Enhancing Privacy Through DMMA: Decision-Making Model for Authentication

Maksym Slavnenko[1], Veronika Kuchta[2], Yevhen Zolotavkin[1,3], Jongkil Jay Jeong[1,3], and Robin Ram Mohan Doss[1]

[1] Deakin University, Geelong, Australia
[2] The University of Queensland, Brisbane, Australia
[3] Cyber Security Cooperative Research Centre, Melbourne, Australia

Abstract. A number of recently introduced formats for cryptographically protected digital credentials enable privacy preserving Attribute-Based Authentication (ABA) where users obtain total control over the features that are stored and released from Digital Credential Wallets (DCW) based on their choice. This choice implies that the decision-making process is placed solely in the hands of the user, which we demonstrate through our Decision-Making Model for Authentication (DMMA) may result in additional privacy threats pertaining to Relying Party (RP) linking users without their consent. The degree of such malicious activity is measured by the criterion of unlinkability which is synthesized based on the definitions of international standard ISO 27551 as well as the information theoretic measure of conditional entropy. The DMMA further demonstrates how users may be provided practical recommendations over the optimal selection of credentials for ABA, and suggests how this can be implemented as a feature within DCW. In order to achieve that, the task is formalized as a non-cooperative coordination game where players (targets of the attack) maximize their utilities by using credentials *interchangeably*. The experiment demonstrates that a number of privacy-enhancing equilibria can be achieved under conditions that vary based on the distribution of user attributes as well as other information provided in the game.

Keywords: Unlinkability · Decision making · Game theory · Verifiable Credentials · Authentication.

1 Introduction

Verifiable digital credentials including digital certificates, tokens, signed documents and personal credentials are starting to play a vital role in everyday transactions once again [22, 31]. One of the primary benefits of such credential is the ability for an assertion owner to selectively disclose and hide Personally Identifiable Information (PII). However, this contrasts with the position of a Relying Party (RP) (e.g. communication partner, service provider etc.) who often demands additional PII in order to establish the required level of trust thus making anonymity through verifiable digital credentials difficult to achieve in practice.

In order to address this issue in authentication and access control systems, a number of prominent cryptographic solutions have been proposed in the past including Idemix, U-Prove, Privacy-preserving Attribute-Based Credentials [5, 23, 4]. Unfortunately, these technical solutions have not become mainstream due to reasons such as the substantial computational complexity and reliance on a trusted third party [28].

To mitigate some of these issues, there have been significant effort into developing the latest standards and specifications for verifiable digital assertions such as IRMA and Verifiable Credentials [31, 1]. They allow users to perform selective disclosure of the information related to the subject of the assertion (e.g. claim) using Zero Knowledge Proofs as a building block which allow to prove knowledge or a secret without revealing it [33, 27]. On the other hand, these techniques do not allow users to control all the components of digital assertions, which may cause the situations where users are still discriminated by RP based on the differences in *assertion metadata*, for instance [16, 24].

1.1 Scope of the paper and contribution

In this paper, we propose the Decision Making Model for Authentication (DMMA) which determines usage of multiple authentication credentials 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 [16, 12]. We focus on the scenario where every user possesses two different credentials.

The interactions between a user (U), the attribute provider (AP) and the relying party (RP) should be considered in the context of 'RP+AP-U' model for attribute-based authentication defined in [15]. We will further interpret the letter 'U' as a specific user, such as *Alice* or *Bob*. Usually, in these authentication systems, original credentials are supplied by AP with whom a user must first register. A user will then utilize obtained credentials to prepare various proofs (*realizations*) to present to 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 1.1).

Our decision making model 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. 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 [31, 1, 33].

1.2 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 [33, 31]. 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.

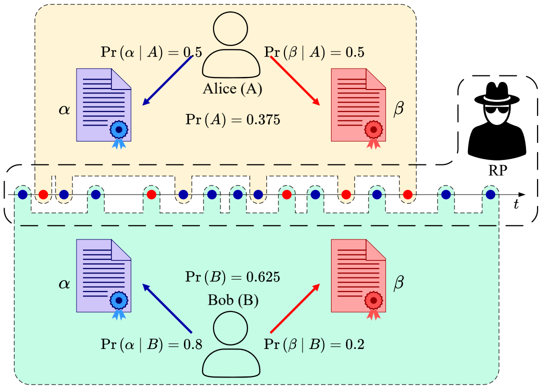

$\Pr(\alpha \mid A) = 0.5$
$\Pr(\beta \mid A) = 0.5$
Alice (A)
$\Pr(A) = 0.375$
α
β
RP
t
$\Pr(B) = 0.625$
Bob (B)
α
β
$\Pr(\alpha \mid B) = 0.8$
$\Pr(\beta \mid B) = 0.2$

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

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 [26].

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 $\Pr\big(A \mid \alpha\big)$, $\Pr\big(A \mid \beta\big)$, $\Pr\big(B \mid \alpha\big)$, $\Pr\big(B \mid \beta\big)$. This in turn depends on the *decisions* made by A and B in the game. For A, these decision includes defining $\Pr\big(\alpha \mid A\big)$, $\Pr\big(\beta \mid A\big)$, $\Pr\big(\alpha \mid A\big) + \Pr\big(\beta \mid A\big) = 1$. For B these decision includes defining $\Pr\big(\alpha \mid B\big)$, $\Pr\big(\beta \mid B\big)$, $\Pr\big(\alpha \mid B\big) + \Pr\big(\beta \mid B\big) = 1$. Using the example of decisions on fig. 1 we conclude that RP infers that $\Pr(\alpha) = \frac{11}{16}$, $\Pr(\beta) = \frac{5}{16}$, $\Pr\big(A \mid \alpha\big) = \frac{3}{11}$, $\Pr\big(A \mid \beta\big) = \frac{3}{5}$, $\Pr\big(B \mid \alpha\big) = \frac{8}{11}$, $\Pr\big(B \mid \beta\big) = \frac{2}{5}$. We further interpret the *privacy meaning* of such decisions based on statistical characteristics available to RP.

1.3 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 system [15, 25]. 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* [15].

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 1.1). 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 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 1.1) remain unclear. As a result, *Alice* and *Bob* require additional concepts to produce *best responses* in the game.

