2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) | 978-1-6654-1658-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/TRUSTCOM53373.2021.00055

Enhancing Privacy Through DMMA: Decision-Making Method for Authentication

Maksym Slavnenko[1], Yevhen Zolotavkin[1,2], Jay Jeong[1,2], Veronika Kuchta[3] and Robin Doss[1,2]
[1]Strategic Centre for Cyber Security Research & Innovation (CSRI)
Deakin University, Geelong, Australia
[2] Cyber Security Cooperative Research Centre (CSCRC), Perth, Australia
[3]School of Mathematics and Physics, The University of Queensland, St Lucia, Australia
{m.slavnenko, first.last,} @deakin.edu.au, v.kuchta@uq.edu.au

***Abstract*—Attribute-Based Authentication (ABA) is becoming more prevalent in everyday interactions. In this paper, we propose the Decision-Making Method for Authentication (DMMA) to address the privacy concerns in ABA. The need for DMMA is supported through multiple observations. First, in practice, the indistinguishability of crypto-proof-based assertions (that are posessed by different users) fails with non-zero probability. This explains why cryptographic means alone are insufficient to provide a substantial level of unlinkability in ABA systems with n users. Second, each user in ABA possesses multiple credentials: they can be used interchangeably to get access to the service(s) which is provided by a relying party (RP). DMMA addresses the challenge of interchangeable usage. As an initial step, we synthesized the criterion of unlinkability: it is based on the definitions of international standard ISO 27551 as well as the information theoretic measure of conditional entropy. We then use that criterion to formalize the task of authentication as a non-cooperative coordination game. In this game, players (targets of the attack) maximize their utilities by using their assertions *interchangeably*. The experiment demonstrates that a number of equilibria with substantially higher unlinkability can be achieved. Unlinkability vary depending on: i) the information (and its trustworthiness) about the moves of the other players in the game; ii) the statistical distribution of user attributes. DMMA demonstrates how users may be provided recommendations over the optimal selection of assertions for ABA. These recommendations can have a practical impact if DMMA is implemented as a feature within Digital Credential Wallets (DCWs).**

***Index Terms*—Unlinkability, Decision making, Game theory, Verifiable Credentials, Authentication.**

I. Introduction

Verifiable digital credentials play a vital role in everyday transactions [1], [2]. They include digital certificates, tokens, signed documents and personal credentials. One of the primary benefits of such credentials is their suitability for user-centric Identity and Access Management (IAM): once being cryptographically signed by a trusted issuer, a user can manage them. The latter can derive assertions (realizations) from these credentials and submit them to a Relying Party (RP) without any assistance from an Identity Provider (IdP). Attribute-based authentication (ABA) is a special case of user-centric IAM where the claims within credentials consider various user attributes, e.g. age and name. ABA allows preserving privacy better if access policies set by an RP are known to a user. For instance, the owner can selectively disclose information that is necessary for an RP and hide valuable Personally Identifiable Information (PII). This obfuscation often contrasts with the position of an RP, who may wish to obtain additional information (e.g. in the form of metadata) with the aim to increase the trustworthiness of the assertions. Such demand for detail from an RP makes user anonymity hardly achievable in practice.

A. Current privacy trends in ABA

A number of prominent cryptographic solutions have been proposed in the past including Idemix, U-Prove, Privacy-preserving Attribute-Based Credentials [3]–[5]. Unfortunately, these technical solutions have not become mainstream due to reasons such as the substantial computational complexity and reliance on a trusted third party [6].

In order to mitigate some of these issues, there has been significant effort in developing the latest standards and specifications for verifiable digital assertions such as IRMA and Verifiable Credentials [2], [7]. They allow users to perform selective disclosure of the information related to the subject of the assertion (e.g. claim). This can be done using zero knowledge proofs (ZKP) as an approach that proves a secret without revealing it [8]. However, these techniques do not permit users to control all the components of digital assertions, which may cause situations where users are still discriminated by an RP based on the differences in *assertion metadata*, for instance [9], [10]. This implies that unlinkability in ABA may be impeded because absolute indistinguishability of user assertions is unattainable. To improve user privacy under such limitation, we consider an alternative game-theoretical approach in this paper.

B. Scope of the paper and contribution

In this paper, we aim to enhance user privacy in ABA through Decision Making Method for Authentication (DMMA). We demonstrate the necessary **evidence** based on:

(a) **context** of ABA system with the *threat model* where Attribute Provider (AP) and RP collude with the aim to link authentication events of a user (U);
(b) the new **assumption** that access policy established by an RP can be satisfied with any one of 2 assertions possessed by U;

2324-9013/21/$31.00 ©2021 IEEE
DOI 10.1109/TrustCom53373.2021.00055

(c) strategy (**methodology**) that includes unlinkability definitions and tests from ISO/IEC DIS 27551, information theory to formalize test criterion using conditional entropy, and game theory to find equilibria in non-cooperative coordination games (which we call *naïve* and *tenable*) with incomplete information which are played by n players (e.g. users) in ABA system. Naïve and tenable games differ due to the information that is available to the players.

As a result of applying methodology listed in item (c) we contribute to the existing research in privacy of authentication systems with:

- new utility for unlinkability that is derived from the information-theoretic criterion;
- new model (DMMA), which defines optimal strategies for n players as well as corresponding equilibria in the games.

DMMA demonstrates that in the system of n users who authenticate to the same RP, realizations should be used *interchangeably*. This is arguably the first non-cryptographic attempt to improve user privacy using such '*multi-attribute*' assumptions by applying game-theoretic techniques. In the experimental part of the paper we demonstrate the evidence of 'enhancing user unlinkability in ABA' by testing the proposed DMMA.

C. Roadmap

The paper is structured as follows. Section II contains preliminaries and simplified illustrations for a game-theoretical approach to the problem of unlinkability. Section III provides theoretic results that explain why conditional entropy should be used to express RP's "guessing" performance (as per listing 1) followed by the Decision-Making Method for Authentication (DMMA) which is based on two different non-cooperative coordination games with incomplete information which we dub '*naïve*' and '*tenable*', respectively. Here we discuss possible distributions of the attributes owned by the players as well as the availability of common information about decisions made by the users in the game. In section IV, we conduct computational experimentation on the properties of DMMA intending to find various equilibria that demonstrate the privacy advantages of rational interchangeable attribute usage. Next, we discuss the key results of our study in section V. In section VI, we revise the work of other authors that is related to our study. We finally conclude by highlighting the main contributions in section VII.

II. Preliminaries

The interactions between a user (U), the attribute provider (AP) and the relying party (RP) are anayzed in the context of '(RP-AP-U)-model' for ABA that is defined in [11]. We will further interpret the letter 'U' as a specific user, such as *Alice* or *Bob*. Multiple authentication credentials are controlled by *Alice* and *Bob* who engage in a digital transaction with an RP. This need for multi-attribute usage is supported by numerous examples from practice including multiple certificates issued to the same entity by different Certificate Authorities (CA) as well as multiple digital credentials (such as driver license, passport, club membership) issued by various official sources [9], [12]. We focus on the scenario where every user possesses two different credentials. Credentials are supplied by the AP with whom a user must first register. A user will then utilize obtained credentials to prepare various assertions (e.g. *realizations*, proofs) to present to the RP. Privacy-preserving techniques, such as selective disclosure and/or ZKP, are usually used for realization.

Some of the resulting realizations may be identical for both *Alice* and *Bob* such that RP can not differentiate them. On the other hand, some others can be easily differentiated because the credentials they were derived from may originate from different authoritative sources (e.g. driver license and club membership). *Alice* and *Bob's* task is to coordinate usage of indistinguishable realizations to achieve a better degree of privacy expressed via criterion for unlinkability (see definition 2 and listing 1). To conserve space, we do not discuss details of the format and process of preparation for the presentations supplied by the AP while assuming that our results are equally applicable to VC and IRMA [2], [7], [8].

A. Simplified model for 2-player game

A simplified illustration of the concept of interchangeable usage of realizations for authentication purposes is depicted in fig. 1. *Alice* and *Bob* authenticate to RP using both of their realizations α and β that must satisfy policy P set by RP. α and β can be, for instance, derived from digital credentials such as *driver license* and *club membership*, respectively. In this example, realizations of the same category (e.g. from driver license) **can not** be distinguished by RP because of selective disclosure and ZKP used during authentication. Therefore 'α submitted by *Alice*' and 'α submitted by *Bob*' can not be differentiated by RP [2], [8]. Conversely, realizations from different categories (e.g. α versus β) can be well-distinguished by RP who plays an adversarial role in the model and tries to '*link*' all the authentication events initiated by the same user.

