacy, which contrasts with current practices of **Identity Providers (IdP)** in federated Identity Management systems. This privacy preserving features of VC are supposed to be enabled by the holders of the credentials that aligns with user-centric paradigm: They decide upon PII and control its disclosure in independent manner. This control is, however, not arbitrary, since resulting assertions must satisfy RP's trust requirements as discussed in Reference [39]. Improved security of identity management is among other advantages of user-centric paradigm: it mitigates the risk pertaining to a single point of failure (which is common for IdPs).

In VC-compatible assertions, control over information that is disclosed to RP can be exercised through ZKP, which enables a legitimate holder to produce **Verifiable Presentation (VP)**. The benefits of such approach can facilitate anonymous authentication for the addressee (e.g., future VC holder) the issuer can sign VC using, for example, Camenisch-Lysyanskaya or Boneh-Boyen signatures [11, 15]. This allows the holder (e.g., user) to produce multiple VPs where he/she selectively discloses the information about the attributes in the VC. The signatures produced by the holder for different VPs do not correlate, which has been widely advocated by the community as the main privacy preserving feature. This often creates a ground for the narrative that RP cannot link assertions from the same user if ZKP is used.

Unfortunately, benefits of anti-correlation properties of ZKP system are undermined by AVM and **Attribute Schema Metadata (ASM)**, which are present in VP [39]. This causes distinguishability: Interchangeable usage of attributes should therefore be practiced to avoid adverse effects on users' unlinkability in VC-compatible attribute-based authentication systems. To see why, let us analyze the case depicted on Figure 10[1] (see Appendix B for VC/VP details).

[1] Enlarged version of the diagram: https://cloudstor.aarnet.edu.au/plus/s/u9yU3cLSQOleHcW.

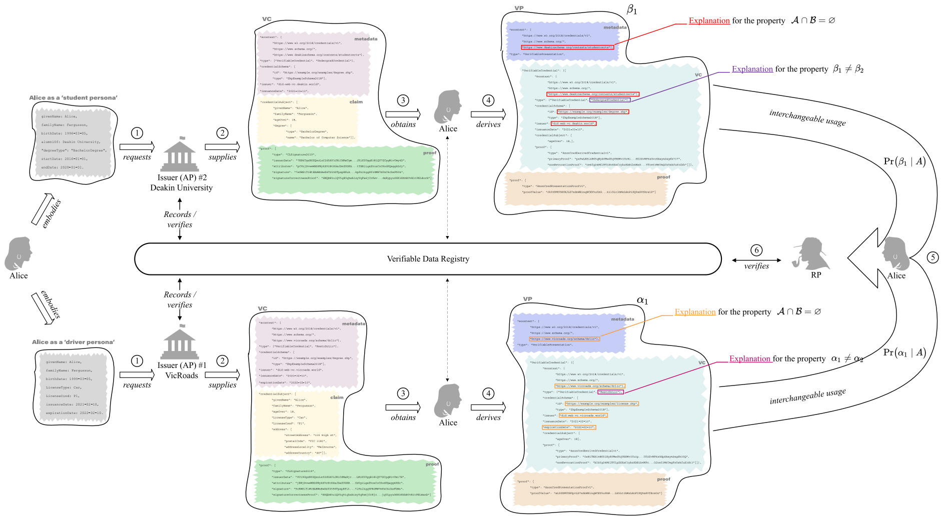


Fig. 10. Application of the proposed model for VC usage scenario.

Taking into account the flow in the (AP, RP, U) system where user *Alice* obtains her credentials from APs and RP grants her access to the products (e.g., allows to buy alcohol online) based on the policy P requiring that any user must be *"older than 18."* In this specific scenario, RP accepts (e.g., trusts the issuers of) the digital driver licenses (issued by VicRoads, Australia) and digital tertiary qualification certificates (issued by Australian universities). *Alice* embodies different personae, which allows her to claim credentials from two different issuers (see Figure 10). At ①, she requests VCs from AP #1 and AP #2. All the aspects of user authentication demanded by APs, and consider that these steps are successfully executed are omitted to ensure simplicity. This eventuates in ②, during which VCs are supplied to *Alice*. For better privacy, VCs can remain completely encrypted during that transmission step. This can be accomplished by using, for instance, a JSON Web Encryption file format [53]. To support verifiability of VCs, information about public keys of both APs must be recorded in a publicly accessible medium. The function of such medium is performed by *Verifiable Data Registry*, which can be implemented using Decentralized PKI, for example [58].

At ③, *Alice* obtains VCs and decrypts them if necessary. Major components within VC include *metadata*, *claims*, and *proof* (see Table 3). Properties of these components explain why assertions from different APs may differ. A claim is a statement about a subject, and a synonymous term *'subject attribute'* can be used as well. These statements are composed using attribute-value pairs: They are placed within `credentialSubject` and define its properties. In spite of being the most informative part of VC, claims (e.g., `credentialSubject`) may be insufficient for unambiguous interpretation, which is required by RP. Therefore, metadata (*meta* for short) carry functions of interpretation and validation of attribute-value pairs within `credentialSubject` [72]. As such, functions of meta in VC can be explained through the combination of AVM and ASM [39]. The purpose of the proof is to assert claims and meta in a way that is unique for the specific issuer (e.g., to uniquely authenticate AP). This is required to establish trust to the information in VC for both *Alice* and RP (since they already trust AP).

It can be observed (③, Figure 10) that multiple objects within *meta* and `credentialSubject` differ when VCs from VicRoads and Deakin University are compared. For example, for the two VCs obtained by *Alice*, at least one object within `@context` property differ: The credential

Table 3. Main Objects and Properties of VC and VP

Data concept	Object/property	Purpose	Presence	Category	Format example
VC, *meta*	`@context`	maps aliases to Uniform Resource Identifiers (URI)	compulsory	static	ordered set of URIs [8]
VC, *meta*	`id`	unambiguously refers to the credential	optional	static/dynamic	single URI (including DID) [8, 38]
VC, *meta*	`type`	defines type for objects within VC	compulsory	static	terms defined according to JSON-LD grammar [55]
VC, *claim*	`credentialSubject`	specifies the subject of the claim	compulsory	static/dynamic	a set of objects with properties related to the subject
VC, *meta*	`issuer`	specifies the issuer of VC	compulsory	static	URI or other id-object such as JWK or DID [3, 52]
VC, *meta*	`issuanceDate`	expresses the date and time when a VC becomes valid	compulsory	dynamic	combination of date and time strings, RFC3339 [56]
VC, *proof*	`proof`	issuer's assertion about information in VC	optional/compulsory	dynamic	RSA digital signatures, such as RsaSignature2018 [60]
VP, *meta*	`id`	unambiguously refers to the presentation	optional	static/dynamic	single URI (including DID) [8, 38]
VP, *meta*	`type`	defines type for objects within VP	compulsory	static	terms defined according to JSON-LD grammar [55]
VP, *meta*	`verifiableCredential`	construction of one or more VCs or derived VCs	optional	dynamic	VC data model v1.0 [72]
VP, *meta*	`holder`	specifies the entity that is generating VP	optional	static/dynamic	URI or other id-object such as JWK or DID [3, 52]
VP, *proof*	`proof`	authenticates VP holder to the verifier	optional/compulsory	dynamic	RSA digital signatures, such as RsaSignature2018 [60]

for driver license contains “`https://.../drlic`” while the credential for tertiary qualification contains “`https://.../studentcerts`”. Each of the other properties within metadata including `type`, `credentialSchema`, `issuer`, and `issuanceDate` also differ for those two VCs. It is remarkable that driver license VC contains `expirationDate` property, while the tertiary qualification certificate does not. Differences between those the VCs become even more apparent if the contents of `credentialSubject` are compared.