Listing 1.1: 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 $\mathrm{U}_0,\ \mathrm{U}_1$ and policy $\mathrm{P}$ ;
2    AP and RP execute the setup phase (if any) ;
3    AP and $\mathrm{U}_0$ execute the user registration phase (if any) ;
4    AP and $\mathrm{U}_1$ execute the user registration phase (if any) ;
5    RP, AP and $\mathrm{U}_0$ execute the authentication phase ;
6    RP, AP and $\mathrm{U}_\mathrm{b}$ execute the authentication phase, $\mathrm{b} \in \{0, 1\}$, $\Pr\big(\mathrm{b} = 0\big) = 0.5$ ;
7    $\mathfrak{A}$ returns a guess $\mathrm{b}' \in \{0, 1\}$ on the value of $\mathrm{b}$ ;
8    if $\Pr\big(\mathrm{b}' = \mathrm{b}\big) \to 0.5$ return true ;
9    else return false .
```

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 1.1) [15]. This makes it suitable to measure performance in the game such that players can directly derive their utilities from it.

Example 1. *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* $\Pr\big(\alpha \mid B\big) = 0.5$ *and* $\Pr\big(\beta \mid B\big) = 0.5$*. Based on that RP would infer that* $\Pr(\alpha) = 0.5$*,* $\Pr(\beta) = 0.5$*,* $\Pr\big(A \mid \alpha\big) = \Pr\big(A \mid \beta\big) = 0.375$*,* $\Pr\big(B \mid \alpha\big) = \Pr\big(B \mid \beta\big) = 0.625$ *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* $\Pr\big(\alpha \mid A\big) = 1$ *and Bob plays* $\Pr\big(\beta \mid B\big) = 1$*.*

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 you improve unlinkability in attribute-based authentication systems when multiple-credentials are held by a user?

1.4 Roadmap

The paper is structured as follows. Section 2 provides theoretic results that explain why conditional entropy should be used to express RP's 'guessing' performance (as per listing 1.1) followed by the Decision-Making Model for Authentication (DMMA) which is described using 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 3, 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 4. In section 5, we revise the work of other authors that is related to our study. We finally conclude by highlighting the main contributions in section 6.

2 Decision-Making Model for Authentication

In this section, we provide further details pertaining to the Decision Making Model for Authentication (DMMA). For the main notations used in this paper see table 1. 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 1: Main Notations

Notation	**Description**
DMMA	Decision-Making Model 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 $\mathcal{S} \to \mathbb{R}^{\|I\|}$
$\Game = \langle I, \mathcal{T}, \mathcal{S}, \aleph, \varrho, \mathbf{u} \rangle$	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** $[0, 1]^{\|\mathcal{P}_{\mathcal{S}}\|} \to [0, 1]$ over $\vartheta_{\mathcal{S}}$.
$\mathbb{E}[\cdot]$	Expected value
$n \geq 2$	Number of players in the game

2.1 Relation between unlinkability and conditional entropy

Based on the 2-player example provided in the previous section, we label *Alice* and *Bob* using labels $\mathcal{L} = \big\{A, B\big\}$ (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 $\ell = \big\{\alpha, \beta\big\}$ (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 $H\big(\mathtt{L} \mid l\big)$ 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. 2b. 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 bounded by* $H\big(\mathtt{L} \mid l\big)$*.*

Proof. In order to link authentication sessions RP labels them with $\mathtt{L}' \in \mathcal{L}'$, where $\mathcal{L}' = \big\{A', B'\big\}$. 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, $H(\mathtt{L} \mid \mathtt{L}') \geq H(\mathtt{L} \mid l)$.

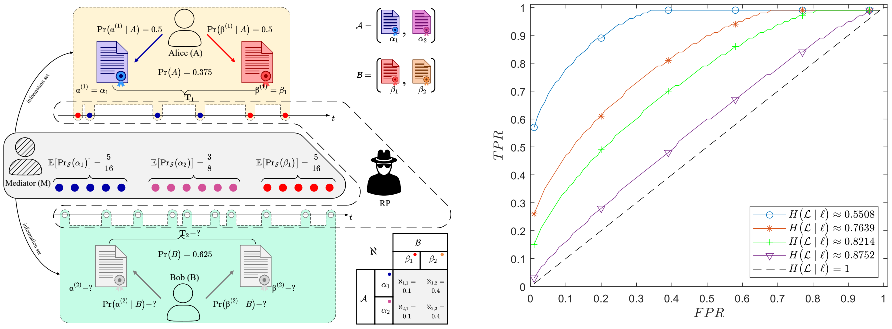
(i) We express linking performance $\mathfrak{P}$ of RP as the difference between True Positive Rate (TPR) and False Positive Rate (FPR): $\mathfrak{P} = \Pr(A' \mid A) - \Pr(A' \mid B) = \frac{\Pr(A', A)}{\Pr(A)} - \frac{\Pr(A', B)}{\Pr(B)}$ which is to be maximized and for which we demand that $\mathfrak{P} \geq 0$. In authentication systems, probability of A, i.e. $\Pr(A)$ and probability of B, i.e. $\Pr\big(B\big)$ 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 $H(\mathtt{L} \mid \mathtt{L}') \geq H(\mathtt{L} \mid l)$. □

As per lemma 1, in order to improve unlinkability at the RP, users A and B could coordinate with one another to increase $H(\mathcal{L} \mid \ell)$. However, questions



$\mathrm{Pr}(\alpha^{(1)} \mid A) = 0.5$
Alice (A)
$\mathrm{Pr}(\beta^{(1)} \mid A) = 0.5$
$\mathrm{Pr}(A) = 0.375$
$\alpha^{(1)} = \alpha_1$
$\beta^{(1)} = \beta_1$
$\mathbf{T}_1$
t
information set
$\mathcal{A} = \left(\alpha_1, \alpha_2\right)$
$\mathcal{B} = \left(\beta_1, \beta_2\right)$
Mediator (M)
$\mathbb{E}[\mathrm{Pr}_{\mathcal{S}}(\alpha_1)] = \frac{5}{16}$
$\mathbb{E}[\mathrm{Pr}_{\mathcal{S}}(\alpha_2)] = \frac{3}{8}$
$\mathbb{E}[\mathrm{Pr}_{\mathcal{S}}(\beta_1)] = \frac{5}{16}$
RP
t
information set
$\mathbf{T}_2 - ?$
$\mathrm{Pr}(B) = 0.625$
Bob (B)
$\alpha^{(2)} - ?$
$\beta^{(2)} - ?$
$\mathrm{Pr}(\alpha^{(2)} \mid B) - ?$
$\mathrm{Pr}(\beta^{(2)} \mid B) - ?$
$\aleph$
$\mathcal{B}$
β_1
β_2
$\mathcal{A}$
α_1
α_2
$\aleph_{1,1} = 0.1$
$\aleph_{1,2} = 0.4$
$\aleph_{2,1} = 0.1$
$\aleph_{2,2} = 0.4$
TPR
FPR
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
$H(\mathcal{L} \mid \ell) \approx 0.5508$
$H(\mathcal{L} \mid \ell) \approx 0.7639$
$H(\mathcal{L} \mid \ell) \approx 0.8214$
$H(\mathcal{L} \mid \ell) \approx 0.8752$
$H(\mathcal{L} \mid \ell) = 1$

(a) Mediated decision made by *Alice* in refined model (incomplete information).

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

Fig. 2: Linking in attribute based authentication system with 2 users.