RP observes 16 authentication events over some time *t* (see fig. 1) out of which 6 events were initiated by *Alice* (*A*) and 10 events were initiated by *Bob* (*B*). Hence, for any non-attributed event observed by RP, probability that it is initiated by *A* is $\Pr(A) = \frac{6}{16} = 0.375$. Probability that the event is initiated by *B* is $\Pr(B) = \frac{10}{16} = 0.625$. Intuitively, the users could modify initial marginal distribution so that $\Pr(A) = \Pr(B)$ to achieve better privacy. In practice, the users' needs in the service provided by RP dictate $\Pr(A)$ and $\Pr(B)$. These needs usually supersede the need to remain private and, hence, $\Pr(A)$ and $\Pr(B)$ must be accepted unaltered [13].

In addition to the information about marginal distribution, RP does observe attributes and differentiates between α and β. As a result, a more refined characteristics that is utilized by RP in his analysis includes $\Pr(A \mid \alpha)$, $\Pr(A \mid \beta)$, $\Pr(B \mid \alpha)$, $\Pr(B \mid \beta)$. This in turn depends on the *decisions* made by *A* and *B* in the game. For *A*, these decisions include defining $\Pr(\alpha \mid A)$, $\Pr(\beta \mid A)$, $\Pr(\alpha \mid A) + \Pr(\beta \mid A) = 1$. For *B*

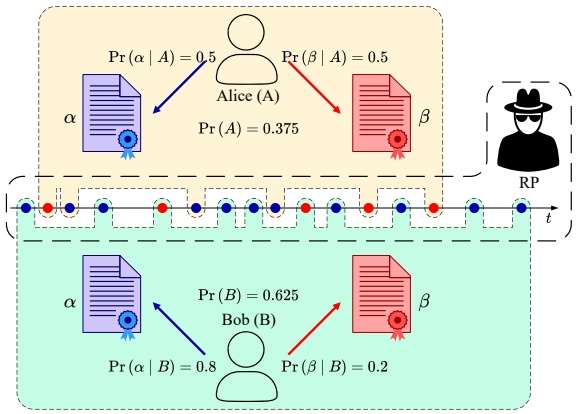


Figure 1: Diagram for interchangeable usage of verifiable realizations (complete information).

these decisions include defining $\Pr(\alpha \mid B)$, $\Pr(\beta \mid B)$, $\Pr(\alpha \mid B) + \Pr(\beta \mid B) = 1$. Using the example of decisions on fig. 1 we conclude that RP infers that $\Pr(\alpha) = \frac{11}{16}$, $\Pr(\beta) = \frac{5}{16}$, $\Pr(A \mid \alpha) = \frac{3}{11}$, $\Pr(A \mid \beta) = \frac{3}{5}$, $\Pr(B \mid \alpha) = \frac{8}{11}$, $\Pr(B \mid \beta) = \frac{2}{5}$. We further interpret the *privacy meaning* of such decisions based on statistical characteristics available to RP.

B. Privacy threats and performance in the game

Anonymity is the strongest notion of privacy. International standard ISO/IEC DIS 27551 'Requirements for attribute-based unlinkable entity authentication' describes various notions of unlinkability that can be used to express degrees of anonymity achievable in authentication systems [11], [14]. A generic definition of unlinkability refers to the inability to *link* authentication protocol executions.

Definition 1 (Linking)**.** *Is the ability for an entity or a group of colluding entities to distinguish protocol executions where an entity role is played by the same entity, from protocol executions where that entity role is played by different entities.*

Assumptions about protocol *correctness* and *unforgeability* are made in ISO/IEC DIS 27551. As such, we make similar assumptions to define performance in the game where all the authentication events are instances of protocol execution. Among all the unlinkability definitions in this standard 'RP+AP-U unlinkability' is characterized by Unlinkability Level (UL) 5 which corresponds to the *highest degree of anonymity* [11].

Definition 2 (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. The target entity role is U.*

The standardization procedure designates a test to decide whether 'inability to link' is met (see listing 1). We further regard that entities U_0 and U_1 in the test are played by *Alice* and *Bob*, respectively, from the game on fig. 1. Among the major similarities between the game and the test are: (*i*) realizations α, β in the game cohere with the set of attributes that satisfy

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

-1 *Output*: `true` or `false`
0 *Test*:
1 Adversary $\mathfrak{A}$ chooses the set of attributes for U_0, U_1 and policy P;
2 AP and RP execute the setup phase (if any);
3 AP and U_0 execute the user registration phase (if any);
4 AP and U_1 execute the user registration phase (if any);
5 RP, AP and U_0 execute the authentication phase;
6 RP, AP and U_b execute the authentication phase, $\mathrm{b} \in \{0,1\}$, $\Pr(\mathrm{b}=0) = 0.5$;
7 $\mathfrak{A}$ returns a *guess* $\mathrm{b}' \in \{0,1\}$ on the value of b;
8 **if** $\Pr(\mathrm{b}' = \mathrm{b}) \to 0.5$ return `true` ;
9 **else** return `false` .

access policy P in the test; (*ii*) authentication events (protocol executions) repeat over time; (*iii*) 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 on practice; (*b*) details of '*guessing*' procedure (line 7, listing 1) remain unclear. As a result, *Alice* and *Bob* require additional concepts to produce *best responses* in the game.

Conditional Entropy is one of the concepts that accords with the introduced game and RP+AP-U unlinkability test. Although numerous information-theoretic measures have been applied to unlinkability in the past, we are the first to demonstrate how conditional entropy limits RP's 'guessing' efficiency (line 7, listing 1) [11]. This makes it suitable to measure performance in the game such that players can directly derive their utilities from it. For example, according to the RP's inference from fig. 1 the resulting conditional entropy is 0.7915. This value can be substantially increased if, for instance, *Alice* does not change her decisions while *Bob* plays $\Pr(\alpha \mid B) = 0.5$ and $\Pr(\beta \mid B) = 0.5$. Based on that RP would infer that $\Pr(\alpha) = 0.5$, $\Pr(\beta) = 0.5$, $\Pr(A \mid \alpha) = \Pr(A \mid \beta) = 0.375$, $\Pr(B \mid \alpha) = \Pr(B \mid \beta) = 0.625$ which results in conditional entropy as high as 0.9544. As can be seen, this requires *coordination* between *Alice* and *Bob* because, for instance, *Bob's* decision depends on the one produced by *Alice*. In extreme cases of miscoordination between players, conditional entropy is 0, meaning that RP can link them with absolute success. This happens if, for example, *Alice* plays $\Pr(\alpha \mid A) = 1$ and *Bob* plays $\Pr(\beta \mid B) = 1$.

To coordinate, players should know the attributes of each other and the decisions made by their counterparts. This information may be non-deterministic and expressed in the form of priors or beliefs. Hence, the rest of the paper will attempt to answer the following **Research Question (RQ)**: *How can a user improve unlinkability in ABA systems when multiple-credentials are held by that user?*

III. Decision-Making Method for Authentication

In this section, we provide further details pertaining to the Decision Making Method for Authentication (DMMA). For

the main notations used in this paper see table I. We first demonstrate that conditional entropy is appropriate to express attacker's performance in linking authentication events to *Alice* and *Bob*. This will be used to derive players' utilities in the game. Then, we analyze the details of refined 2-player model where *Alice* and *Bob* aim at maximizing conditional entropy through coordination in various game-theoretic scenarios.

Table I: Main Notations

Notation	Description
DMMA	Decision-Making Method for Authentication
RP, AP	Relying Party, Attribute Provider
ROC	Receiver Operating Characteristics
DCW	Digital Credentials Wallet
$H(\cdot\|\cdot)$	Conditional entropy.
$\ell = \mathcal{A} \cup \mathcal{B}$	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.
$\mathtt{L} \in \{A, B\}$	Discrete random variable in set $\mathcal{L}$.
$I = \{1, ..., n\}$	Set of indices of the players.
$\mathcal{A} = \{\alpha_1, \ldots, \alpha_l\}$, $\mathcal{B} = \{\beta_1, \ldots, \beta_m\}$	Categories of attribute realizations.
$\alpha^{(i)} \in \mathcal{A}$, $\beta^{(i)} \in \mathcal{B}$	Random attributes of player i
$\mathbf{T}_i = (\alpha^{(i)}, \beta^{(i)})$	Player's i types.
$\mathcal{T} = \{\mathbf{T}_1, ...\mathbf{T}_n\}$	Set of all players' types $\mathbf{T}_i$, $i \in I$.
$\Pr_{\mathcal{S}}(\alpha^{(i)})$, $\Pr_{\mathcal{S}}(\beta^{(i)})$	Marginal probabilities at RP
$\mathcal{P}'_{\mathcal{S}} \subseteq \ell$	Set of attributes observable at RP
$\vartheta_{\mathcal{S}}$	Random vector of marginal probabilities at RP
$\mathcal{S} = \{s_1, ..., s_n\}$	Set of all continuous strategies for the players.
$\mathbf{u}$	Payoff vector $\mathcal{S} \to \mathbb{R}^{\lvert I\rvert}$
$\eth = \langle I, \mathcal{T}, \mathcal{S}, \aleph, \varrho, \mathbf{u}\rangle$	Bayesian game over the sets $I, \mathcal{T}, \mathcal{S}, \aleph, \varrho, \mathbf{u}$.
ϱ	Joint **pdf** $\mathcal{T}_{-i} \times \mathcal{S}_{-i} \to [0, 1]$
$\aleph$	Discrete joint **pmf** $\mathcal{A} \times \mathcal{B} \to [0, 1]$.
φ	Joint **pdf** $[0, 1]^{\lvert\mathcal{P}_{\mathcal{S}}\rvert} \to [0, 1]$ over $\vartheta_{\mathcal{S}}$.
$\mathbb{E}[\cdot]$	Expected value
$n \geq 2$	Number of players in the game