To improve her privacy in the attribute-based authentication system, *Alice* eliminates from her assertions as much of PII as possible. For this, she first modifies existing VC, VC ⇒ $\dot{\mathrm{VC}}$. This modification is due to the changes in original claims cl of that VC: *Alice* removes redundant information from cl, $cl \Rightarrow \dot{cl}$. Since structural parts of original VC change, a new proof $\dot{pr}(\dot{cl}, \dot{mt})$ must be produced for $\dot{\mathrm{VC}}$ where meta remains unchanged, e.g., $\dot{mt} = mt$. This modified $\dot{\mathrm{VC}}$ then becomes a part of $\ddot{\mathrm{VP}}$. To compose a valid $\ddot{\mathrm{VP}}$ it is also required to add $\ddot{mt}$ and to produce $\ddot{pr}(\ddot{mt}, \dot{\mathrm{VC}})$ (see ④, Figure 10):

$$\ddot{\mathrm{VP}} = \left[\ddot{pr}\left(\ddot{mt}, \underbrace{\left[\dot{pr}\left(\dot{cl}, \dot{mt}\right), \dot{cl}, \dot{mt}\right]}_{\dot{\mathrm{VC}}}\right),\ \ddot{mt}, \dot{\mathrm{VC}}\right].$$

As per the derivation procedure, $\dot{cl}$ becomes the only part of the resulting assertion where attributes of the subject are stated explicitly. The rule $\dot{cl} \subseteq cl$ must be obeyed during the derivation procedure. For instance, it can be seen that for both VPs derived by *Alice* that she has $\dot{cl} \leftarrow$ {“`ageOver" : 18`}. This conforms with the derivation rule and greatly reduces PII, which makes claims alone indistinguishable when driver license VP is compared with tertiary qualification VP. Nevertheless, the remaining payloads in both VPs contain enough information to differentiate assertions possessed by *Alice*.

This is because ZKP and selective disclosure can be performed by the holder over claims only (e.g., `credentialSubject`). As a result, *Alice* assertions that originate from VicRoads and Deakin University will always differ: This supports property $\mathcal{A} \cap \mathcal{B} = \varnothing$ in GAPAS. In addition, assertions that originate from AP #1 may differ for *Alice* and other users. This is due to the meta property `type` containing object `‘‘RestrDrLic’’`, which is used only with restricted driver licenses. In contrast, for unrestricted driver licenses `type` will contain `‘‘UnrestrDrLic’’` (not displayed on Figure 10) instead of `‘‘RestrDrLic’’` (for the details see **Example #2**).

This observation supports $\alpha_1 \neq \alpha_2$, which is important for our decision-making model. Also, assertions produced from restricted and unrestricted driver licenses belong to the same category $\mathcal{A}$, because they are issued by AP #1, but *Alice* cannot hold both valid α_1 and α_2. This also supports our assumption for joint distribution of assertions among players, which is defined by matrix $\aleph$. Similarly, we observe that VP produced by *Alice* from tertiary qualification certificate contains `UndergradCredential` as part of the `type`. There are other users who possess credentials from AP #2 and whose degree is postgraduate. As a result, VPs of those users will have `type`, which is different (not shown on Figure 10) to *Alice's* type. This is because those users' types contain `PostgradCredential` in contrast to `UndergradCredential`. This conforms with $\beta_1 \neq \beta_2$ while both β_1, β_2 belong to the same category $\mathcal{B}$. *Alice* can have only one assertion from $\mathcal{B}$.

Based on the discussed properties of assertions, we can uniquely specify *Alice* type in the game (not to be confused with `type` in VC/VP) as $\mathbf{t} = (\alpha_1, \beta_1)$ (see Figure 10). Depending on the information that is available to *Alice* about other players she can play either *naïve* or *tenable* variant of the game. This implies that during the sessions when she authenticates to RP she will use her assertions α_1 and β_1 *interchangeably* according to probabilities $\Pr(\alpha_1 \mid A)$ and $\Pr(\beta_1 \mid A)$, respectively (see ⑤, Figure 10). Finally, in each of authentication sessions RP can verify whether assertion submitted by *Alice* is valid (see ⑥, Figure 10).

6.2 Interchangeable Usage of Assertions in DCW

VCs and VPs can be held within a piece of software or a hardware device—also known as a DCW [68, 74]. Although different types of DCW's exist, the majority of them satisfy basic requirements related to the task of entity authentication. This includes (i) receiving and securely storing credentials (also includes requesting a credential from AP in some cases) and (ii) selective disclose of credential information, e.g., VC → VP. This implies that stages ③ and ④ on the diagram (see Figure 10) comprise common functionalities of DCWs, which provides a user-friendly yet secure way of authentication through Verifiable Credentials [68].

With the aim to improve user unlinkability while maintaining ease of use for VCs we propose to also incorporate *interchangeable usage* of assertions, which corresponds to stage ⑤ (see Figure 10) into the design of DCW. For example, best responses for the player in *tenable* variant of the game can be governed by Algorithm 1, which is derived from corresponding equilibrium conditions (see Corollary 1 and Equation (16)). It is based on maximin principle and does not require any additional information except $\aleph$. From Figure 10, *Alice* with type $\mathbf{t} = (\alpha_1, \beta_1)$ uses matrix $\aleph = \left(\begin{smallmatrix} 0.106 & 0.017 \\ 0.759 & 0.118 \end{smallmatrix}\right)$ (see Example #2 in Section 3.3). According to Algorithm 1 she calculates $\theta(1,1) = 1$, $\theta(1,2) = 1$, and $\theta(2,1) = 0$, from which she obtains $s = \frac{(\theta(1,1)\ \ \theta(1,2))\times(0.106\ \ 0.017)^{\mathrm{T}}}{(\theta(1,1)\ \ \theta(1,2))\times(0.106\ \ 0.017)^{\mathrm{T}}+(1-\theta(1,1)\ \ 1-\theta(2,1))\times(0.106\ \ 0.759)^{\mathrm{T}}} = \frac{0.123}{0.123+0.759} \approx 0.139$.

To ensure that for each of authentication sessions assertion (VP) α_1 is selected randomly with probability $\Pr(\alpha_1 \mid A) = 0.139$, and β_1 is selected randomly with probability $\Pr(\beta_1 \mid A) = 0.861$, user *Alice* runs uniform random generator with support on $[0, 1]$. If random number s^* from the generator is less or equal to s, then *Alice* authenticates to RP with the assertion derived from her driver license (e.g., α_1). Otherwise, she authenticates with the assertion derived from her tertiary degree certificate (e.g., β_1). If the proposed algorithm is implemented in DCW, then all the mentioned decisions will be made by the software that does not require *Alice* participation.

7 RELATED WORK

Below, we provide an analysis of sources contributing to the questions of privacy, unlinkability, and anonymity. First, we outline works that rather formalize the above-mentioned definitions.