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 $\ell = \left\{\alpha, \beta\right\}$; and (ii) the information $\mathrm{Pr}\left(\alpha \mid B\right)$ and $\mathrm{Pr}\left(\beta \mid B\right)$ is known to *Alice* and *Bob*, respectively. We will further elaborate on this issue by analyzing a refined model.

2.2 Refinement of 2-player model for the game

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 [16]. 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 it distinguishable as α_1 and α_2 when issued by different states. This difference dictates the need to refine the model introduced previously.

We therefore acknowledge that extended set ℓ of attributes is used on practice where each user may possess only a subset within ℓ. There are multiple use cases that support such construct.

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 1 corresponds to *Alice* and 2 corresponds to *Bob*. We then demand that *Alice* and *Bob* possess one realization from each category. Random variables $\upalpha^{(1)}, \upalpha^{(2)}$ for *Alice* and *Bob*, respectively, take realizations from $\mathcal{A}$. In a similar way $\upbeta^{(1)}, \upbeta^{(2)}$ take realizations from $\mathcal{B}$. 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.

Bayesian game [4] 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 Bayesian game $\Game_B = \langle I, \mathcal{T}, \mathcal{S}, \aleph, \varrho, \mathbf{u} \rangle$. 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* 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*. This is reflected in the following assumption.

Assumption 1. *Neither exact type of player i nor the number of attributes that she uses is known to $-i$.*

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 $\Pr\big(\alpha^{(i)} = \alpha_\iota, \beta^{(i)} = \beta_\rho\big) = \aleph_{\iota,\rho}$, and $i \in I$; $\iota \in \big\{1, ..., |\mathcal{A}|\big\}$, $\rho \in \big\{1, ..., |\mathcal{B}|\big\}$ (see $\aleph$ on fig. 2a). We use $\mathcal{S} = \{s_1, s_2\}$, $s_1 = \Pr\big(\alpha^{(1)} \mid A\big)$ and $s_2 = \Pr\big(\alpha^{(2)} \mid B\big)$ 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* $\mathop{\mathbb{E}}\limits_{\aleph,\varrho_{\mathcal{T}_{-i}}}\big[u_1\big]$.

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

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*.

[4] For the main notations see table 1

Mediated game In a mediated game synchronization can be achieved if information that is *sufficient* for calculation of *best response* is provided to the players in contrast to a Bayesian game where this information is provided in the form of beliefs (priors) ϱ_{-i}.

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, $\Pr_{\mathcal{S}}\big(\alpha^{(1)}\big)$ 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
$\vartheta_{\mathcal{S}} = \Big(\Pr_{\mathcal{S}}\big(\alpha^{(1)}\big), \Pr_{\mathcal{S}}\big(\alpha^{(2)}\big), \Pr_{\mathcal{S}}\big(\beta^{(1)}\big), \Pr_{\mathcal{S}}\big(\beta^{(2)}\big)\Big)$ of marginal probabilities at RP, $\vartheta_{\mathcal{S}} \in [0,1]^{|\mathcal{P}_{\mathcal{S}}|}$. These beliefs are captured by joint **pdf** $\varphi : [0,1]^{|\mathcal{P}_{\mathcal{S}}|} \to [0,1]$.

(3) We then introduce *Mediator* (M for short) who provides information about these beliefs to the players (e.g. synchronizes φ among them).

Assumption 2. *Priors over $\vartheta_{\mathcal{S}}$ are known to the players.*

This information, for instance, can also be in the form of compact statistical characteristic, such as expectation $\underset{\varphi}{\mathbb{E}}[\vartheta_{\mathcal{S}}]$ (see fig. 2a). 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.

2.3 Utility and best responses in the game with $n \geq 2$ players

We will further analyze game with mediator $\Game_M = \langle I, \mathcal{T}, \mathcal{S}, \aleph, \varrho, \varphi, \mathbf{u}\rangle$ where $n = |I|$, $n \geq 2$. We will also generalize our results for extended attribute categories $\mathcal{A} = \left\{\alpha_1, ..., \alpha_\iota, ..., \alpha_l\right\}$ and $\mathcal{B} = \left\{\beta_1, ..., \beta_\rho, ..., \beta_m\right\}$.

Definition 5 (DMMA)**.** *Decision Making Model for Authentication enables user i to increase unlinkability by maximising u_i in $\Game_M$ based on values s_i and $1 - s_i$.*

Defining expressions for expected utility and best responses requires specifying conditional entropy $H\big(\mathtt{L} \mid l\big)$ that the attacker has in regards to linking the users. This expression for entropy depends on what is known to RP who registers authentication events. The next corollary follows directly from definition 2 and helps us to specify $H\big(\mathtt{L} \mid l\big)$.

Corollary 1. *The set of players' types $\mathcal{T}$ is known to RP.*

According to conditional entropy under corollary 1, collective unlinkability of the entire authentication system with n users is:

$$\mathcal{C} = H\big(\mathtt{L} \mid l\big) = -\sum_{i=1}^{n} \Pr(i)\bigg(\Pr\big(\alpha^{(i)} \mid i\big)\log\frac{\Pr\big(\alpha^{(i)} \mid i\big)}{\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)}\Pr(i)$$
$$+\Pr\big(\beta^{(i)} \mid i\big)\log\frac{\Pr\big(\beta^{(i)} \mid i\big)}{\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)}\Pr(i)\bigg) = \sum_{i=1}^{n}\Pr(i)\log\frac{1}{\Pr(i)} \tag{1}$$
$$-\sum_{i=1}^{n}\Pr(i)\bigg(\Pr\big(\alpha^{(i)} \mid i\big)\log\frac{\Pr\big(\alpha^{(i)} \mid i\big)}{\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)} + \Pr\big(\beta^{(i)} \mid i\big)\log\frac{\Pr\big(\beta^{(i)} \mid i\big)}{\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)}\bigg)\,,$$

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

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

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