A. Relation between unlinkability and conditional entropy

Based on the 2-player example provided in the previous section, we label *Alice* and *Bob* using labels $\mathcal{L} = \{A, B\}$ (random variable $\mathtt{L} \in \mathcal{L}$ denoting either A or B), respectively. The set of Alice's and Bob's attributes will be denoted as $\ell = \{\alpha, \beta\}$ (random variable $l \in \ell$ denoting either α or β), respectively. We then argue that irrespective of the linking method deployed by RP, conditional entropy $H(\mathtt{L} \mid l)$ can be used to characterize the best performance achievable by that linking method. This is supported by the following lemma as well as can be observed from Receiver Operating Characteristics (ROC) curves on fig. 2. ROCs are graphical representations of the performance of a classification model at all possible classification thresholds. The graphical plot contains two parameters: the True Positive Rate (TPR) and the False Positive Rate (FPR).

Lemma 1. *Best linking performance is limited by* $H(\mathtt{L} \mid l)$ *(for proof see section A).*

The plots of ROC curves also illustrate effect of conditional entropy on fig. 2: for a given FPR the highest possible TPR decreases with $H(\mathtt{L} \mid l)$. Based on lemma 1 we propose that criterion $\mathcal{C} = H(\mathtt{L} \mid l)$ is used by A and B to produce decisions which improve unlinkability in ABA. To maximize

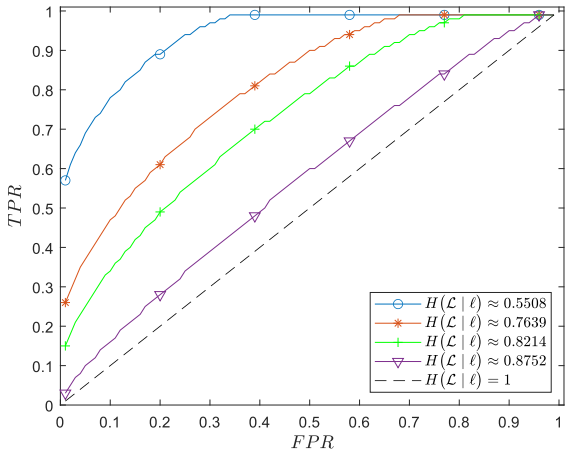


Figure 2: ROC curves for optimal linking function built by RP.

$\mathcal{C}$ users A and B could coordinate with one another. However, questions surrounding *how this coordination is carried out* still remains. Therefore, we will proceed to gradually refine the initial model introduced on fig. 1 to ensure that this is addressed. Substantial simplification for that initial model is achieved due to assumptions that: (*i*) both *Alice* and *Bob* use the same set of attributes $\ell = \{\alpha, \beta\}$; and (*ii*) the information $\Pr(\alpha \mid B)$, $\Pr(\beta \mid B)$ is known to *Alice*, and information $\Pr(\alpha \mid A)$, $\Pr(\beta \mid A)$ is known to *Bob*. We will further elaborate on this issue by analyzing a refined model.

B. The need to refine 2-player model

In real-world authentication systems, credentials are predominantly passed along with additional metadata that makes it possible to discriminate users even when the same type of credential is used. For example, although a driving license is a credential that may prove specific user's claims, it also contains metadata on what state issued it [9]. To illustrate, we present the following example: a driving license is an attribute α as per the previous diagram (see fig. 1), but state metadata makes distinguishable driver licenses issued by different states. Therefore, we will also distinguish different realizations α_1 and α_2 that both belong to a more general category 'driver licenses' (further $\mathcal{A}$). These differences dictate the need to refine the model: we will use extend set ℓ of attributes. On practice, each user may possess only a subset within ℓ.

We define the set of all possible realizations in the game as $\ell = \mathcal{A} \cup \mathcal{B}$ consisting of categories (subsets) $\mathcal{A} = \{\alpha_1, \alpha_2\}$ and $\mathcal{B} = \{\beta_1, \beta_2\}$, $\mathcal{A} \cap \mathcal{B} = \varnothing$ ($\mathcal{A}$ and $\mathcal{B}$ are not to be confused with user's labels A and B). For convenience of notations, we use the set of indices $I = \{1, 2\}$, where index $i = 1$ corresponds to *Alice* and $i = 2$ corresponds to *Bob*. We then demand that *Alice* and *Bob* possess one realization from each category. Random variables $\alpha^{(1)}, \alpha^{(2)}$ for *Alice* and *Bob*, respectively, take realizations from $\mathcal{A}$. In a similar way $\beta^{(1)}, \beta^{(2)}$ take realizations from $\mathcal{B}$. Different kinds of

decisions are made by player i in ABA which is summarized on fig. 3.

C. Decision making framework for ABA

In realistic settings, player i relies on information that needs to be collected from different sources. The player may place different levels of *trust* into these sources. For example, in ABA systems with $n \geq 2$ users information on 'how often users authenticate to RP' may be available in the form of distribution, $\forall i \ \Pr(i)$. This can be attained from different surveys, e.g. asking '*how often do you authenticate to digital platform X ?*'. Also, information about categories $\mathcal{A}, \mathcal{B}, ...$ of realizations (assertions) and how these realizations are distributed among the users (joint distribution $\aleph$ further), can be obtained based on the issuance (e.g. APs practices) and acquisition of these realizations by the users (e.g. adoption). All the mentioned information is verifiable (through census or other public statistics) and can be trusted. On the diagram (see fig. 3) we denote this high level of trust as 'trust I'.

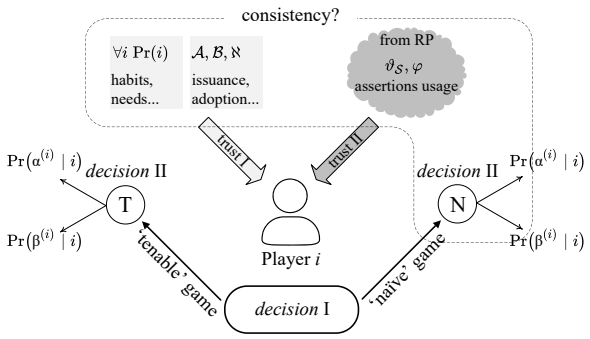


Figure 3: Decisions of player i and information available in ABA.

If players do not communicate with each other, they may rely on the information that is gathered at RP side. RP should not be involved in the dissemination of this information because he is the potential adversary in 'RP-AP-U' model. Instead, this statistics can be distributed to the users by an independent mediator who runs a proxy. As such, users willing to receive information also need to authenticate through that proxy. The construction of such scheme, nevertheless, goes beyond the scope of our paper. We assume that players obtain vector $\vartheta_{\mathcal{S}}$ of '*marginal probabilities at RP*'. This is the information about collective effect of interchangeable usage of realizations. For example, from fig. 1 $\vartheta_{\mathcal{S}} = \big(\Pr_{\mathcal{S}}(\alpha), \Pr_{\mathcal{S}}(\beta)\big)$ where $\Pr_{\mathcal{S}}(\alpha) = \frac{11}{16}$ and $\Pr_{\mathcal{S}}(\beta) = \frac{5}{16}$. This example exhibits the situation when $\vartheta_{\mathcal{S}}$ is known to the players with certainty. Alternatively, this knowledge may be in the form of priors (e.g. some probabilistic distribution φ) over $\vartheta_{\mathcal{S}}$. The information about $\vartheta_{\mathcal{S}}, \varphi$ has a different level of trust because it is provided by the mediator: on fig. 3 it is denoted as 'trust II'.