ALGORITHM 1: Maximin decision algorithm

```
input  : $\aleph$, $\{\alpha_\iota, \beta_\rho\}$
output : $\mathrm{VP}(\alpha_\iota)$ or $\mathrm{VP}(\beta_\rho)$
begin
    $\mathbf{h}(\iota,\cdot) \leftarrow \left\{\bigcup_{\nu=1}^{m} \aleph_{\iota,\nu}\right\}$, $\mathbf{h}(\cdot,\rho) \leftarrow \left\{\bigcup_{\tau=1}^{l} \aleph_{\tau,\rho}\right\}$;
    $\boldsymbol{\theta}(\iota,\cdot) \leftarrow \{0\}_m$, $\boldsymbol{\theta}(\cdot,\rho) \leftarrow \{0\}_l$, $s \leftarrow 0$, $s^* \leftarrow 0$;
    for $\nu \leftarrow 1$ to $m$ do
        if $\sum_{\substack{\psi=1\\ \psi\neq\nu}}^{m} \aleph_{\iota,\psi} \geq \sum_{\substack{\gamma=1\\ \gamma\neq\iota}}^{l} \aleph_{\gamma,\nu}$ then $\theta(\iota,\nu) \leftarrow 0$;
        else $\theta(\iota,\nu) \leftarrow 1$;
    for $\tau \leftarrow 1$ to $l$ do
        if $\sum_{\substack{\psi=1\\ \psi\neq\rho}}^{m} \aleph_{\tau,\psi} \geq \sum_{\substack{\gamma=1\\ \gamma\neq\tau}}^{l} \aleph_{\gamma,\rho}$ then $\theta(\tau,\rho) \leftarrow 0$;
        else $\theta(\tau,\rho) \leftarrow 1$;
    $s \leftarrow \frac{\boldsymbol{\theta}(\iota,\cdot)\times\mathbf{h}(\iota,\cdot)^{\mathrm{T}}}{\boldsymbol{\theta}(\iota,\cdot)\times\mathbf{h}(\iota,\cdot)^{\mathrm{T}}+\left(\{1\}_l-\boldsymbol{\theta}(\cdot,\rho)\right)\times\mathbf{h}(\cdot,\rho)^{\mathrm{T}}}$;
    $s^* \leftarrow UniRand([0,1])$;
    if $s^* \leq s$ then output $\mathrm{VP}(\alpha_\iota)$;
    else output $\mathrm{VP}(\beta_\rho)$;
```

Second, we deliberate upon game-theoretical approaches that aim at improving some of these characteristics through non-cooperative interactions.

7.1 Definitions of Unlinkability

In common privacy context unlinkability is strongly related to anonymity. For example, Reference [65] states " ...Unlinkability is a sufficient condition of anonymity, but it is not a necessary condition," while [50] equals anonymity to RP+AP-U-unlinkability.

Some definitions of unlinkability are called "games" [50, 62, 64, 73]. Such selection of terminology seems unjustified: These constructs are barely related to game theory, since observability of the strategies played by the players as well as their payoffs remain unclear. To avoid confusion with game theory, we will further call them "tests." One common idea behind these tests is to define unlinkability as the result of an interaction between an attacker (whose goal is to distinguish between the actions of different agents) and a challenger. Depending on the variant of the test, either two or three agents are selected by the challenger. He then presents to an attacker results of the sessions with explicit assurance that either one or two among these agents were selected, respectively. The attacker then needs to guess the label of the entity, or the relation between the labels (e.g., "same" or "different"), respectively. Unlinkability is achieved if the attacker's performance is not statistically better than a random guess.

Models based on logical description for unlinkability have also been used by the community [13, 40, 61]. For example, in Reference [40] the authors propose a framework for reasoning about anonymity in particular. Their framework employs the modal logic of knowledge within the context of the runs and the systems framework but does not consider quantitative measurements of anonymity. The authors of Reference [13] abstract model based on epistemic logic, with natural and intuitive definitions in terms of the attacker's knowledge. In addition to unlinkability they also identified a dual notion of inseparability that may be of importance for some special cases on practice. However, majority of the authors in this category do not examine probabilistic descriptions of unlinkability. This sets forth impossibility of a single (or integral) measure for unlinkability. As a result, many realistic scenarios with complex usage patterns cannot be encompassed by these descriptions for the sake of further optimization, for instance.

References [7, 33, 73] quantify the linkability of items in the system using information-theoretic descriptions. For example, a basic information-theoretic notion for unlinkability is given in Reference [73] where the authors utilize Shannon entropy to measure unlinkability of elements within one set as well as between the sets. One limitation of this approach is that no context is considered by the authors. This was rectified in Reference [33] whereseven special cases of context information were considered. Nevertheless, this information is provided in the form of "hints" (known to the attacker) about target partitions and cannot be easily generalized for all kinds of context. The authors of Reference [7] measure unlinkability using mutual information between the actual communication that took place, and the information the adversary knows about it. In that way, their definition obviously considers context information in the most general way. However, mutual information may be heavily affected by priors, which is impractical for decisions making in non-cooperative environment where players may not communicate.

Finally, a number of papers survey criteria and measures that may be useful in a wide range of privacy applications [21, 26, 75]. For example, survey [75] spans across multiple privacy domains and can serve as a general framework for privacy measurements. In particular, the authors propose an extensive taxonomy of privacy metrics, which is classified by output and describes 17 entropy-based measures, to name a few.

In summary, one of the main limitations of the analyzed sources is the lack of attention to the problem of interchangeable usage of assertions. Some of the information-theoretic measures such as in Reference [75] are universal. However, possible application of these measures to the problem of interchangeable usage is not suggested by the authors. Existing definitions are therefore insufficient to optimize unlinkability in the environment where multiple assertions of a user can satisfy access policy P that is established by RP.

7.2 Game-theoretical Approaches

Game theory studies incentives and interactions between rational agents [35]. In some cases these interactions can be characterized by a stable state of Nash equilibrium where agents do not deviate from the strategies they play. These outcomes are also considered as the most likely and hence users should evaluate their anticipated privacy level with an equilibrium in mind.

A number of papers apply game theory to address privacy issues that occur in practice [34, 44, 59]. It should, however, be taken into account that their solutions are usually problem specific and cannot be easily applied across multiple domains. For example, problems of pseudonym change in mobile networks were investigated by the authors of References [34, 44]. In Reference [34], the authors elaborate on a user-centric location privacy model that takes into account the beliefs of users about the tracking power of the adversary, the degree of anonymity that users obtain in the mix zones as well as the cost and time of pseudonym change. Results from their study define an equilibrium where the strategies played by the users can be decided when their utilities are compared with a threshold value. In Reference [44] authors analyze a game where local adversary is equipped with multiple eavesdropping stations to track mobile users who deploy mix zones to protect their location privacy. The authors predict the strategies of both players and derive the strategies at equilibrium in complete and incomplete information scenarios that is quantified based on real road-traffic information.

However, there is a number of game-theoretic papers with less distinct contribution to practical aspects of the privacy. They nevertheless may be useful for coordination scenarios in attribute-based authentication [36, 54]. For instance, the authors of Reference [54] discuss a game with mediating mechanism that can improve the outcome of the game when compared to Bayes Nash Equilibrium. Authors demonstrate that any algorithm that computes a correlated equilibrium of a complete information game while satisfying a variant of differential privacy can be used as a recommended mechanism satisfying desired incentive properties.

To sum up, an obvious limitation of the existing game-theoretical solutions is the absence of models that adequately cover interchangeable usage of assertions. Properties of information-theoretical measures command that games with continuous strategies (and not mixed strategies!) must be analyzed in the presence of multiple alternatives for the players. This component is missing from game-theoretical applications for privacy. Also, majority of the sources gravitate toward the games where information sets can be provided to the users. As such they ignore cases of severe uncertainty. There are several limitations for this sole line of thoughts. First, a mechanism that provides information to the players (similar to *mediator* in “naïve game”) must be designed. Second, players must place trust on that mechanism.

8 DISCUSSION & CONCLUSION

In this article, we contribute toward a solution (evidence) for the problem of unlinkability in ABA. We demonstrate how GAPAS can address the problem through the interchangeable usage of different assertions possessed by every user. The differences between assertions are inevitable: They contain static elements (metadata) that the users cannot alter (see Table 3). The *academic novelty* of our article is supported by (a) unique set of assumptions for ABA, (b) new dilemma of interchangeable usage of assertions that stems from (a), and (c) new methodology to solve the dilemma. The *generalizability* of GAPAS follows from the generalizability of ABA Context, Assumptions & Threat. In defining these aspects, we cover a domain of situations in ABA that applies to federated IAM such as eIDAS: This is justified in Sections 1 and 2 [4, 30]. Hence, we make a meaningful contribution to both *theory* and *practice* of privacy assurance in ABA.