Assumption 3. *Priors over $\vartheta_{\mathcal{S}}$ satisfy the following scaling constraint for all $i \in I$:*

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

Based on the corollary, for instance, expected utilities $\mathbb{E}\big[u_i\big]$ of the players derived from $H\big(\mathtt{L} \mid l\big)$ under assumption that for all i, where $1 \le i \le n$ holds $\Pr(i) = \frac{1}{n}$ are provided in the following Lemma:

Lemma 2. *Expected utility for player i is*

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

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

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

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

$$\mathbb{E}\Big[s_i\log\tfrac{s_i}{\Pr_{\mathcal{S}}(\alpha^{(i)})}\Big] = \mathbb{E}\big[s_i\log s_i - s_i\log\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big] = s_i\log s_i - s_i\mathbb{E}\big[\log\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big]\,.$$

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

$$\begin{aligned}\log\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big) \;\approx\;& \log\mathbb{E}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big]\\ +\tfrac{\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)-\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]}{\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]} -& \tfrac{\left(\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)-\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]\right)^2}{2\mathbb{E}\left[\Pr_{\mathcal{S}}\left(\alpha^{(i)}\right)\right]^2}\,.\end{aligned}$$

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

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

and, we obtain that

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

In a similar way, we arrive at

$$\mathbb{E}\Big[(1-s_i) \log \tfrac{1-s_i}{\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)}\Big] \approx (1-s_i) \log \tfrac{1-s_i}{\mathbb{E}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big]} + (1-s_i) \tfrac{\operatorname{Var}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big]}{2\mathbb{E}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big]^2} \ .$$

Lastly, by considering assumption 3 we obtain that $\mathbb{E}\left[u_i\right]$ equals to

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

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

According eqs. (2) and (3), expected unlinkability $\mathbb{E}\left[\mathcal{C}\right]$, of the whole system is:

$$\mathbb{E}\left[\mathcal{C}\right] \approx \log n + \sum_i \mathbb{E}\left[u_i\right] \ . \tag{4}$$

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

Corollary 2. *For* ***continuous*** *strategies* ***best response of player*** *i is defined as*

$$s_i^{\mathrm{b}} = \frac{\mathbb{E}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big]}{\mathbb{E}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big] + \mathbb{E}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big]} \ , \tag{5}$$

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

$$s_i^{\mathrm{b}} = \begin{cases} 1 & \text{if } \mathbb{E}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big] > \mathbb{E}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big] \ ; \\ 0 & \text{if } \mathbb{E}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big] < \mathbb{E}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big] \ , \end{cases} \tag{6}$$

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

The proof for corollary 2 is straightforward and we omit it here.

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

$$\forall i, j\big(\mathbf{T}_j = \mathbf{T}_i\big) \implies \mathbb{E}\big[u_j^{\mathrm{b}}\big] = \mathbb{E}\big[u_i^{\mathrm{b}}\big] = \log\Big(\mathbb{E}\big[\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(i)}\big)\big] + \mathbb{E}\big[\mathrm{Pr}_{\mathcal{S}}\big(\beta^{(i)}\big)\big]\Big)\ . \tag{7}$$

The result of remark 1 follows directly from how we define $\vartheta_{\mathcal{S}}$ and corollary 2. Also, it is a common game-theoretical premise, but we have arrived at it by first accepting assumption 2.

Best response can be, for instance, illustrated using diagram on fig. 2a. It is clear that type $\mathbf{T}_1$ of *Alice* has realization $\{\alpha_1, \beta_1\}$ while $\mathop{\mathbb{E}}_{\varphi}[\vartheta_{\mathcal{S}}] = \big(\frac{5}{16}, \frac{3}{8}, \frac{5}{16}, 0\big)$. Hence, her best response calculated in accordance to eq. (5) 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.

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 [13]. 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. 2a). We admit that

$$\begin{aligned}&\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(1)}\big) = \mathrm{Pr}\big(\alpha^{(1)}, A\big) + \mathrm{Pr}\big(\alpha^{(1)}, B\big) = \\ &\mathrm{Pr}\big(A\big)\mathrm{Pr}\big(\alpha^{(1)} \mid A\big) + \mathrm{Pr}\big(B\big)\mathrm{Pr}\big(\alpha^{(1)} = \alpha^{(2)}\big)\mathrm{Pr}\big(\alpha^{(2)} \mid B\big)\ ,\end{aligned} \tag{8}$$

from which we obtain

$$\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(1)}\big)\big] = \mathrm{Pr}\big(A\big)s_1 + \mathrm{Pr}\big(B\big)\Big(\aleph_{1,1}\mathop{\mathbb{E}}_{\varrho_{1,1}}\big[s_2\big] + \aleph_{1,2}\mathop{\mathbb{E}}_{\varrho_{1,2}}\big[s_2\big]\Big)\ , \tag{9}$$

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

$$\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\beta^{(1)}\big)\big] = \mathrm{Pr}\big(A\big)(1 - s_1) + \mathrm{Pr}\big(B\big)\Bigg(\aleph_{1,1}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{1,1}}\big[s_2\big]\Big) + \aleph_{2,1}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{2,1}}\big[s_2\big]\Big)\Bigg)\ , \tag{10}$$