Based on the information that is collected by i from the described sources she makes '*decision* I' as to which variant of the game to play: it is either 'naïve' or 'tenable' game. Within each game player i makes '*decision* II': she defines $\Pr\big(\alpha^{(i)} \mid i\big)$ and $\Pr\big(\beta^{(i)} \mid i\big)$. *Decision* II is based on best response principle and must be *consistent* with the information provided to the players. Tenable game utilizes 'worst case' (maximin) scenario to estimate $\vartheta_{\mathcal{S}}, \varphi$: this technique can be applied even if 'trust II' is absent [15], [16]. In contrast, naïve game requires trust II to use $\vartheta_{\mathcal{S}}, \varphi$ that are provided by mediator. Consistency is a necessary condition for trust II in that game: without consistency trust is impossible. We discuss advantages and limitations of naïve and tenable games without providing any recommendations as to how *decision* I should be made. The following **Decision-Making Method for Authentication** (DMMA) embodies flow of the diagram on fig. 3 and supports decision making in ABA:

1) i acquires $\big(\alpha^{(i)}, \beta^{(i)}\big)$ from AP;
2) i obtains information about $\forall j \ \Pr(j)$, $\mathcal{A}, \mathcal{B}, \aleph$, $\vartheta_{\mathcal{S}}, \varphi$;
3) **decision I**: i assigns either `N` or `T` to variable `g`;
4) `if g == T` i calculates 'worst case' $\varpi_{\alpha^{(i)}} = \Pr_{\mathcal{S}_w}\big(\alpha^{(i)}\big)$, $\varpi_{\beta^{(i)}} = \Pr_{\mathcal{S}_w}\big(\beta^{(i)}\big)$ and go to step 6; `else` go to step 5;
5) i estimates
$$\varpi_{\alpha^{(i)}} = \mathop{\mathbb{E}}_{\varphi}\Big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\Big], \varpi_{\beta^{(i)}} = \mathop{\mathbb{E}}_{\varphi}\Big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\Big];$$
6) **decision II**: i calculates best response
$$s_i^{\mathrm{b}} = \Pr\big(\alpha^{(i)} \mid i\big) = \frac{\varpi_{\alpha^{(i)}}}{\varpi_{\alpha^{(i)}} + \varpi_{\beta^{(i)}}}.$$
`if g == T` terminate `else` go to step 7;
7) `if` information at step 2 is inconsistent with the best responses of all n assign `g` := `T` and go to step 4; `else` terminate.

Further we will analyze games with incomplete information that are suitable for the method. For example, we will: *a*) demonstrate why step 5 is important for Bayesian and Mediated games; *b*) explain expression for best response in step 6 using $H\big(\mathtt{L} \mid l\big)$; *c*) formalize consistency requirement in step 7; *d*) explain how to obtain $\varpi_{\alpha^{(i)}}$, $\varpi_{\beta^{(i)}}$ in step 4.

D. Applying games with incomplete information

In order to encompass uncertainty (incomplete information) that *Alice* and *Bob* have about the decisions of each other we consider the following game-theoretical approaches with incomplete information including **Bayesian**, **Mediated** and **Maximin** games.

1) Bayesian game: [1] The game depicted on fig. 1 is of complete information, which is impractical for authentication systems where players do not share with each other information about their attributes and decisions. This can be addressed by a variant of Bayesian game where information about characteristics of the players is represented in the form of *beliefs* (or *priors*) which are defined using statistical distributions.

Let us consider the following game $\Im = \langle I, \mathcal{T}, \mathcal{S}, \aleph, \varrho, \mathbf{u} \rangle$. We will use set $\mathcal{T} = \{\mathbf{T}_1, \mathbf{T}_2\}$ of random vectors $\mathbf{T}_1 = \big(\alpha^{(1)}, \beta^{(1)}\big)$, $\mathbf{T}_2 = \big(\alpha^{(2)}, \beta^{(2)}\big)$ which represent the *type*

[1] For the main notations see table I

of each player. Realization of type $\mathbf{T}_1$ is known to *Alice*, and realization of type $\mathbf{T}_2$ is known to *Bob*. However, $\mathbf{T}_1$ appears random to *Bob*, and $\mathbf{T}_2$ appears random to *Alice*. We describe these random vector realizations using discrete joint **p**robability **m**ass **f**unction (**pmf**) $\aleph : \mathcal{A} \times \mathcal{B} \to [0,1]$ where $\Pr\left(\alpha^{(i)} = \alpha_\iota, \beta^{(i)} = \beta_\rho\right) = \aleph_{\iota,\rho}$, and $i \in I$; $\iota \in \left\{1, ..., |\mathcal{A}|\right\}$, $\rho \in \left\{1, ..., |\mathcal{B}|\right\}$ (see $\aleph$ on fig. 4). We use $\mathcal{S} = \{s_1, s_2\}$, $s_1 = \Pr\left(\alpha^{(1)} \mid A\right)$ and $s_2 = \Pr\left(\alpha^{(2)} \mid B\right)$ to describe *decisions* (in pure continuous strategies) of *Alice* and *Bob*, respectively. Because s_1 is random for *Bob* and s_2 is random for *Alice* we use continuous **p**robability **d**ensity **f**unction (**pdf**) $\varrho_{\mathcal{T}_{-i}} : \mathcal{T}_{-i} \times \mathcal{S}_{-i} \to [0,1]$, to describe decision of (all) other player(s) $-i$ whose type(s) is/are $\mathcal{T}_{-i}$. We also consider that information carried out by $\aleph$ and $\varrho_{\mathcal{T}_{-i}}$ is *symmetric* for *Alice* and *Bob*. Finally, we define the vector $\mathbf{u} : \mathcal{S} \to \mathbb{R}^{|I|}$ of utilities for the players in the game where component u_i specifies the utility of i. Based on $\mathbf{T}_1$, s_1, $\aleph$, $\varrho_{\mathcal{T}_{-i}}$, player *Alice* calculates her *expected utility* $\mathop{\mathbb{E}}_{\aleph,\varrho_{\mathcal{T}_{-i}}}\left[u_1\right]$.

Definition 3 (Best response)**.** *The best response s_i^{b} of player i must satisfy* $\mathop{\mathbb{E}}_{\aleph,\varrho_{\mathcal{T}_{-i}}}\left[u_i^{\mathrm{b}}\right] \geq \mathop{\mathbb{E}}_{\aleph,\varrho_{\mathcal{T}_{-i}}}\left[u_i\right]$.

Definition 4 (Bayes Nash Equilibrium – BNE)**.** *It is the condition of the game where every i plays s_i^{b}.*

The state of equilibrium is a stable (e.g. long-lasting) state. As such, characteristics of authentication systems, including its unlinkability can be estimated in this state. Due to this, we analyze equilibria states only. Multiple equilibria (where $\varrho_{\mathcal{T}_{-i}}$ may differ) can exist in the game and one of the main criticisms of Bayesian games is the necessity to synchronize information about $\varrho_{\mathcal{T}_{-i}}$ across all the players (unlike $\aleph$ which is defined by AP, is known to the players, and remains unchanged). This can be addressed in *mediated games*.

2) Mediated game: In a mediated game, synchronization can be achieved if information that is *sufficient* for calculation of *best response* is directly provided to the players. This contrasts with the Bayesian game where i requires priors (or beliefs) about decisions of $-i$ players.

Intuition for a mediation game can be explained in the following 3 steps:

(1) As a result of players executing their decisions $\mathcal{S}$, RP observes the set of realizations $\mathcal{P}_{\mathcal{S}}^{'} \subseteq \mathcal{P}_{\mathcal{S}}$ where $\mathcal{P}_{\mathcal{S}} = \left\{\alpha^{(1)}, \alpha^{(2)}, \beta^{(1)}, \beta^{(2)}\right\}$, $\mathcal{P}_{\mathcal{S}} \subseteq \ell$. Cardinality of $\mathcal{P}_{\mathcal{S}}^{'}$ satisfies $1 \leq \left|\mathcal{P}_{\mathcal{S}}^{'}\right| \leq 4$, depending on the number of attributes that are used by *Alice* and *Bob* as well as the number of realization of these attributes that match. For example, when both *Alice* and *Bob* use their realizations interchangeably (e.g. each player uses 2 attributes), and none of their realizations match, the cardinality of $\mathcal{P}_{\mathcal{S}}^{'}$ is 4. However, if the players use the same realization (across all of their authentication sessions) the cardinality of $\mathcal{P}_{\mathcal{S}}^{'}$ is 1.

(2) We define *marginal probabilities at RP* (subscript $\mathcal{S}$) such that, for instance, $\Pr_{\mathcal{S}}\left(\alpha^{(1)}\right)$ is the probability that a random authentication event at RP is executed using realizations which is indistinguishable from realization of $\alpha^{(1)}$. Players may also have *beliefs* about random probability vectors $\vartheta_{\mathcal{S}} = \left(\Pr_{\mathcal{S}}\left(\alpha^{(1)}\right), \Pr_{\mathcal{S}}\left(\alpha^{(2)}\right), \Pr_{\mathcal{S}}\left(\beta^{(1)}\right), \Pr_{\mathcal{S}}\left(\beta^{(2)}\right)\right)$ of marginal probabilities at RP, $\vartheta_{\mathcal{S}} \in [0,1]^{|\mathcal{P}_{\mathcal{S}}|}$. These beliefs are captured by joint **pdf** $\varphi : [0,1]^{|\mathcal{P}_{\mathcal{S}}|} \to [0,1]$.