8.1 Theoretical Implications

We defined criterion C from which users derive their utilities in non-cooperative coordination games. This development was motivated by our analysis of existing criteria and measures for user unlinkability, during which we observed a substantial gap. For example, test for RP+AP-U unlinkability described in ISO/IEC 27551 (see Listing 1) does not specify *how* users select attributes for authentication if multiple attribute realizations (e.g., α_ι and β_ρ) are available [50].

In addition, it neither specifies *how* RP links users nor *what* is the estimate of the rate of successful linking. The proposed criterion C is based on the information-theoretical measure of conditional entropy. The main advantage of C is that it can be used to estimate the best linking performance of malicious RP (see Lemma 1). To explain the linking procedure, on Figure 3(a) we assumed that certain statistics about user decisions must be known to calculate $C = H(\mathsf{L} \mid l)$. However, the results of Lemma 1 do not depend on *whether* these statistics are known to RP and *how* he can use it: C is the upper estimate of linking ability.

Lemma 2 further demonstrates how personal expected utilities of the players can be derived from C as well as the assumptions needed for that. To calculate their optimal strategies (e.g., best responses) a player i who controls their personal assertions $\alpha^{(i)}$ and $\beta^{(i)}$ (e.g., whose type is $\mathbf{t}_i = (\alpha^{(i)}, \beta^{(i)})$) must know marginal probabilities at RP, $\Pr_{\mathcal{S}}(\alpha^{(i)})$, $\Pr_{\mathcal{S}}(\beta^{(i)})$ (see Figure 5). These marginal probabilities depend on statistical distribution $\aleph$ of types of other players in the system as well as their best responses. Inter-dependencies of personal best responses for multiple players explain why unlinkability in attribute-based authentication systems needs to be addressed using GAPAS. The way of *how* information about $\Pr_{\mathcal{S}}(\alpha^{(i)})$, $\Pr_{\mathcal{S}}(\beta^{(i)})$ is communicated to i is of paramount importance to the game and its equilibria. As such, this '*how to communicate information*' question is a stepping- stone to answer the **R**esearch **Q**uestion

"How should users use their assertions to maximize RP+AP-U unlinkability?"

GAPAS provides the answer to RQ. To accommodate likely scenarios, two types of non-cooperative coordination games with incomplete information were analyzed: (i) *naïve* game, where mediator M provides necessary information to the players, and (ii) *tenable* game, where no information about $\Pr_{\mathcal{S}}(\alpha^{(i)})$, $\Pr_{\mathcal{S}}(\beta^{(i)})$ is available. To the best of our knowledge, we are the first to apply game-theoretical approaches to the task of interchangeable attribute usage. For the naïve game, we applied principles of correlated equilibria [6, 54]. This allows us to achieve consistency of priors (e.g., information about $\Pr_{\mathcal{S}}(\alpha^{(i)})$, $\Pr_{\mathcal{S}}(\beta^{(i)})$).

In addition, multiple equilibria are possible in the system (see Equation (10)): By selecting this information, we maximize the overall unlinkability C for the system of n players. One of the limitations of this approach is that information provided by mediator M needs to be trusted by the players. This is avoided in the tenable game where we apply the Wald maximin principle to maximize the utility of each player i. That player calculates "worst-case scenario" values $\Pr_{\mathcal{S}}(\alpha^{(i)})$, $\Pr_{\mathcal{S}}(\beta^{(i)})$ based on $\aleph$ [71]. This worst-case scenario guarantees that the utility of i cannot be lower even if other $n-1$ players attempt to minimize the expected utility (over $\aleph$) of i whose type is unknown to them. Such approach is widely used in robust decision-making and is suitable for the situations that are characterized by severe uncertainty such as the situation when a player i does not know $\Pr_{\mathcal{S}}(\alpha^{(i)})$, $\Pr_{\mathcal{S}}(\beta^{(i)})$. The downside of this approach is that the utility of i is lower than in correlated equilibrium.

From Section 5, we observe that playing tenable game comes at the cost (see Figure 6) [71]. Also, there is a clear contrast between unlinkability when users are guided by different principles of assertion usage. Both naïve and tenable games belong to *rational* principle of assertion usage. In Section 5, we compare them with *alternative* principles of usage (see Figures 7 and 8). From the results, it is clear that rational principles of interchangeable usage where users *coordinate* have a substantial benefit over other alternative scenarios. This is because GAPAS optimizes impact on unlinkability (through best responses) produced by every individual user i.

We also gain insights into how the distribution $\aleph$ of attribute realizations (and, hence, user types $\mathbf{t}_i$) used by the users impacts their unlinkability. Indistinguishability may be insufficient for unlinkability. However, better indistinguishability (e.g., larger anonymity sets) can provide better unlinkability. This is because the task of coordination in GAPAS becomes easier if user assertions belong to larger anonymity sets. The size of anonymity sets can be calculated from $\aleph$, the number of users n in the system, and their decisions $\mathcal{S}$. For instance, it can be seen that for naïve and tenable games, expected unlinkability is lower toward the center of corresponding heatmaps on Figure 6(a) and (b). This is because that area represents more diverse distributions $\aleph$ (more equally sized anonymity sets), which further constrains the coordination effect. In contrast, the outer areas of these maps represent the cases when the majority of the players have the same type.

Last, our approach can be ***generalized*** to include cases when users possess more than 2 different assertions. For instance, if user i controls $\{\alpha^{(i)}, \beta^{(i)}, \gamma^{(i)}\}$, then her best response would require that $\frac{\Pr(\alpha^{(i)}|i)}{\Pr_{\mathcal{S}}(\alpha^{(i)})} = \frac{\Pr(\beta^{(i)}|i)}{\Pr_{\mathcal{S}}(\beta^{(i)})} = \frac{\Pr(\gamma^{(i)}|i)}{\Pr_{\mathcal{S}}(\gamma^{(i)})}$.

8.2 Implications for Practice

First, our study contributes to privacy assurance: Our argument is based on information theory, and game theory [46, 48]. This argument is needed in ABA, because existing best practices and standards such as ISO 27551 do not stipulate how evidence can be supplied. For instance, it is unclear *how* users need to use their assertions (if a user has more than two of them) during the test nor *how* RP makes his *"guess."* Criterion C coupled with naïve or tenable games can now be used to address this limitation. This is because rational decision-making unambiguously specifies the actions of the users the resulting unlinkability can now be rated. A certain threshold T_C can be used to decide whether an attribute-based authentication system conforms to requirements or not.

Second, recommendations can be provided based on GAPAS for (a) AP to issue credentials and (b) RP to set access policies. As we have observed in Section 6 metadata can become the reason for the distinguishability of assertions. Therefore, any unnecessary detalization or redundancy should be avoided by AP who issues credentials. For example, content of the properties `@context`, `type` within VC should be kept as minimal as possible. Across all APs, it is better to use the same `credentialSchema`. If it is not critical for LoA of credentials or internal protocols and procedures of AP, then time must not appear in `issuanceDate` property, e.g., it should always be in the format `YYYY-MM-DD`. This would ensure that a larger number of users have identical metadata. The role of RP is also important: he defines access policies P, which also affects user types $\mathbf{t}_i$ in an attribute-based authentication system. For example, if for the use case in Section 6, then RP only allows assertions from AP VicRoads (and does not allow assertions from AP Deakin University) the distribution of user types in the system will become $\aleph = (0.122 \;\; 0.878)$. This will affect decisions that the users can make as well as equilibria and resulting unlinkability C. Hence, GAPAS can be used to recommend (or "benchmark") policies. Privacy-respecting RPs can use this benchmarking to the greater benefit of the users. However, malicious RPs may deliberately constrain access policies with the aim to make linking easier (e.g., to reduce unlinkability). Therefore, additional mechanisms that may incentivize or penalize RP for such malicious behaviour are yet to be further studied [57].