$$\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(2)}\big)\big] = \mathrm{Pr}\big(B\big)\Big(\aleph_{2,1}\mathop{\mathbb{E}}_{\varrho_{2,1}}\big[s_2\big] + \aleph_{2,2}\mathop{\mathbb{E}}_{\varrho_{2,2}}\big[s_2\big]\Big)\ , \tag{11}$$

$$\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\beta^{(2)}\big)\big] = \mathrm{Pr}\big(B\big)\Bigg(\aleph_{1,2}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{1,2}}\big[s_2\big]\Big) + \aleph_{2,2}\Big(1 - \mathop{\mathbb{E}}_{\varrho_{2,2}}\big[s_2\big]\Big)\Bigg)\ . \tag{12}$$

Example 2. *Notably, M on fig. 2a 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* $\underset{\varphi}{\mathbb{E}}[\vartheta_{\mathcal{S}}]$ *is* $\underset{\varphi}{\mathbb{E}}\big[\mathrm{Pr}_{\mathcal{S}}\big(\beta^{(2)}\big)\big] = 0$ *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* $\underset{\varrho_{2,2}}{\mathbb{E}}\big[s_2\big] = 1$ *we obtain from eq.* (11)

$$\underset{\varrho_{2,1}}{\mathbb{E}}\big[s_2\big] = \frac{1}{\aleph_{2,1}}\left(\frac{\underset{\varphi}{\mathbb{E}}\big[\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(2)}\big)\big]}{\mathrm{Pr}\big(B\big)} - \aleph_{2,2}\right) = 2 \,, \tag{13}$$

which is impossible because $0 \leq s_2 \leq 1$.

2.4 Equilibria for mediated game with $n \geq 2$ players

With the aim to design coordination mechanism where M is consistent we assign to $\underset{\varphi}{\mathbb{E}}[\vartheta_{\mathcal{S}}]$ values from corresponding complete information Nash equilibria [18].

Definition 6 (Consistent expectations)**.** *Best responses* $\mathcal{S}^{\mathrm{b}}$ *of all* n *players must satisfy for all* $i \in I$*:*

$$\Big(\mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\alpha^{(i)}\big) = \mathbb{E}\big[\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(i)}\big)\big]\Big) \wedge \Big(\mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\beta^{(i)}\big) = \mathbb{E}\big[\mathrm{Pr}_{\mathcal{S}}\big(\beta^{(i)}\big)\big]\Big). \tag{14}$$

One way to enable eq. (14) it is to find complete information Nash equilibrium in the form, $\forall i \in I$:

$$\begin{cases} s_i^{\mathrm{b}} = \dfrac{\mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\alpha^{(i)}\big)}{\mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\alpha^{(i)}\big)+\mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\beta^{(i)}\big)} \,, \\ \mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\alpha^{(i)}\big) = \sum\limits_{j=1}^{n} \mathrm{Pr}\big(\alpha^{(i)} = \alpha^{(j)}\big)\mathrm{Pr}(j)s_j^{\mathrm{b}} \,, \\ \mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\beta^{(i)}\big) = \sum\limits_{j=1}^{n} \mathrm{Pr}\big(\beta^{(i)} = \beta^{(j)}\big)\mathrm{Pr}(j)(1-s_j^{\mathrm{b}}) \,. \end{cases} \tag{15}$$

Mediator M then forms priors on $\vartheta_{\mathcal{S}}$ accordingly. We will next formulate eq. (15) 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 $\mathbf{\Omega}_{\alpha_\iota}$ and $\mathbf{\Omega}_{\beta_\rho}$ such that $\forall \phi, \xi \in I$

$$\big(\alpha^{(\xi)} = \alpha_\iota\big) \iff \big(\xi \in \mathbf{\Omega}_{\alpha_\iota}\big) \,, \; \big(\beta^{(\phi)} = \beta_\rho\big) \iff \big(\phi \in \mathbf{\Omega}_{\beta_\rho}\big) \,.$$

We denote $\mathbf{\Omega}_{\alpha_\iota,\beta_\rho} = \mathbf{\Omega}_{\alpha_\iota} \cap \mathbf{\Omega}_{\beta_\rho}$ for which holds: $\mathbf{\Omega}_{\alpha_\iota} = \bigcup\limits_{\rho=1}^{m} \mathbf{\Omega}_{\alpha_\iota,\beta_\rho}$ and $\mathbf{\Omega}_{\beta_\rho} = \bigcup\limits_{\iota=1}^{l} \mathbf{\Omega}_{\alpha_\iota,\beta_\rho}$. For $i \in \mathbf{\Omega}_{\alpha_\iota,\beta_\rho}$ we set $\mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\alpha^{(i)}\big) = \frac{1}{n}\sum\limits_{\xi \in \mathbf{\Omega}_{\alpha_\iota}} s_\xi^{\mathrm{b}}$ and $\mathrm{Pr}_{\mathcal{S}^{\mathrm{b}}}\big(\beta^{(i)}\big) = \frac{1}{n}\sum\limits_{\phi \in \mathbf{\Omega}_{\beta_\rho}} \big(1 - s_\phi^{\mathrm{b}}\big)$. In line with remark 1 all players whose indices are in $\mathbf{\Omega}_{\alpha_\iota,\beta_\rho}$

produce the same best response denoted as $\theta_{\iota,\rho}$, and, as a result $\Pr_{\mathcal{S}^{\mathrm{b}}}\big(\alpha^{(i)}\big) = \frac{1}{n}\sum_{\nu=1}^{m}|\Omega_{\alpha_\iota,\beta_\nu}|\theta_{\iota,\nu}$, while $\Pr_{\mathcal{S}^{\mathrm{b}}}\big(\beta^{(i)}\big) = \frac{1}{n}\sum_{\tau=1}^{l}|\Omega_{\alpha_\tau,\beta_\rho}|\big(1-\theta_{\tau,\rho}\big)$. Without loss of generality, for large $n >> l \times m$, we have $\aleph_{\iota,\rho} = \frac{1}{n}|\boldsymbol{\Omega}_{\alpha_\iota,\beta_\rho}|$. Validity of eq. (15) is guaranteed if for all ι, ρ:

$$\theta_{\iota,\rho} = \frac{\sum_{\nu=1}^{m}\aleph_{\iota,\nu}\theta_{\iota,\nu}}{\sum_{\nu=1}^{m}\aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum_{\tau=1}^{l}\aleph_{\tau,\rho}(1-\theta_{\tau,\rho})} \ . \tag{16}$$

Next, we will discuss the solutions for eq. (16) in pure continuous (e.g. authentication with 2 attributes, $\forall\iota,\rho\big(\theta_{\iota,\rho}\in[0,1]\big)$) and pure discrete (e.g. authentication with 1 attribute, $\forall\iota,\rho\big(\theta_{\iota,\rho}\in\{0,1\}\big)$) strategies. In addition, we will consider the maximin scenario for the case when assumption 2 does not hold.

Players use 2 attributes interchangeably In order to provide consistent priors to the players, mediator M solves eq. (16) and takes 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 complete information Nash equilibria in (pure) continuous strategies are then represented by the following system of quadratic equations:

$$\begin{cases} \theta_{1,1}\Big(\sum\limits_{\nu=2}^{m}\aleph_{1,\nu}\theta_{1,\nu} + \sum\limits_{\tau=2}^{l}\aleph_{\tau,1}(1-\theta_{\tau,1})\Big) = \sum\limits_{\nu=2}^{m}\aleph_{1,\nu}\theta_{1,\nu} \ , \\ \quad\vdots \\ \theta_{\iota,\rho}\Big(\sum\limits_{\substack{\nu=1\\ \nu\neq\rho}}^{m}\aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum\limits_{\substack{\tau=1\\ \tau\neq\iota}}^{l}\aleph_{\tau,\rho}(1-\theta_{\tau,\rho})\Big) = \sum\limits_{\substack{\nu=1\\ \nu\neq\rho}}^{m}\aleph_{\iota,\nu}\theta_{\iota,\nu} \ , \\ \quad\vdots \\ \theta_{l,m}\Big(\sum\limits_{\nu=1}^{m-1}\aleph_{l,\nu}\theta_{l,\nu} + \sum\limits_{\tau=1}^{l-1}\aleph_{\tau,m}(1-\theta_{\tau,m})\Big) = \sum\limits_{\nu=1}^{m-1}\aleph_{l,\nu}\theta_{l,\nu} \ . \end{cases} \tag{17}$$

Players use 1 attribute In these settings, as a result of applying constraint $\theta_{\iota,\rho}\in\{0,1\}$, each player only plays discrete pure strategy. If all players with the same $\big(\iota,\rho\big)$ produce same best response, 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 $\big(\iota,\rho\big)$ play different pure discrete strategies. This is only possible if players of type $\big(\iota,\rho\big)$ are indifferent as to which among 2 strategies to play (see corollary 2). Taking into account all possible attribute realizations, this latter condition is represented by the following linear system:

$$\begin{cases} \sum_{\nu=1}^{m} \aleph_{1,\nu}\bar{\theta}_{1,\nu} & = \displaystyle\sum_{\tau=1}^{l} \aleph_{\tau,1}(1-\bar{\theta}_{\tau,1})\ , \\ & \vdots \\ \sum_{\nu=1}^{m} \aleph_{\iota,\nu}\bar{\theta}_{\iota,\nu} & = \displaystyle\sum_{\tau=1}^{l} \aleph_{\tau,\rho}(1-\bar{\theta}_{\tau,\rho})\ , \\ & \vdots \\ \sum_{\nu=1}^{m} \aleph_{l,\nu}\bar{\theta}_{l,\nu} & = \displaystyle\sum_{\tau=1}^{l} \aleph_{\tau,m}(1-\bar{\theta}_{\tau,m})\ . \end{cases} \tag{18}$$

Equilibria established for eq. (16) are necessary, but not sufficient for truthfulness. This implies that players must *trust M*, and for that reason we further dub mediated game '*Naïve game*'. The issue of trust can be addressed if maximin principle is applied.

2.5 Equilibria for maximin game with $n \geq 2$ players

Here we presume that player i makes decision under uncertainty about the priors φ on $\vartheta_{\mathcal{S}}$. One way to address this uncertainty is to apply Wald maximin principle requiring i to consider the worst-case scenario played by $-i$ [29]. Expected utility of i is then

$$\mathbb{E}_w[u_i] = \max_{s_i} \min_{\mathcal{S}_{-i}} \mathop{\mathbb{E}}_{\aleph,\varphi} [u_i]\ , \tag{19}$$

which can be ruminated as a special case of the 'naïve' game where i calculates her best response using corollary 2. Instead of $\mathbb{E}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big)\big]$ and $\mathbb{E}\big[\Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big]$ she substitutes worst possible estimates $\mathbb{E}\big[\Pr_{\mathcal{S}^w}\big(\alpha^{(i)}\big)\big]$ and $\mathbb{E}\big[\Pr_{\mathcal{S}^w}\big(\beta^{(i)}\big)\big]$, respectively. We denote worst case decisions of all $-i$ players using $\mathcal{S}^w_{-i}$ and stress that $\mathcal{S}^w_{-i}$ unambiguously defines φ. We then use eq. (7) to obtain:

$$\mathcal{S}^w_{-i} = \arg\min_{\mathcal{S}_{-i}} \mathop{\mathbb{E}}_{\aleph}\big[u_i^{\mathrm{b}}\big] \sim \arg\min_{\mathcal{S}_{-i}} \mathop{\mathbb{E}}_{\aleph}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big) + \Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big]. \tag{20}$$

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

$$\begin{aligned} \mathop{\mathbb{E}}_{\aleph}\big[\Pr_{\mathcal{S}}\big(\alpha^{(i)}\big) + \Pr_{\mathcal{S}}\big(\beta^{(i)}\big)\big] &= \mathop{\mathbb{E}}_{\aleph}\big[\aleph_{\iota,\rho} + \sum_{\substack{\nu=1 \\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum_{\substack{\tau=1 \\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}(1-\theta_{\tau,\rho})\big] \\ = \sum_{\iota=1}^{l}\sum_{\rho=1}^{m} \Big(\aleph^2_{\iota,\rho} + \aleph_{\iota,\rho} \sum_{\substack{\nu=1 \\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} &+ \aleph_{\iota,\rho} \sum_{\substack{\tau=1 \\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}\big(1-\theta_{\tau,\rho}\big)\Big)\ . \end{aligned} \tag{21}$$

In the last line of eq. (21) we ignore $\aleph^2_{\iota,\rho}$ for the calculation of

$$\arg\min_{\boldsymbol{\theta}} \sum_{\iota=1}^{l} \sum_{\rho=1}^{m} \aleph_{\iota,\rho} \Big(\sum_{\substack{\nu=1 \\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu}\theta_{\iota,\nu} + \sum_{\substack{\tau=1 \\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho}(1-\theta_{\tau,\rho}) \Big) \ , \tag{22}$$

from which we conclude that for all ι, ρ:

$$\theta_{\iota,\rho} = \begin{cases} 0, \text{ if } \sum_{\substack{\nu=1 \\ \nu\neq\rho}}^{m} \aleph_{\iota,\nu} \geq \sum_{\substack{\tau=1 \\ \tau\neq\iota}}^{l} \aleph_{\tau,\rho} \ ; \\ 1, \qquad\qquad\qquad \text{otherwise} \ . \end{cases} \tag{23}$$

Example 3. *Maximin game can then be considered a special case of mediated game on fig. 2a 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.* (9) *and* (10) *she would then obtain*

$$\begin{aligned} &\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(1)}\big)\big] = \mathrm{Pr}\big(A\big)s_1 + \mathrm{Pr}\big(B\big)\aleph_{1,2} \\ &\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\beta^{(1)}\big)\big] = \mathrm{Pr}\big(A\big)(1-s_1) + \mathrm{Pr}\big(B\big)\big(\aleph_{1,1} + \aleph_{2,1}\big) \ , \end{aligned} \tag{24}$$

which is substituted into eq. (5) *to obtain*

$$s_1^{\mathrm{b}} = \frac{\mathrm{Pr}\big(A\big)s_1^{\mathrm{b}} + \mathrm{Pr}\big(B\big)\aleph_{1,2}}{\mathrm{Pr}\big(A\big) + \mathrm{Pr}\big(B\big)\big(\aleph_{1,1} + \aleph_{1,2} + \aleph_{2,1}\big)} = \frac{\aleph_{1,2}}{\aleph_{1,1} + \aleph_{1,2} + \aleph_{2,1}} = \frac{2}{3} \ , \tag{25}$$

that maximizes her expected utility from eq. (23)*. This result means that over the period of time observable in fig. 2a 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'.*