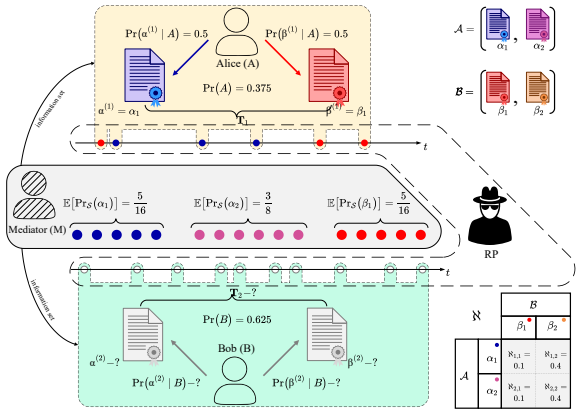


Figure 4: Mediated decision made by *Alice* in refined model (incomplete information).

(3) We then introduce *Mediator* (*M* for short) who provides information about these beliefs to the players (e.g. synchronizes φ among them). This information, for instance, can also be in the form of compact statistical characteristic, such as expectation $\mathbb{E}[\vartheta_{\mathcal{S}}]$ (see fig. 4). Later it will be demonstrated (for $n \geq 2$ players) that this characteristic is sufficient to calculate expected utilities, best responses and, hence, is sufficient for establishing equilibrium.

3) Utility and best responses in mediated game with $n \geq 2$ players: We briefly outline major points in relation to our analysis of the coordination game with $n \geq 2$ players while more detailed description can be found in the full version of the paper [17].

Defining expressions for expected utility and best responses requires specifying conditional entropy $H\left(\mathtt{L} \mid l\right)$. At this stage, we do not have any preferences as for distribution $\left\{\Pr(i)\right\}_{1 \leq i \leq n}$. We therefore assume that $\forall i \Pr(i) = \frac{1}{n}$. We then define user utilities based on the following Lemma:

Lemma 2. *Expected utility for player i is (for the proof see [17])*

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

from which, *best response* of player i is:

$$s_i^{\mathrm{b}} = \frac{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right] + \mathbb{E}\left[\Pr_{\mathcal{S}}\left(\beta^{(i)}\right)\right]}, \tag{2}$$

Expected unlinkability $\mathbb{E}\left[\mathcal{C}\right]$, $\mathcal{C} = H\left(\mathtt{L} \mid l\right)$ of the whole system is

$$\mathbb{E}\left[\mathcal{C}\right] = \log n + \sum_i \mathbb{E}\left[u_i\right]. \tag{3}$$

For instance, from the diagram on fig. 4 it is clear that type $\mathbf{T}_1$ of *Alice* has realization $\{\alpha_1, \beta_1\}$ while $\mathop{\mathbb{E}}_{\varphi}[\vartheta_{\mathcal{S}}] = \left(\frac{5}{16}, \frac{3}{8}, \frac{5}{16}, 0\right)$. Hence, her best response calculated in accordance to eq. (2) is $s_1^{\mathrm{b}} = \frac{\frac{5}{16}}{\frac{5}{16}+\frac{5}{16}} = 0.5$. Mediated games provide overall better performance compared to Bayesian games. On the other hand, decision of *Alice* is heavily reliant on *M* implying that its *truthfulness* is of great importance for the system. For i, an assurance that *M* is truthful supports i's trust (see 'trust II' on fig. 3).

4) Consistency in mediated game: Consistency of the information provided by *M* in the context of other information available to the players is one of the *necessary* conditions for its truthfulness. As such, coordination mechanism facilitating unlinkability in authentication systems must be consistent.

For instance, let us demonstrate the requirement for consistency in the case when *Alice* type is $\{\alpha_1, \beta_1\}$ and she knows $\aleph$, φ, ϱ (see fig. 4). We admit that

$$\begin{aligned}&\Pr_{\mathcal{S}}\big(\alpha^{(1)}\big) = \Pr\big(\alpha^{(1)}, A\big) + \Pr\big(\alpha^{(1)}, B\big) = \\ &\Pr(A)\Pr\big(\alpha^{(1)} \mid A\big) + \Pr(B)\Pr\big(\alpha^{(1)} = \alpha^{(2)}\big)\Pr\big(\alpha^{(2)} \mid B\big)\,,\end{aligned} \tag{4}$$

from which we obtain

$$\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\alpha^{(1)}\big)\big] = \Pr(A)s_1 + \Pr(B)\Big(\aleph_{1,1}\mathop{\mathbb{E}}_{\varrho_{1,1}}[s_2] + \aleph_{1,2}\mathop{\mathbb{E}}_{\varrho_{1,2}}[s_2]\Big)\,, \tag{5}$$

where $\varrho_{\iota,\rho}$ is a short notation for the distribution of *Bob's* decisions when his type is $\mathbf{T}_2 = \{\alpha_\iota, \beta_\rho\}$. In a similar way

$$\begin{aligned}&\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\beta^{(1)}\big)\big] = \Pr(A)(1 - s_1) + \\ &\Pr(B)\left(\aleph_{1,1}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{1,1}}[s_2]\Big) + \aleph_{2,1}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{2,1}}[s_2]\Big)\right)\,,\end{aligned} \tag{6}$$

$$\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\alpha^{(2)}\big)\big] = \Pr(B)\Big(\aleph_{2,1}\mathop{\mathbb{E}}_{\varrho_{2,1}}[s_2] + \aleph_{2,2}\mathop{\mathbb{E}}_{\varrho_{2,2}}[s_2]\Big)\,, \tag{7}$$

$$\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\beta^{(2)}\big)\big] = \Pr(B)\left(\aleph_{1,2}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{1,2}}[s_2]\Big) + \aleph_{2,2}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{2,2}}[s_2]\Big)\right) \tag{8}$$

Notably, *M* on fig. 4 is inconsistent with the information about $\aleph$ that is available to *Alice*. In order to demonstrate this we first notice that the 4^{th} component of $\mathop{\mathbb{E}}_{\varphi}[\vartheta_{\mathcal{S}}]$ is $\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\beta^{(2)}\big)\big] = 0$ meaning that $s_2 = 1$ for the cases when either $\mathbf{T}_2 = \{\alpha_1, \beta_2\}$ or $\mathbf{T}_2 = \{\alpha_2, \beta_2\}$. Then, for consistency it is required that, for example, taking into account $\mathop{\mathbb{E}}_{\varrho_{2,2}}[s_2] = 1$ we obtain from eq. (7)

$$\mathop{\mathbb{E}}_{\varrho_{2,1}}[s_2] = \frac{1}{\aleph_{2,1}}\left(\frac{\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\alpha^{(2)}\big)\big]}{\Pr(B)} - \aleph_{2,2}\right) = 2\,, \tag{9}$$

which is impossible because $0 \le s_2 \le 1$. With the aim to design coordination mechanism where *M* is consistent we assign to $\mathop{\mathbb{E}}_{\varphi}[\vartheta_{\mathcal{S}}]$ values from corresponding complete information Nash equilibria [18]. It, nevertheless, should be noted that consistency is necessary but not sufficient for truthfulness implying that players must *trust M* (see 'trust II' on fig. 3). For that reason we further dub mediated game '*Naïve game*'. The issue of trust can be addressed in the next paragraph.