Finally, GAPAS can assist firms in developing various identity agents and DCWs. For example, software implementation of Algorithm 1 would not require changes in authentication protocols and procedures such as OAuth 2.0 and OIDC. In addition, communication and computation overheads associated with GAPAS are minimal. To make a decision in a naïve game, a user's DCW that incorporates GAPAS would require information set from mediator M. Such set is provided once only, and its size equals the number of different realizations in the system, e.g., $|\mathcal{A}|+|\mathcal{B}|$. A tenable game does not require communication. Also, only naïve game requires that equilibrium is computed by M. This needs to be done once only: Obtained information set (e.g., $\Pr(\alpha^{(i)})$ and $\Pr(\beta^{(i)})$) is communicated to every user i. To calculate that set M solves Equation (10). This can be done by the trust-region algorithm [22]. In both naïve and tenable variants of the game, each player produces the best response, which is a single operation (see Corollary 1). However, the tenable game calculates the "worst-case scenario" (see Equation (16) and Algorithm 1). For every user, the total complexity of that procedure is $x(l+m)$, where x is the number of different realizations controlled by a user, $l \times m$ is the dimension of matrix $\aleph$. Each player produces the best response only once in both naïve and tenable games; the decision is then applied across all authentication sessions.

8.3 Limitations & Future Studies

We acknowledge there are certain limitations to our study, which we aim to address through future research. First, there is a need for further inquiry about the settings in the user-centric approach adopted in this article. Second, there is a need to explore the possibilities of non-user-centric approaches to counter the threat of RP and AP collusion in ABA.

In a user-centric approach, mediator M is an essential prerequisite for the *naïve* game, which provides better unlinkability compared to *tenable* game. The construction of an efficient and privacy-preserving M is, however, beyond the scope of our article. This task may be associated with additional privacy challenges. Specifically, some sources suggest that to prevent players from learning (much) more about any player's type outside of the coalition than they could have learned in the original (non-coordinated) settings, the designed proxy M should satisfy differential privacy conditions [54]. Also, questions associated with the cost of developing and maintenance of M are yet to be explored. The immaturity of the latter question is, nevertheless, compensated in our article: We analyze *tenable* game, which does not require further studies.

Exploration of non-user-centric approaches may also bring fruitful results. We study unlinkability in ABA under a specific threat: RP and AP collude against users. Despite being considered by ISO/IEC 27551 as one of the major threats in ABA, the threat is yet to be better understood. In particular, "Asset, Threat and Vulnerability" risk identification method may be used for this purpose [47]. This may require an understanding of asset values (e.g., valuations for user privacy) as well as the likelihood of occurrence of such risk. The values in risk assessment, for instance, can be reduced by reducing the likelihood of risk occurrence. This can be done through disincentivizing RP and AP to collude with each other, which contrast with our user-centric approach to the problem in this article. For example, game-theoretical approaches of mechanism design can be used to incentivize AP and RP to betray (e.g., to report to the Authorities) each other should sufficient evidence be accumulated by these parties [5, 28].

APPENDICES

A PROOFS

Lemma 1. *Best linking performance decreases with $H(\mathsf{L} \mid l)$ (for details see Appendix A).*

Proof. To link authentication sessions RP labels them with $\mathsf{L}' \in \mathcal{L}'$, where $\mathcal{L}' = \{A', B'\}$. We divide the proof into two parts: (i) We demonstrate that for the best linking performance RP aims to minimize $H(\mathsf{L} \mid \mathsf{L}')$ and (ii) $H(\mathsf{L} \mid \mathsf{L}') \geq H(\mathsf{L} \mid l)$.

(i) We express linking performance $\mathfrak{P}$ of RP as the difference between TPR and FPR: $\mathfrak{P} = \Pr(A' \mid A) - \Pr(A' \mid B) = \frac{\Pr(A', A)}{\Pr(A)} - \frac{\Pr(A', B)}{\Pr(B)}$, which is to be maximized and for which we demand that $\mathfrak{P} \geq 0$. In authentication systems, probability of A, i.e., $\Pr(A)$ and probability of B, i.e., $\Pr(B)$ are decided by the users and hence cannot be affected by RP. We further demonstrate that either increase of the probability that both events A' and A occur, i.e., $\Pr(A', A)$ or decrease of the probability that both events A' and B occur, i.e., $\Pr(A', B)$ reduces $H(\mathsf{L} \mid \mathsf{L}')$. We note that conditional entropy

$$H(\mathsf{L} \mid \mathsf{L}') = \sum_{\mathsf{L} \in \mathcal{L}} \sum_{\mathsf{L}' \in \mathcal{L}'} \Pr(\mathsf{L}, \mathsf{L}') \log \frac{\Pr(\mathsf{L}')}{\Pr(\mathsf{L}, \mathsf{L}')}$$

is unimodal on $\Pr(A', A)$ (something similar must be stated about $\Pr(A', B)$) by analyzing its first derivative $\partial \frac{H(\mathsf{L}|\mathsf{L}')}{\partial \Pr(A, A')} = \log(\Pr(A, A') + \Pr(B, A')) + \log \Pr(A, B') - \log \Pr(A, A') - \log(\Pr(A, B') + \Pr(B, B'))$ and finding its unique extremum at $\frac{\Pr(A, A')}{\Pr(A, A') + \Pr(A, B')} = \frac{\Pr(B, A')}{\Pr(B, A') + \Pr(B, B')}$. The denominators in the latter equation are equal to $\Pr(A)$ and $\Pr(B)$, respectively. As a result, $\mathfrak{P} = 0$ at this extremum, and, due to unimodality of $H(\mathsf{L} \mid \mathsf{L}')$ on $\Pr(A', A)$ (and on $\Pr(A', B)$), maximization of $\mathfrak{P}$ requires minimization of $H(\mathsf{L} \mid \mathsf{L}')$.

(ii) For any deterministic linking algorithm $c : \ell \rightarrow \mathcal{L}'$ it is true that $H(\mathsf{L}' \mid l) = 0$, and hence, $H(\mathsf{L}', l) = H(l)$. Next, according to the properties of joint entropy, $H(\mathsf{L}, \mathsf{L}', l) \geq H(\mathsf{L}, l)$ from which follows that $H(\mathsf{L} \mid \mathsf{L}', l) \geq H(\mathsf{L} \mid l)$. According to the conditional entropy properties, we also have $H(\mathsf{L} \mid \mathsf{L}') \geq H(\mathsf{L} \mid \mathsf{L}', l)$, which finally implies $H(\mathsf{L} \mid \mathsf{L}') \geq H(\mathsf{L} \mid l)$. □

Lemma 2. *Expected utility for player i is (for details see Appendix A)*

$$\mathbb{E}\left[u_i\right] \approx -s_i \log \frac{s_i}{\mathbb{E}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big]} - \Big(1 - s_i\Big) \log \frac{1 - s_i}{\mathbb{E}\Big[\Pr_{\mathcal{S}}\Big(\beta^{(i)}\Big)\Big]}. \tag{3}$$

Proof. According to Equation (2), the expected value of unlinkability for the entire system is $\mathbb{E}\left[C\right] = \log n + \frac{1}{n} \sum_i^n \mathbb{E}\left[u_i\right]$ where

$$\mathbb{E}\left[u_i\right] = -\mathbb{E}\left[s_i \log \frac{s_i}{\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)}\right] - \mathbb{E}\left[(1 - s_i) \log \frac{1 - s_i}{\Pr_{\mathcal{S}}(\beta^{(i)})}\right].$$