3 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.

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 [6]. It is remarkable that (according to eqs. (3) to (5)) $\mathrm{Pr}(i)$, $\aleph$, $\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\alpha^{(i)}\big)\big]$, $\mathop{\mathbb{E}}_{\varphi}\big[\mathrm{Pr}_{\mathcal{S}}\big(\beta^{(i)}\big)\big]$ 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. (3) we derive best response expressions that are identical among players i whose types $\mathbf{T}_i = \big\{\alpha^{(i)}, \beta^{(i)}\big\}$ 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.

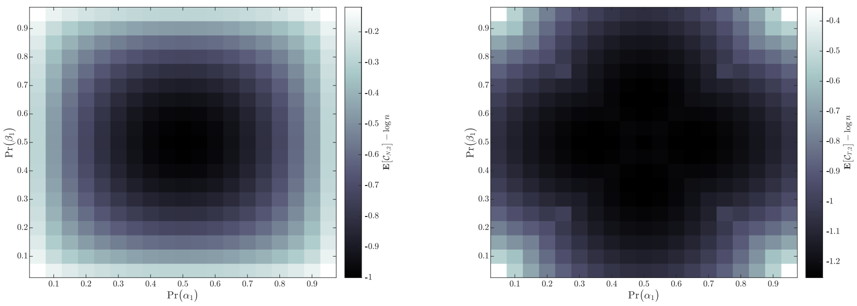
3.1 Experiment organization

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 2.4 and 2.5). 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 eq. (4)) 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 $\mathbb{E}\big[\mathcal{C}_{\{\kappa,\mu\}^r}\big]$ 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 3-5). 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..

3.2 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. 3). For each possible $\aleph$ in naïve game we solved complete information Nash equilibria (see eq. (17)) to find $\underset{\varphi}{\mathbb{E}}[\vartheta_{\mathcal{S}}]$ 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 (see eq. (23)). Then, best response of i, and $\mathbb{E}[\mathcal{C}_{T,2}]$ are calculated (see eq. (5)). As can be observed from comparison of fig. 3a and fig. 3b naïve game provides substantially better unlinkability.

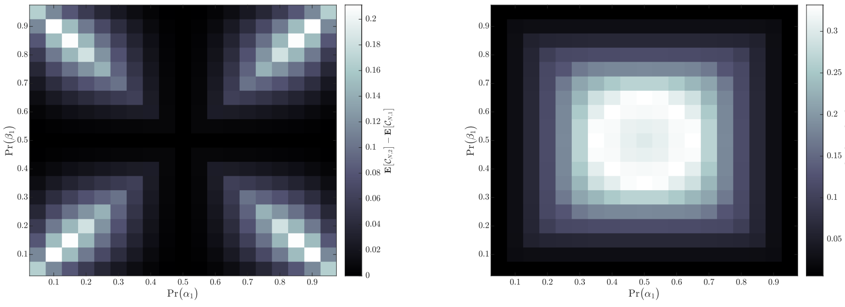


(a) Naïve game, players use 2 attributes interchangeably, e.g. $\theta_{\iota,\rho} \in [0, 1]$.

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

Fig. 3: 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 eq. (18)). The benefits of using 2 attributes (unconstrained rationality) versus 1 attribute (constrained rationality) can be observed by comparing residual unlinkabilities on fig. 4 which are greater than 0 for the both heatmaps.

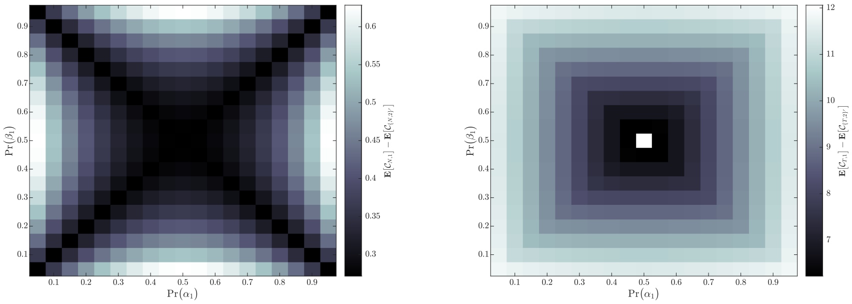


(a) Difference between expected unlinkability $\mathbb{E}\left[\mathcal{C}_{N,2}\right]$ and $\mathbb{E}\left[\mathcal{C}_{N,1}\right]$.

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

Fig. 4: Residual expected unlinkability in 'Naïve game' and 'Tenable' games.

We conducted a range of experiments with randomized moves which results are presented on fig. 5. 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).



(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 $\mathbb{E}\left[\mathcal{C}_{\{T,2\}^r}\right]$.

Fig. 5: Residual expected unlinkability in 'Naïve' and 'Tenable' games.

4 Discussion

The cross-comparison of results from section 3 demonstrates how proposed DMMA impacts the rate of user *unlinkability* 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 3 we differentiate between *rational* and *alternative* principles of usage (see figs. 4 and 5). 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 [7]. We contemplate that expectation $\underset{\varphi}{\mathbb{E}}[\vartheta_{\mathcal{S}}]$ calculated using priors φ over the vector $\vartheta_{\mathcal{S}}$ 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. (5)). Second, this may be shared with the players in differentially private form [2, 18]. 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. 3) [29]. We, nevertheless, do not provide recommendation as to which among the naïve and tenable game scenario to chose for ABA systems. This is because these different decision making concepts require various levels of *trust*: 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. 3a and 3b. 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.

5 Related Work

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 [17]. 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 [15, 25]. 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 [14, 21, 20]. One such test is ISO/IEC DIS 27551 "Requirements for attribute-based unlinkable entity authentication" as per listing 1.1, 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 [3, 8]. For example, a basic information-theoretic notion for unlinkability is presented by [30] 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 [9] 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 [32], where they propose an extensive taxonomy of privacy metrics which is classified by output and, for instance, describes 17 different entropy-based measures.

Game Theory Applications to Privacy A number of papers apply game theory to address privacy issues either based on problems derived from practice [19, 10, 14], or focusing on the theoretical aspects of game theory [11, 18].

From a practical approach, there are several studies that explored the challenges pertaining to pseudonym change in mobile networks were investigated by the authors of [10, 14]. In [10], 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 [14] 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, studies have examined the coordination scenarios which impact privacy in general [11, 18]. 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.

6 Conclusion

As per highlighted in this paper, the privacy dilemma of interchangeable usage of multiple assertions possessed by a user is something that can cause serious privacy issues. While substantial efforts of the research community have been directed toward making assertions indistinguishable, the question around '*How to best use these assertions if indistinguishability fails with non-zero probability?*' have been largely ignored. Therefore, we recommended a rational decision-making approach as a means to address this question.