5) Maximin game: Here we consider a *trustless* environment where players' decisions in the game are not based on external information except $\Pr(A)$, $\Pr(B)$, and $\aleph$. Players produce their best responses in accordance to Walds' maximin principle where they optimize utilities for the worst case scenario [15]. For example, let us ponder over best expected utility of *Alice* for the worst case scenario w:

$$\mathbb{E}_w[u_1^{\mathrm{b}}] = \max_{s_1}\min_{\varrho_{\mathbf{T}_2}}\mathop{\mathbb{E}}_{\aleph,\varrho_{\mathbf{T}_2}}[u_1] = \max_{s_1}\sum_{\iota}\sum_{\rho}\aleph_{\iota,\rho}\min_{\varrho_{\iota,\rho}}\mathop{\mathbb{E}}_{\varrho_{\iota,\rho}}[u_1]\,. \tag{10}$$

In the full version of this paper [17] we show that the solution for $\varrho_{\iota,\rho}$ is a degenerate distribution where the only possible outcome for *Bob* with type $\mathbf{T}_2 = \{\alpha_\iota, \beta_\rho\}$ is $s_2 = 0$, if $\sum_{\rho}\aleph_{\iota,\rho} \ge \sum_{\iota}\aleph_{\iota,\rho}$, and $s_2 = 1$, otherwise. Maximin game can then be considered a special case of mediated game on fig. 4 where information about $\mathop{\mathbb{E}}_{\varphi}[\vartheta_{\mathcal{S}}]$ is calculated by *Alice* instead of *M*. According to maximin principle she would then assume that *Bob* produces decisions: (*i*) $s_2 = 0$ if $\mathbf{T}_2 = \{\alpha_1, \beta_1\}$; (*ii*) $s_2 = 1$ if $\mathbf{T}_2 = \{\alpha_1, \beta_2\}$; (*iii*) $s_2 = 0$ if $\mathbf{T}_2 = \{\alpha_2, \beta_1\}$; (*iv*) $s_2 = 1$ if $\mathbf{T}_2 = \{\alpha_2, \beta_2\}$. Using eqs. (5) and (6) she would then obtain

$$\begin{aligned}&\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\alpha^{(1)}\big)\big] = \Pr(A)s_1 + \Pr(B)\aleph_{1,2} \\ &\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\beta^{(1)}\big)\big] = \Pr(A)(1 - s_1) + \Pr(B)\big(\aleph_{1,1} + \aleph_{2,1}\big)\,,\end{aligned} \tag{11}$$

which is substituted into eq. (2) to obtain

$$\begin{aligned}s_1^{\mathrm{b}} = \frac{\Pr(A)s_1^{\mathrm{b}} + \Pr(B)\aleph_{1,2}}{\Pr(A) + \Pr(B)\big(\aleph_{1,1} + \aleph_{1,2} + \aleph_{2,1}\big)} = \\ \frac{\aleph_{1,2}}{\aleph_{1,1} + \aleph_{1,2} + \aleph_{2,1}} = \frac{2}{3}\,,\end{aligned} \tag{12}$$

that maximizes her expected utility from eq. (10). This result means that over the period of time observable in Figure 4 she would use $\alpha^{(1)}$ in four authentication sessions while $\beta^{(1)}$ would be used only twice. Due to its trustless property we will further call maximin game '*tenable game*'.

IV. Experiment

To evaluate the impact of DMMA on privacy in ABA we asses our game-theoretical results by conducting numerical evaluations for the system with $n \gg 2$ users

a) The goal of experiment.: We address RQ by comparing: (*i*) unlinkability in ABA as per naïve game (e.g. game with mediator) with the unlinkability in ABA as per tenable game (e.g. maximin); (*ii*) unlinkability in ABA where users are guided by rational principles such as best responses in the games with the unlinkability in ABA where users make '*alternative*' decisions. To find solutions for nonlinear systems we run our experiment in Matlab using the trust region algorithm [19]. It is remarkable that (according to eqs. (1) to (3)) $\Pr(i)$, $\aleph$, $\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big]$, $\mathop{\mathbb{E}}_{\varphi}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big]$ are the only information which is required to make a decision as for attribute usage in ABA while ϱ, φ are not required. Based on eq. (1) 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.

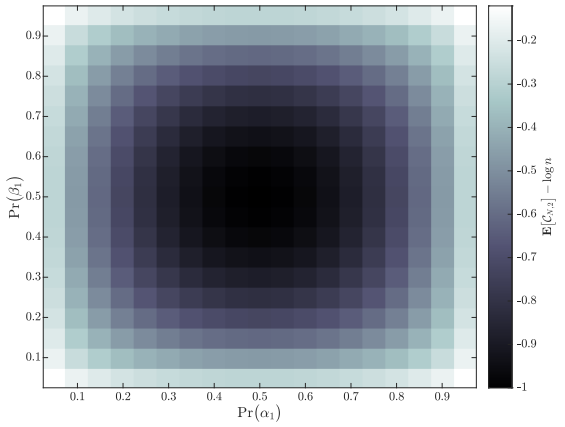
A. Experiment organization

For baseline scenarios, we consider 'unrestricted rationality' where 2 attribute realizations $\{\alpha^{(i)}, \beta^{(i)}\}$ available to player i can be used interchangeably in naïve and tenable games (see sections III-D2 and III-D5). 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 paper we identify: (*a*) 'restricted rationality' where users play naïve or tenable game but (in contrast to interchangeable usage) select and always use the same realization out of 2 realizations available to them; (*b*) '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 (as defined in (3)) in rational scenarios is denoted by $\mathbb{E}[\mathcal{C}_{\kappa,\mu}]$ 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 $\mathbb{E}\left[\mathcal{C}_{\{\kappa,\mu\}^r}\right]$ is for expected unlinkability measured under random moves scenario (index 'r').

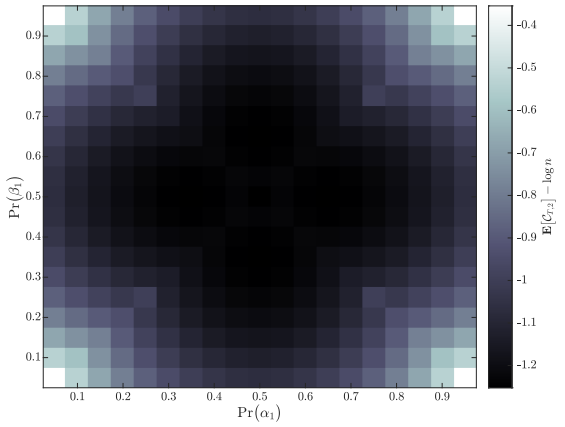
In order to produce *outputs* in the form of expected unlinkability, our experiment requires the following *inputs*: 1) $\{\kappa, \mu\}$ or $\{\kappa, \mu\}^r$; and 2) $\Pr(i)$, for all players i and the **pmf** $\aleph$. For all the instances of experiment, we consider n users and $\Pr(i) = \frac{1}{n}$ 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 5-7). Coordinates $\big(\Pr(\alpha_1), \Pr(\beta_1)\big)$ of each point on the map define a corresponding 2×2 matrix $\aleph$: $\aleph = \big[\Pr(\alpha_1),\ 1 - \Pr(\alpha_1)\big]^{\mathrm{T}} \times \big[\Pr(\beta_1),\ 1 - \Pr(\beta_1)\big]$ where both $\Pr(\alpha_1)$, $\Pr(\beta_1)$ were quantized with 0.05 step on interval $[0,\ \ 1]$. Color intensity corresponds to unlinkability..

B. Results

We first calculated equilibria for our baseline scenarios of naïve and tenable games where players can use both of their attribute realizations interchangeably (see fig. 5). For each possible $\aleph$ in naïve game we solved complete information Nash equilibria to find $\mathop{\mathbb{E}}_{\varphi}[\vartheta_{\mathcal{S}}]$ that need to be communicated to the players by mediator. Among all the possible solutions we selected those maximizing $\mathbb{E}[\mathcal{C}_{N,2}]$. For each possible $\aleph$ in tenable game we calculated worst case condition that may be created for player i by others $n - 1$ players. Then, best response of i, and $\mathbb{E}[\mathcal{C}_{T,2}]$ are calculated (see eq. (2)). As can be observed from comparison of fig. 5a and fig. 5b naïve game provides substantially better unlinkability.



(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. $\theta_{\iota,\rho} \in [0, 1]$

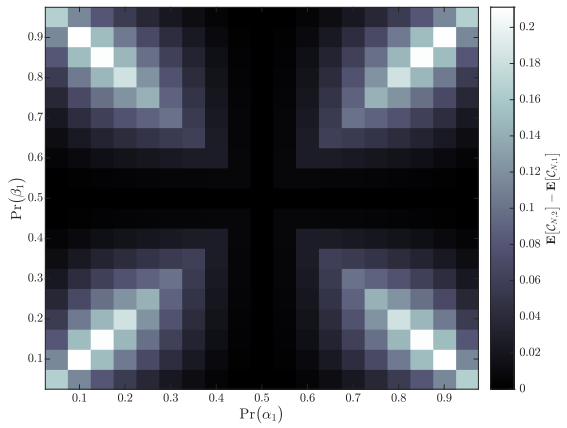
Figure 5: Comparison of expected unlinkability in naïve and tenable baseline scenarios.

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 full version of the paper [17]). The benefits of using 2 attributes (unconstrained rationality) versus 1 attribute (constrained rationality) can be observed by comparing residual unlinkabilities on fig. 6 which are greater than 0 for the both heatmaps.