We further process the first term of the right-hand side of the latter equation:

$$\begin{aligned} \mathbb{E}\left[s_i \log \tfrac{s_i}{\Pr_{\mathcal{S}}(\alpha^{(i)})}\right] &= \mathbb{E}\left[s_i \log s_i - s_i \log \Pr_{\mathcal{S}}(\alpha^{(i)})\right] \\ &= s_i \log s_i - s_i \mathbb{E}\Big[\log \Pr_{\mathcal{S}}(\alpha^{(i)})\Big]. \end{aligned}$$

Taylor expansion of $\log \Pr_{\mathcal{S}}(\alpha^{(i)})$ around $x_0 = \mathbb{E}[\Pr_{\mathcal{S}}(\alpha^{(i)})]$ produces

$$\begin{aligned} \log \Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big) \ \approx & \quad \log \mathbb{E}\Big[\Pr_{\mathcal{S}}(\alpha^{(i)})\Big] \\ + \frac{\Pr_{\mathcal{S}}(\alpha^{(i)}) - \mathbb{E}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big]}{\mathbb{E}\Big[\Pr_{\mathcal{S}}(\alpha^{(i)})\Big]} - & \quad \frac{\Big(\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big) - \mathbb{E}\Big[\Pr_{\mathcal{S}}(\alpha^{(i)})\Big]\Big)^2}{2\mathbb{E}\Big[\Pr_{\mathcal{S}}(\alpha^{(i)})\Big]^2}. \end{aligned}$$

By taking expectation over the right-hand side of the equation we arrive at

$$\mathbb{E}\Big[\log \Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big] \approx \log \mathbb{E}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big] - \frac{\mathrm{Var}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big]}{2\mathbb{E}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big]^2},$$

and we obtain that

$$\mathbb{E}\left[s_i \log \frac{s_i}{\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)}\right] \approx s_i \log \frac{s_i}{\mathbb{E}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big]} + s_i \frac{\mathrm{Var}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big]}{2\mathbb{E}\Big[\Pr_{\mathcal{S}}\Big(\alpha^{(i)}\Big)\Big]^2}.$$

Similarly, the following equation also holds:

$$\begin{aligned} \mathbb{E}\left[(1 - s_i) \log \tfrac{1 - s_i}{\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)}\right] \approx & \ (1 - s_i) \log \tfrac{1 - s_i}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]} \\ & + (1 - s_i) \tfrac{\mathrm{Var}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]}{2\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]^2}. \end{aligned}$$

Last, by considering Assumption 7 we obtain that $\mathbb{E}\left[u_i\right]$ equals

$$-\mathbb{E}\left[s_i \log \frac{s_i}{\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)}\right] - \mathbb{E}\left[(1 - s_i) \log \frac{1 - s_i}{\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)}\right] - \frac{\mathrm{const}}{2},$$

and, due to indifference of the constant term to the actions of i, we exclude it from further considerations and, hence, demonstrate the validity of Equation (3). □

B LISTINGS FOR VC/VP

```
{
"@context": [
    "https://www.w3.org/2018/credentials/v1",
    "https://www.schema.org/",
    "https://www.deakinschema.org/contexts/studentcerts"],
"type": "VerifiablePresentation",
    "VerifiableCredential": [{
        "@context": [
            "https://www.w3.org/2018/credentials/v1",
            "https://www.schema.org/",
            "https://www.deakinschema.org/contexts/studentcerts"],
        "type": ["VerifiableCredential", "UndergradCredential"],
        "credentialSchema": {
            "id": "https://example.org/examples/degree.zkp",
            "type": "ZkpExampleSchema2018"},
        "issuer": "did:web:vc.deakin.world",
        "issuanceDate": "2021-03-10",
        "credentialSubject": {
            "ageOver": 18,},
        "proof": {
            "type": "AnonCredDerivedCredentialv1",
            "primaryProof": "ps9wLNSi48K5qNyAVMwdYqVHSMv1Ur8i ... Hf2ZvWF6zGvcSAsym2sgSk737
                ",
            "nonRevocationProof": "ce6fg24MfJPU1HvSXsf3ybzKARib4WxG ...
                VTce53M6UwQCxYshCuS3d2h"}}],
"proof": {
    "type": "AnonCredPresentationProofv1",
    "proofValue": "JkYdYMUYHURJLD7xdnWRinqWCEY5u5hG ... k115Lt3hMzLHoPiPQ9sSVfRrs1P"}
}
```

Listing 2. *Alice*'s VP from Deakin University.

```
{
"@context": [
    "https://www.w3.org/2018/credentials/v1",
    "https://www.schema.org/",
    "https://www.vicroads.org/schema/drlic"],
"type": "VerifiablePresentation",
    "VerifiableCredential": [{
        "@context": [
            "https://www.w3.org/2018/credentials/v1",
            "https://www.schema.org/",
            "https://www.vicroads.org/schema/drlic"],
        "type": ["VerifiableCredential", "RestrDrLic"],
        "credentialSchema": {
            "id": "https://example.org/examples/license.zkp",
            "type": "ZkpExampleSchema2018"},
        "issuer": "did:web:vc.vicroads.world",
        "issuanceDate": "2021-02-10",
        "expirationDate": "2022-02-10",
        "credentialSubject": {
            "ageOver": 18},
        "proof": {
            "type": "AnonCredDerivedCredentialv1",
            "primaryProof": "Ox8iTNSi48K5iHyAVMwdYqVHSMv1Uu1p ... Uf2ZvWF6zGdpSAsym2sgSk35Q
                ",
            "nonRevocationProof":     "bI6fg24MfJPU1pZSXsf3ybzKARib4WPc ...
                OJce53M6UwgFxYshCuS3dt3"}}],
"proof": {
    "type": "AnonCredPresentationProofv1",
    "proofValue": "aLYdYMUYHUpvLD7xdnWRinqWCEY5u9hW ... Lk5Lt3hMzLHoFIPQ9sSVfRrsOz"}
}
```

Listing 3. *Alice*'s VP from VicRoads.

REFERENCES

[1] Gergely Alpár, Fabian van den Broek, Brinda Hampiholi, Bart Jacobs, Wouter Lueks, and Sietse Ringers. 2017. IRMA: Practical, decentralized and privacy-friendly identity management using smartphones. In *Proceedings of the Hot Topics in Privacy Enhancing Technologies (HotPETs'17).*

[2] Sebastian Angel and Srinath Setty. 2016. Unobservable communication over fully untrusted infrastructure. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16).* USENIX Association, 551–569. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/angel.

[3] A. Beduschi, J. Cinnamon, J. Langford, C. Luo, and D. Owen. 2017. Building digital identities: The challenges, risks and opportunities of collecting behavioural attributes for new digital identity systems. 40 pages.

[4] Diana Berbecaru and Cesare Cameroni. 2020. ATEMA: An attribute enablement module for attribute retrieval and transfer through the eIDAS Network. In *Proceedings of the 24th International Conference on System Theory, Control and Computing (ICSTCC'20).* IEEE, 532–539. https://doi.org/10.1109/ICSTCC50638.2020.9259642

[5] Joyce Berg, John Dickhaut, and Kevin McCabe. 1995. Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 1 (1995), 122–142. https://doi.org/10.1006/game.1995.1027

[6] Dirk Bergemann and Stephen Morris. 2016. Bayes correlated equilibrium and the comparison of information structures in games. *Theor. Econ.* 11, 2 (2016), 487–522. https://doi.org/10.3982/TE1808

[7] Ron Berman, Amos Fiat, and Amnon Ta-Shma. 2004. Provable unlinkability against traffic analysis. In *Proceedings of the International Conference on Financial Cryptography.* Springer, 266–280.