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 including 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) to ensure that the issues pertaining to privacy highlighted in this study are mitigated [27].

References

1. Alpár, G., Broek, F. van den, Hampiholi, B., Jacobs, B., Lueks, W., and Ringers, S.: IRMA: practical, decentralized and privacy-friendly identity management using smartphones. In: 10th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2017) (2017)
2. Bergemann, D., and Morris, S.: Bayes correlated equilibrium and the comparison of information structures in games. Theoretical Economics 11(2), 487–522 (2016)
3. Berman, R., Fiat, A., and Ta-Shma, A.: Provable unlinkability against traffic analysis. In: International Conference on Financial Cryptography, pp. 266–280 (2004)
4. Camenisch, J., Krenn, S., Mikkelsen, A.L.G., Neven, G., and Pedersen, M.: D3. 1: Scientific Comparison of ABC Protocols. Part I-Formal Treatment of Privacy-Enhancing Credential Systems. Project deliverable in ABC4Trust (2014)
5. Camenisch, J., and Van Herreweghen, E.: Design and implementation of the idemix anonymous credential system. In: Proceedings of the 9th ACM conference on Computer and communications security, pp. 21–30 (2002)
6. Coleman, T.F., and Li, Y.: An interior trust region approach for nonlinear minimization subject to bounds. SIAM Journal on optimization 6(2), 418–445 (1996)
7. Cooper, R., DeJong, D.V., Forsythe, R., and Ross, T.W.: Communication in Coordination Games. The Quarterly Journal of Economics 107(2), 739–771 (1992)
8. Dwork, C., and Nissim, K.: Privacy-Preserving Datamining on Vertically Partitioned Databases. In: 24th Annual International Cryptology Conference (CRYPTO 2004). LNCS, vol. 3152, pp. 528–544. Springer, Heidelberg (2004)
9. Franz, M., Meyer, B., and Pashalidis, A.: Attacking unlinkability: The importance of context. In: International Workshop on Privacy Enhancing Technologies, pp. 1–16 (2007)
10. Freudiger, J., Manshaei, M.H., Hubaux, J.-P., and Parkes, D.C.: On non-cooperative location privacy: a game-theoretic analysis. In: Proceedings of the 16th ACM conference on Computer and communications security, pp. 324–337 (2009)
11. Ghosh, A., and Ligett, K.: Privacy as a coordination game. In: 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1608–1615 (2013)
12. Grassi, P., Fenton, J., Lefkovitz, N., Danker, J., Choong, Y.-Y., Greene, K., and Theofanos, M.: Digital Identity Guidelines: Enrollment and Identity Proofing. Tech. rep., National Institute of Standards and Technology (2017)
13. Harsanyi, J.C.: Games with Incomplete Information Played by "Bayesian" Players, I–III Part I. The Basic Model. Management Science 14(3), 159–182 (1967)
14. Humbert, M., Manshaei, M.H., Freudiger, J., and Hubaux, J.-P.: Tracking games in mobile networks. In: International Conference on Decision and Game Theory for Security, pp. 38–57 (2010)

15. ISO Central Secretary: Information technology – Requirements for attribute-based unlinkable entity authentication. en. Standard ISO/IEC DIS 27551, Geneva, CH: International Organization for Standardization (2020)
16. Karegar, F., Striecks, C., Krenn, S., Hörandner, F., Lorünser, T., and Fischer-Hübner, S.: Opportunities and Challenges of CREDENTIAL. In: IFIP International Summer School on Privacy and Identity Management, pp. 76–91 (2016)
17. Katos, V.: Managing IS Security and Privacy. In: Cyber Crime: Concepts, Methodologies, Tools and Applications, pp. 1246–1254. IGI Global(2012)
18. Kearns, M., Pai, M., Roth, A., and Ullman, J.: Mechanism design in large games: Incentives and privacy. In: Proceedings of the 5th conference on Innovations in theoretical computer science, pp. 403–410 (2014)
19. Liu, X., Liu, K., Guo, L., Li, X., and Fang, Y.: A game-theoretic approach for achieving k-anonymity in location based services. In: 2013 Proceedings IEEE INFOCOM, pp. 2985–2993 (2013)
20. Maronna, R., Martin, D., and Yohai, V.: Robust Statistics: Theory and Methods. Wiley (2006)
21. Neubauer, M.: Modelling of pseudonymity under probabilistic linkability attacks. In: 2009 International Conference on Computational Science and Engineering, pp. 160–167 (2009)
22. Ontario, G. of: Ontario's Verifiable Businesses. https://www.von.gov.on.ca/en/home (2020)
23. Paquin, C.: U-prove technology overview v1. 1. Microsoft Corporation Draft Revision 1 (2011)
24. Pashalidis, A., and Mitchell, C.J.: Limits to anonymity when using credentials. In: International Workshop on Security Protocols, pp. 4–12 (2004)
25. Pfitzmann, A., and Hansen, M.: A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. (2010)
26. Preibusch, S., Kübler, D., and Beresford, A.R.: Price versus privacy: an experiment into the competitive advantage of collecting less personal information. Electronic Commerce Research 13(4), 423–455 (2013)
27. Quintyne-Collins, M., Vescent, H., O'Donnell, D., Slepak, G., Brown, M., Allen, C., and Ruther, M.: Digital Credential Wallets
28. Sabouri, A.: Understanding the determinants of privacy-ABC technologies adoption by service providers. In: Conference on e-Business, e-Services and e-Society, pp. 119–132 (2015)
29. Sniedovich, M.: Wald's mighty maximin: a tutorial. ITOR 23(4), 625–653 (2016)
30. Steinbrecher, S., and Köpsell, S.: Modelling unlinkability. In: International Workshop on Privacy Enhancing Technologies, pp. 32–47 (2003)
31. W3C: Verifiable Credentials Data Model v1.0. https://www.w3.org/TR/vc-data-model/ (2020)
32. Wagner, I., and Eckhoff, D.: Technical privacy metrics: a systematic survey. ACM Computing Surveys (CSUR) 51(3), 1–38 (2018)
33. Zhang, Z., Król, M., Sonnino, A., Zhang, L., and Rivière, E.: EL PASSO: Privacy-preserving, Asynchronous Single Sign-On. arXiv preprint arXiv:2002.10289 (2020)