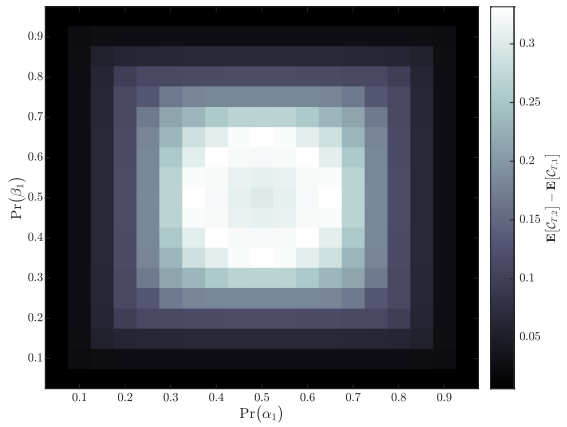
We conducted a range of experiments with randomized moves which results are presented on fig. 7. For the 2-attribute randomized game, each player i decides $0 \leq s_i \leq 1$ at random in accordance to uniform distribution on $[0, 1]$. As can be seen from the residuals of expected unlinkabilities, even constrained rationality (1 attribute usage) scenario outperforms scenario where 2 realizations are used randomly (chaotically).

V. Discussion

The cross-comparison of results from section IV demonstrates how proposed DMMA impacts the rate of user *unlink-*

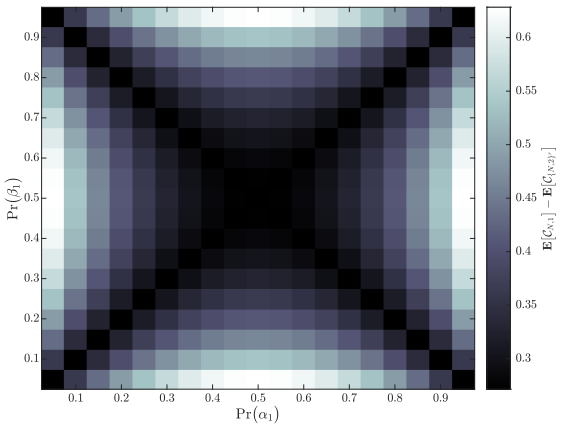


(a) Difference between expected unlinkability $\mathbb{E}\left[\mathcal{C}_{N,2}\right]$ and $\mathbb{E}\left[\mathcal{C}_{N,1}\right]$.

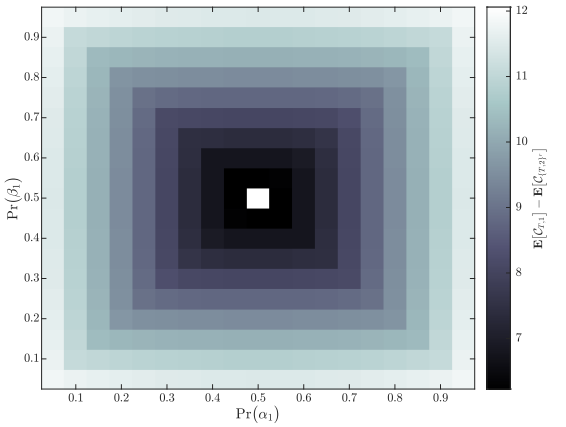


(b) Difference between expected unlinkability $\mathbb{E}\left[\mathcal{C}_{T,2}\right]$ and $\mathbb{E}\left[\mathcal{C}_{T,1}\right]$.

Figure 6: Residual expected unlinkability in 'Naïve game' and 'Tenable' games.



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



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

Figure 7: Residual expected unlinkability in 'Naïve' and 'Tenable' games.

ability in ABA systems. In this section we further contribute to RQ by discussing the details of (*i*) usage of attribute realizations in addition to (*ii*) distribution of attribute realizations among the users.

As per DMMA, there is a clear contrast between unlinkability in ABA systems where users are guided by different principles of ***realization usage***. In section IV we differentiate between *rational* and *alternative* principles of usage (see figs. 6 and 7). From the results it is clear that rational principles of interchangeable usage where users *coordinate* have substantial benefit over other alternative scenarios. This is because non-cooperative game-theoretical approaches optimize impact on unlinkability (through *best responses*) produced by every individual user i taking into account best responses of other users. In spite of this we emphasize that game-theoretical approaches do differ and, hence, their impacts on unlinkability in ABA systems are not equal. This difference is due to various amount of *context information* that is available to users in *naïve* and *tenable* games. In addition, this *context information* for coordination can be supplied to the players in various ways [20]. We contemplate that expectation $\mathbb{E}_{\varphi}[\vartheta_S]$ calculated using priors φ over the vector ϑ_S of marginal probabilities for attribute realizations observable by RP (in '*naïve*' variant of the game) can become a viable option in support of better decisions. First, this information is sufficient for each player to produce best response (see eq. (2)). Second, this may be shared with the players in differentially private form [18], [21]. However, if this information is not available, player i may resort to best response under the worst case scenario (e.g. '*tenable*' game) which comes at the cost of lower unlinkability compared to naïve scenario (see fig. 5) [15]. We, nevertheless, do not provide recommendation as to which among the naïve and tenable game scenario to chose for ABA systems (see '*decision* I' on fig. 3). This is because these different decision making concepts require various levels of *trust* (see 'trust II' on fig. 3): naïve game is reliant on mediator *M*, while tenable

game can be executed in a trustless environment.

In addition, we also gain insights into how the ***distribution of attribute realizations*** used by the users impacts their unlinkability. Properties of joint *distribution* $\aleph$ substantially affect expected unlinkability in ABA systems. For example, it can be seen that for naïve and tenable games expected unlinkabilities are lower towards the center of corresponding heatmaps on figs. 5a and 5b. This is because that area represents more diverse distributions which further constrains coordination effect. In contrast, outer areas of these maps represent the cases when majority of the players have the same type.

VI. Related Work

1) Privacy and Unlinkability: Unlinkability refers to the ability for a user to perform actions and undertake tasks without others being able to link these actions together [22]. In the context of authentication, multiple studies have identified how unlinkability significantly impacts user privacy, as it is one of the primary conditions of remaining anonymous within a digital environment [11], [14]. Below, we synthesis the main applications of unlikability in the context of privacy and authentication based on studies to date.

Firstly, studies have applied unlinkability tests to determine whether an attacker is able to guess the label of the entity or the relation between them (i.e. 'link'), and contrasted these 'guesses' with an attacker acting at random [23]–[25]. One such test is ISO/IEC DIS 27551 "Requirements for attribute-based unlinkable entity authentication" as per listing 1, which recognizes and explicitly defines the threat of linkability and profiling pertaining to authentication for the system consisting of AP, users U_0, U_1, and RP.

Secondly, studies have also quantified the linkability of items in a system by applying information-theoretical descriptions [26], [27]. For example, a basic information-theoretic notion for unlinkability is presented by [28] where they utilize Shannon entropy to measure unlinkability of elements within one set as well as between the sets. Further improvements to this notion were then added by [29] where they provided specific context information across 7 special cases. It must be stressed that the hints that the attacker gathers to create relational links about the user cannot be generalized and must be determined based on a case-by-case basis. This is exemplified in studies such as [30]. The authors propose an extensive taxonomy of privacy metrics which, for instance, describes 17 different entropy-based measures.

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 [30] 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 are used.

2) Game Theory Applications to Privacy: A number of papers apply game theory to address privacy issues either based on problems derived from practice [23], [31], [32], or focusing on the theoretical aspects of game theory [18], [33].

From a practical approach, there are several studies that explored the challenges pertaining to pseudonym change in mobile networks [23], [32]. In [32], the authors elaborate on user-centric location privacy model which 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 [23] authors analyze a game where local adversary is equipped with multiple eavesdropping stations to track mobile users who deploy mix zones in order 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 which is quantified based on real road-traffic information. From a theoretical perspective, a number of different studies have examined the coordination scenarios which impact privacy in general [18], [33]. For instance, authors of [18] discuss a game with mediating mechanism that can improve the outcome of the game when compared to Bayes Nash Equilibrium (BNE). It also demonstrates 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.

Summary: 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.

VII. Conclusion

In this paper, we demonstrated that unlinkability in ABA can be improved. This requires that necessary attention is given to the aspects of interchangeable usage of assertions possessed by a user. That line of thoughts comprises a new research direction. It contrasts with the traditional approach to the problem of unlinkability that is practiced within the research community – to make assertions indistinguishable. Due to this, the question '*How to best use these assertions if indistinguishability fails with non-zero probability?*' has been largely ignored. With the aim to contribute to this new topic we proposed a framework and DMMA in section III: it is based on rational decision-making approaches.

Using conditional entropy, we measured the strongest notion of unlinkability specified by ISO 27551 for *attribute based authentication*. We believe that this is the most optimal benchmark because it allows to encompass various levels of *context information* that may be available to adversary as well as players in real world settings. Players' utilities and their best responses are then derived for two (*naïve* and *tenable*) different instances of non-cooperative coordination game with incomplete information.