[8] T. Berners-Lee, R. Fielding, and L. Masinter. 2005. Uniform resource identifier (URI): Generic syntax. Retrieved from https://tools.ietf.org/html/rfc3986.

[9] E. Birrell and F. B. Schneider. 2013. Federated identity management systems: A privacy-based characterization. *IEEE Secur. Priv.* 11, 5 (2013), 36–48.

[10] Kaitlin R. Boeckl and Naomi B. Lefkovitz. 2020. *NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management, Version 1.0.* Special Publication. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.01162020

[11] Dan Boneh, Xavier Boyen, and Hovav Shacham. 2004. Short group signatures. In *Proceedings of the 24th Annual International Cryptology Conference (CRYPTO'04)*, Matt Franklin (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 41–55.

[12] E. J. Borowski and J. M. Borwein. 2002. *Collins Dictionary of Mathematics.* HarperCollins.

[13] Mayla Brusó, Konstantinos Chatzikokolakis, Sandro Etalle, and Jerry Den Hartog. 2012. Linking unlinkability. In *Proceedings of the International Symposium on Trustworthy Global Computing*, Catuscia Palamidessi and Mark D. Ryan (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 129–144.

[14] J. Camenisch, S. Krenn, A. Lehmann G. L. Mikkelsen, G. Neven, and M. Ø. Pedersen. 2014. D3. 1: Scientific comparison of ABC protocols. Part I-Formal Treatment of Privacy-Enhancing Credential Systems. Project deliverable in ABC4Trust (2014).

[15] Jan Camenisch and Anna Lysyanskaya. 2002. A signature scheme with efficient protocols. In *Proceedings of the 3rd International Conference on Security in Communication Networks (SCN'02), Revised Papers* , Lecture Notes in Computer Science, Vol. 2576. Springer, 268–289.

[16] Tom Carter. 2007. An introduction to information theory and entropy. Complex Systems Summer School, Santa Fe.

[17] Sam Castle, Fahad Pervaiz, Galen Weld, Franziska Roesner, and Richard Anderson. 2016. Let's talk money: Evaluating the security challenges of mobile money in the developing world. In *Proceedings of the 7th Annual Symposium on Computing for Development.* Association for Computing Machinery, New York, NY, Article 4, 10 pages. https://doi.org/10.1145/3001913.3001919

[18] David W. Chadwick, Romain Laborde, Arnaud Oglaza, Remi Venant, Samer Wazan, and Manreet Nijjar. 2019. Improved identity management with verifiable credentials and FIDO. *IEEE Commun. Stand. Mag.* 3, 4 (2019), 14–20.

[19] Mike Clark. 2019. German government adds iPhone NFC identity card reading to digital ID app. Retrieved from https://www.nfcw.com/2019/10/01/364573/german-government-adds-iphone-nfc-identity-card-reading-to-digital-id-app/.

[20] Sarah Clark. 2020. Germany to begin rollout of open national digital identity service "later this year". Retrieved from https://www.nfcw.com/2020/07/29/367360/germany-to-begin-rollout-of-open-national-digital-identity-service-later-this-year/.

[21] Sebastian Clauß and Stefan Schiffner. 2006. Structuring anonymity metrics. In *Proceedings of the 2006 Workshop on Digital Identity Management, Alexandria, VA, USA, November 3, 2006*, Ari Juels, Marianne Winslett, and Atsuhiro Goto (Eds.). ACM, 55–62. https://doi.org/10.1145/1179529.1179539

[22] Thomas F. Coleman and Yuying Li. 1996. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.* 6, 2 (1996), 418–445.

[23] European Commission. 2018. Looking ahead: The user experience of eIDAS-based eID. Value Proposition of eIDAS eID.

[24] Alissa Cooper, Hannes Tschofenig, Bernard D. Aboba, Jon Peterson, John Morris, Marit Hansen, and Rhys Smith. 2013. *Privacy Considerations for Internet Protocols*. Request for Comments IETF RFC 6973. The Internet Engineering Task Force, Wilmington, DE. https://doi.org/10.17487/RFC6973

[25] Matthew Davie, Dan Gisolfi, Daniel Hardman, John Jordan, Darrell O'Donnell, and Drummond Reed. 2019. The trust over ip stack. *IEEE Commun. Stand. Mag.* 3, 4 (2019), 46–51.

[26] Claudia Diaz, Stefaan Seys, Joris Claessens, and Bart Preneel. 2002. Towards measuring anonymity. In *Proceedings of the International Workshop on Privacy Enhancing Technologies* Roger Dingledine and Paul Syverson (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 54–68.

[27] Dizme. 2021. The Key to Digital Identity. Retrieved from https://www.dizme.io/.

[28] Jim Engle-Warnick and Robert L. Slonim. 2004. The evolution of strategies in a repeated trust game. *J. Econ. Behav. Organiz.* 55, 4 (2004), 553–573. https://doi.org/10.1016/j.jebo.2003.11.008 Trust and Trustworthiness.

[29] European Commission. 2020-07-23. Proposal for a European Digital Identity (EUid) and Revision of the eIDAS Regulation. Directorate-General for Communications Networks, Content and Technology (2020-07-23).

[30] European Parliament. 2014-07-23. Regulation (EU) No 910/2014 of the european parliament and of the council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC. Council of the European Union (2014-07-23).

[31] Uriel Feige, Amos Fiat, and Adi Shamir. 1987. Zero Knowledge Proofs of Identity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York*, New York, USA, Alfred V. Aho (Ed.). ACM, 210–217. https://doi.org/10.1145/28395.28419

[32] Joint Task Force. 2020. *Security and Privacy Controls for Information Systems and Organizations*. Special Publication NIST SP 800-53 rev.5. National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.800-53r5

[33] Matthias Franz, Bernd Meyer, and Andreas Pashalidis. 2007. Attacking unlinkability: The importance of context. In *Proceedings of the International Workshop on Privacy Enhancing Technologies*. Springer, 1–16.

[34] Julien Freudiger, Mohammad Hossein Manshaei, Jean-Pierre Hubaux, and David C Parkes. 2009. On non-cooperative location privacy: A game-theoretic analysis. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*. 324–337.

[35] Drew Fudenberg and Jean Tirole. 1991. *Game Theory* (11 ed.). The MIT Press.

[36] Arpita Ghosh and Katrina Ligett. 2013. Privacy as a coordination game. In *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton'13)*. IEEE, 1608–1615.

[37] Sérgio Manuel Nóbrega Gonçalves, Alessandro Tomasi, Andrea Bisegna, Giulio Pellizzari, and Silvio Ranise. 2020. Verifiable Contracting. In *Computer Security*, Ioana Boureanu, Constantin Cătălin Drăgan, Mark Manulis, Thanassis Giannetsos, Christoforos Dadoyan, Panagiotis Gouvas, Roger A. Hallman, Shujun Li, Victor Chang, Frank Pallas, Jörg Pohle, and Angela Sasse (Eds.). Springer International Publishing, Cham, 133–144.

[38] Paul A. Grassi, Michael E. Garcia, and James L. Fenton. 2020. *Digital Identity Guidelines*. Standard NIST SP 800-63-3. National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.800-63-3

[39] Paul A. Grassi, Naomi B. Lefkovitz, Ellen M. Nadeau, Ryan J. Galluzzo, and Abhiraj T. Dinh. 2018. *Attribute Metadata: A Proposed Schema for Evaluating Federated Attributes*. Technical Report NISTIR 8112. National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.IR.8112

[40] Joseph Y. Halpern and Kevin R. O'Neill. 2005. Anonymity and information hiding in multiagent systems. *J. Comput. Secur.* 13, 3 (2005), 483–514.