The equilibria calculated in the experimental part of our paper clearly indicates that the rational approach to the problem outperforms the alternative approaches. For instance, the habitual usage of the same assertion or random usage of many available assertions. As such, we conclude by recommending that the proposed DMMA be adopted by those working on Digital Credential Wallets (DCW). This can improve unlinkability in ABA in a way that is easy and convenient for a user and does not require modifications of existing authentication protocols.

ACKNOWLEDGMENT

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

REFERENCES

[1] G. of Ontario, "Ontario's verifiable businesses," *https://www.von.gov.on.ca/en/home*, 2020. [Online]. Available: https://www.von.gov.on.ca/en/home.

[2] W3C, "Verifiable credentials data model v1.0," *https://www.w3.org/TR/vc-data-model/*, 2020. [Online]. Available: https://www.w3.org/TR/vc-data-model/.

[3] J. Camenisch and E. Van Herreweghen, "Design and implementation of the idemix anonymous credential system," in *Proceedings of the 9th ACM conference on Computer and communications security*, 2002, pp. 21–30.

[4] C. Paquin, "U-prove technology overview v1. 1," *Microsoft Corporation Draft Revision*, vol. 1, 2011.

[5] J. Camenisch, S. Krenn, A. L. G. Mikkelsen, G. Neven, and M. Pedersen, "D3. 1: Scientific comparison of abc protocols," *Part I-Formal Treatment of Privacy-Enhancing Credential Systems. Project deliverable in ABC4Trust*, 2014.

[6] A. Sabouri, "Understanding the determinants of privacy-abc technologies adoption by service providers," in *Conference on e-Business, e-Services and e-Society*, Springer, 2015, pp. 119–132.

[7] G. Alpár, F. van den Broek, B. Hampiholi, B. Jacobs, W. Lueks, and S. Ringers, "Irma: Practical, decentralized and privacy-friendly identity management using smartphones," in *10th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2017)*, 2017.

[8] Z. Zhang, M. Król, A. Sonnino, L. Zhang, and E. Rivière, "El passo: Privacy-preserving, asynchronous single sign-on," *arXiv preprint arXiv:2002.10289*, 2020.

[9] F. Karegar, C. Striecks, S. Krenn, F. Hörandner, T. Lorünser, and S. Fischer-Hübner, "Opportunities and challenges of credential," in *IFIP International Summer School on Privacy and Identity Management*, Springer, 2016, pp. 76–91.

[10] A. Pashalidis and C. J. Mitchell, "Limits to anonymity when using credentials," in *International Workshop on Security Protocols*, Springer, 2004, pp. 4–12.

[11] ISO Central Secretary, "Information technology – requirements for attribute-based unlinkable entity authentication," en, International Organization for Standardization, Geneva, CH, Standard ISO/IEC DIS 27551, 2020. [Online]. Available: https://www.iso.org/standard/72018.html.

[12] P. Grassi, J. Fenton, N. Lefkovitz, J. Danker, Y.-Y. Choong, K. Greene, and M. Theofanos, "Digital identity guidelines: Enrollment and identity proofing," National Institute of Standards and Technology, Tech. Rep., 2017.

[13] S. Preibusch, D. Kübler, and A. R. Beresford, "Price versus privacy: An experiment into the competitive advantage of collecting less personal information," *Electronic Commerce Research*, vol. 13, no. 4, pp. 423–455, 2013.

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

[15] M. Sniedovich, "Wald's mighty maximin: A tutorial," *ITOR*, vol. 23, no. 4, pp. 625–653, 2016. DOI: 10.1111/itor.12248. [Online]. Available: https://doi.org/10.1111/itor.12248.

[16] A. Wald, "Statistical decision functions," *Ann. Math. Statist.*, vol. 20, no. 2, pp. 165–205, Jun. 1949. DOI: 10.1214/aoms/1177730030. [Online]. Available: https://doi.org/10.1214/aoms/1177730030.

[17] M. Slavnenko, V. Kuchta, J. Jeong, Y. Zolotavkin, and R. Doss, "Enhancing privacy through dmma: Decision-making model for authentication," https://doi.org/10.6084/m9.figshare.13560587, Tech. Rep.

[18] M. Kearns, M. Pai, A. Roth, and J. Ullman, "Mechanism design in large games: Incentives and privacy," in *Proceedings of the 5th conference on Innovations in theoretical computer science*, 2014, pp. 403–410.

[19] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on optimization*, vol. 6, no. 2, pp. 418–445, 1996.

[20] R. Cooper, D. V. DeJong, R. Forsythe, and T. W. Ross, "Communication in coordination games," *The Quarterly Journal of Economics*, vol. 107, no. 2, pp. 739–771, May 1992, ISSN: 0033-5533. DOI: 10.2307/2118488.

eprint: https://academic.oup.com/qje/article-pdf/107/2/739/5203164/107-2-739.pdf. [Online]. Available: https://doi.org/10.2307/2118488.

[21] D. Bergemann and S. Morris, "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, vol. 11, no. 2, pp. 487–522, 2016. DOI: 10.3982/TE1808. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.3982/TE1808.

[22] V. Katos, "Managing is security and privacy," in *Cyber Crime: Concepts, Methodologies, Tools and Applications*, IGI Global, 2012, pp. 1246–1254.

[23] M. Humbert, M. H. Manshaei, J. Freudiger, and J.-P. Hubaux, "Tracking games in mobile networks," in *International Conference on Decision and Game Theory for Security*, Springer, 2010, pp. 38–57.

[24] M. Neubauer, "Modelling of pseudonymity under probabilistic linkability attacks," in *2009 International Conference on Computational Science and Engineering*, IEEE, vol. 3, 2009, pp. 160–167.

[25] R. Maronna, D. Martin, and V. Yohai, *Robust Statistics: Theory and Methods*, ser. Wiley Series in Probability and Statistics. Wiley, 2006, ISBN: 9780470010921. [Online]. Available: https://books.google.com.au/books?id=iFVjQgAACAAJ.

[26] R. Berman, A. Fiat, and A. Ta-Shma, "Provable unlinkability against traffic analysis," in *International Conference on Financial Cryptography*, Springer, 2004, pp. 266–280.

[27] C. Dwork and K. Nissim, "Privacy-preserving datamining on vertically partitioned databases," in *24th Annual International Cryptology Conference (CRYPTO 2004)*, ser. Lecture Notes in Computer Science, vol. 3152, Springer Verlag, Aug. 2004, pp. 528–544.

[28] S. Steinbrecher and S. Köpsell, "Modelling unlinkability," in *International Workshop on Privacy Enhancing Technologies*, Springer, 2003, pp. 32–47.

[29] M. Franz, B. Meyer, and A. Pashalidis, "Attacking unlinkability: The importance of context," in *International Workshop on Privacy Enhancing Technologies*, Springer, 2007, pp. 1–16.

[30] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, pp. 1–38, 2018.

[31] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving k-anonymity in location based services," in *2013 Proceedings IEEE INFOCOM*, IEEE, 2013, pp. 2985–2993.

[32] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes, "On non-cooperative location privacy: A game-theoretic analysis," in *Proceedings of the 16th ACM conference on Computer and communications security*, 2009, pp. 324–337.

[33] A. Ghosh and K. Ligett, "Privacy as a coordination game," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IEEE, 2013, pp. 1608–1615.

APPENDIX A
PROOF OF LEMMA 1

Lemma 1. *Best linking performance is limited by* $H(\mathtt{L} \mid l)$ *(for proof see section A).*

Proof. In order to link authentication sessions RP labels them with $\mathtt{L}' \in \mathcal{L}'$, where $\mathcal{L}' = \{A', B'\}$. We divide the proof in 2 parts: (*i*) we demonstrate that for the best linking performance RP aims to minimize $H(\mathtt{L} \mid \mathtt{L}')$; (*ii*) and, $H(\mathtt{L} \mid \mathtt{L}') \geq H(\mathtt{L} \mid l)$.

(*i*) We express linking performance $\mathfrak{P}$ of RP as the difference between True Positive Rate (TPR) and False Positive Rate (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 can not 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 $H(\mathtt{L} \mid \mathtt{L}')$. We note that conditional entropy

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

is unimodal on $\Pr(A', A)$ (similar must be stated about $\Pr(A', B)$) by analyzing its first derivative $\partial \frac{H(\mathtt{L}|\mathtt{L}')}{\partial \Pr(A,A')} = \log\big(\Pr(A, A') + \Pr(B, A')\big) + \log \Pr(A, B') - \log \Pr(A, A') - \log\big(\Pr(A, B') + \Pr(B, B')\big)$ 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 $H(\mathtt{L} \mid \mathtt{L}')$ on $\Pr(A', A)$ (and on $\Pr(A', B)$) maximization of $\mathfrak{P}$ requires minimization of $H(\mathtt{L} \mid \mathtt{L}')$.

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