[41] John C. Harsanyi. 1967. Games with incomplete information played by "Bayesian" players, I–III Part I. the basic model. *Manage. Sci.* 14, 3 (1967), 159–182. https://doi.org/10.1287/mnsc.14.3.159

[42] Nicholas Hopper, Eugene Y. Vasserman, and Eric Chan-TIN. 2010. How much anonymity does network latency leak? *ACM Trans. Inf. Syst. Secur.* 13, 2, Article 13 (March 2010), 28 pages. https://doi.org/10.1145/1698750.1698753

[43] Vincent C. Hu, David F. Ferraiolo, and D. Richard Kuhn. 2019. *Attribute Considerations for Access Control Systems*. Recommendation NIST SP 800-205. National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.800-205

[44] Mathias Humbert, Mohammad Hossein Manshaei, Julien Freudiger, and Jean-Pierre Hubaux. 2010. Tracking games in mobile networks. In *Proceedings of the International Conference on Decision and Game Theory for Security*. Springer, 38–57.

[45] IDunion. 2021. An open ecosystem for trusted identities. Retrieved from https://idunion.org/?lang=en.

[46] ISO Central Secretary. 2012. *Information Technology–Security Techniques–Security Assurance Framework–Part 1: Introduction and Concepts*. Technical Report ISO/IEC TR 15443-1:2012(E). International Organization for Standardization, Geneva, CH.

[47] ISO Central Secretary. 2018. *Information Technology–Security Techniques–Information Security Risk Management*. Standard ISO/IEC 27005:2018(E). International Organization for Standardization, Geneva, CH.

[48] ISO Central Secretary. 2019. *Systems and Software Engineering–Systems and Software Assurance–Part 1: Concepts and Vocabulary*. Standard ISO/IEC/IEEE 15026-1:2019(E). International Organization for Standardization, Geneva, CH.

[49] ISO Central Secretary. 2020. *Information Security, Cybersecurity and Privacy Protection–Evaluation Criteria for IT Security–Part 2: Security Functional Components*. Standard ISO/IEC DIS 15408-2:2020(E). International Organization for Standardization, Geneva, CH.

[50] ISO Central Secretary. 2020. *Information Technology–Requirements for Attribute-based Unlinkable Entity Authentication*. Standard ISO/IEC DIS 27551. International Organization for Standardization, Geneva, CH.

[51] Telecommunication Standardization Sector of ITU. 2017. *ITU-T Focus Group Digital Financial Services: Main Recommendations*. Standard. International Telecommunication Union, Geneva, CH.

[52] M. Jones. 2015. JSON web key (JWK). Retrieved from https://tools.ietf.org/html/rfc7517.

[53] Michael Jones and Joe Hildebrand. 2015. *JSON Web Encryption (JWE)*. Request for Comments IETF RFC 7516. The Internet Engineering Task Force, Wilmington, DE. https://doi.org/10.17487/RFC7516

[54] Michael Kearns, Mallesh Pai, Aaron Roth, and Jonathan Ullman. 2014. Mechanism design in large games: Incentives and privacy. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*. 403–410.

[55] Gregg Kellogg, Pierre-Antoine Champin, and Dave Longley. 2020. *JSON-LD 1.1: A JSON-based Serialization for Linked Data*. Recommendation. World Wide Web Consortium.

[56] G. Klyne and C. Newman. 2002. Date and time on the Internet: Timestamps. Retrieved from https://tools.ietf.org/html/rfc3339.

[57] Michael Kubach, Heiko Roßnagel, and Rachelle Sellung. 2013. Service providers' requirements for eID solutions: Empirical evidence from the leisure sector. *Open Identity Summit* 2013 (2013).

[58] Loic Lesavre, Priam Varin, Peter Mell, Michael Davidson, and James Shook. 2019. A taxonomic approach to understanding emerging blockchain identity management systems. CoRR abs/1908.00929 (2019). arXiv:1908.00929 http://arxiv.org/abs/1908.00929

[59] Xinxin Liu, Kaikai Liu, Linke Guo, Xiaolin Li, and Yuguang Fang. 2013. A game-theoretic approach for achieving k-anonymity in location based services. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM'13)*. IEEE, 2985–2993. https://doi.org/10.1109/INFCOM.2013.6567110

[60] Dave Longley and Manu Sporny. 2020. *RSA Signature Suite 2018*. Specification. World Wide Web Consortium.

[61] Kiraku Minami. 2020. Trace equivalence and epistemic logic to express security properties. In *Proceedings of the International Conference on Formal Techniques for Distributed Objects, Components, and Systems*. Springer, 115–132.

[62] K. Nohl and D. Evans. 2009. Privacy through noise: A design space for private identification. In *Proceedings of the Annual Computer Security Applications Conference*. 518–527.

[63] Nate Otto, Sunny Lee, Brian Sletten, Daniel Burnett, Manu Sporny, and Ken Ebert. 2019. *Verifiable Credentials Use Cases*. Guide. World Wide Web Consortium.

[64] Khaled Ouafi and Raphael C.-W. Phan. 2008. Privacy of recent RFID authentication protocols. In *Proceedings of the International Conference on Information Security Practice and Experience*. Springer, 263–277.

[65] Andreas Pfitzmann and Marit Hansen. 2010. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management.

[66] Walter Priesnitz Filho, Carlos Ribeiro, and Thomas Zefferer. 2019. Privacy-preserving attribute aggregation in eID federations. *Fut. Gener. Comput. Syst.* 92 (2019), 1–16. https://doi.org/10.1016/j.future.2018.09.025

[67] Pavel Pudlák. 2013. Proofs of impossibility. In *Logical Foundations of Mathematics and Computational Complexity*. Springer, 255–364.

[68] Mikerah Quintyne-Collins, Heather Vescent, Darrell O'Donnell, Greg Slepak, Michael Brown, Christoper Allen, and Michael Ruther. [n. d.]. Digital credential wallets.

[69] Drummond Reed, Manu Sporny, Dave Longley, Christopher Allen, Ryan Grant, and Markus Sabadello. 2021. *Decentralized Identifiers (DIDs) v1.0: Core Architecture, Data Model, and Representations*. Recommendation. World Wide Web Consortium.

[70] Daniel Servos and Sylvia L. Osborn. 2017. Current research and open problems in attribute-based access control. *ACM Comput. Surv.* 49, 4 (2017), 1–45.

[71] Moshe Sniedovich. 2016. Wald's mighty maximin: A tutorial. *Int. Trans. Operat. Res.* 23, 4 (2016), 625–653. https://doi.org/10.1111/itor.12248

[72] Manu Sporny, Noble Grant, Dave Longley, Daniel Burnett, and Brent Zundel. 2019. *Verifiable Credentials Data Model v1.0: Expressing Verifiable Information on the Web*. Recommendation. World Wide Web Consortium.

[73] Sandra Steinbrecher and Stefan Köpsell. 2003. Modelling Unlinkability. In *Privacy Enhancing Technologies, Third International Workshop, PET 2003, Dresden, Germany, March 26-28, 2003, Revised Papers (Lecture Notes in Computer Science, Vol. 2760)*, Roger Dingledine (Ed.). Springer, 32–47. https://doi.org/10.1007/978-3-540-40956-4_3

[74] Kalman C. Toth, Ann Cavoukian, and Alan Anderson-Priddy. 2020. Privacy by design architecture composed of identity agents decentralizing control over digital identity. In *Proceedings of the Open Identity Summit*, Heiko Roßnagel, Christian H. Schunck, Sebastian Mödersheim, and Detlef Hühnlein (Eds.). Gesellschaft für Informatik e.V., Bonn, 163–170. https://doi.org/10.18420/ois2020_14

[75] Isabel Wagner and David Eckhoff. 2018. Technical privacy metrics: A systematic survey. *ACM Comput. Surv.* 51, 3 (2018), 1–38.

Received March 2021; revised August 2021; accepted November 2021