EURASIP Journal on
Information Security

RESEARCH Open Access

Metadata filtering for user-friendly centralized biometric authentication

Christian Gehrmann*, Marcus Rodan and Niklas Jönsson

Abstract

While biometric authentication for commercial use so far mainly has been used for local device unlock use cases, there are great opportunities for using it also for central authentication such as for remote login. However, many current biometric sensors like for instance mobile fingerprint sensors have too large false acceptance rate (FAR) not allowing them, for security reasons, to be used in larger user group for central identification purposes. A straightforward way to avoid this FAR problem is to either request a user unique identifier such as a device identifier or require the user to enter a unique user ID prior to making the biometric matching. Usage of a device identifier does not work when a user desires to authenticate on a previously unused device of a generic type. Furthermore, requiring the user at each login occasion to enter a unique user ID, is not at all user-friendly. To avoid this problem, we in this paper investigate an alternative, most user-friendly approach, for identification in combination with biometric-based authentication using metadata filtering. An evaluation of the adopted approach is carried out using realistic simulations of the Swedish population to assess the feasibility of the proposed system. The results show that metadata filtering in combination with traditional biometric-based matching is indeed a powerful tool for providing reliable, and user-friendly, central authentication services for large user groups.

Keywords: Biometric authentication, Metadata filtering, System simulation

1 Introduction

Online authentication for many services is still to a great extent done through usage of username and passwords. Textual password-based authentication suffers from several notable security problems such as low password entropy and poor password management [10, 20, 41]. It has also been found by several studies that sufficiently secure passwords are hard to remember, often resulting in that users select predictable structures when constructing their passwords [20, 38].

A more user-friendly alternative, compared to textual passwords, is the usage of biometric authentication [12] to verify and login a user on a device with biometrics capabilities, only demanding a biometric trait to be presented. One example of a widely adopted biometric trait is the fingerprint [4]. It is also the case that most modern smartphones are equipped with fingerprint sensors. Hence, we have in this paper decided to investigate a fingerprint-based identification service. Biometric authentication is typically only used locally on each device and not for remote login. Fingerprint-based remote login often requires central processing and matching of the fingerprint data. A main advantage with a centralized solution is that it does not hinder users from using any compatible device, thus being more suitable for cases where users are expected to move between different devices. It should be noted though at this point that there exist centralized biometrics-based systems like the Indian Adhaar system [19]. The Adhaar system and similar systems are only providing a match or no match response for a given user identity. In this system, the user (or the one requesting the identification) must enter a unique user ID before the actual biometric matching can take place. Even if this is a very efficient solution, allowing one against one biometric matching, it is not particularly user-friendly. Hence, we are interested in investigating alternative solutions for central biometric authentication not requiring the user to enter a unique user ID and actually trying to find solutions where the user does *not* need to present any additional information at all, in most cases, apart from the biometric during an identification session.

*Correspondence: christian.gehrmann@eit.lth.se
Department of Electrical and Information Technology, Lund University, Box 118, SE-221 99 Lund, Sweden

One reason for the lack of widely used central biometric authentication solutions are privacy issues as a fingerprint template must be available in a central repository. Consequently, there has been quite some resistance in the past against centralized storage and processing of biometric data for commercial systems [28]. However, recent achievements in the area of biometric transforms [29, 33, 40, 45] pave the road for more secure sharing of biometric data. These techniques provide raw data protection through application of non-invertible transformations. Hence, they open up for centralized, privacy-preserving identification services which we are investigating in this paper. It should be noted though that the main focus of this paper is the metadata filtering and not the biometric matching, and the results are independent of the actual use of a biometric transform or not.

With a strong biometric transform in place, the remaining barrier for wide adoption of central biometric authentication that requires a solution, is the ability to maintain acceptable performance as the user population grows. This task is challenging for different reasons. To start with, the manufacturers have been moving toward smaller sensors because of cost pressure[12]. This trend has progressed, meanwhile, attempting to maintain a high level of security. The typical false acceptance rate (FAR) provided by mobile, fingerprint-based authentication, is for instance, about $\frac{1}{50000}$[3]. Meanwhile, being sustainable for biometric matching against a small local database, this makes it hard to use these sensors together with associated algorithms directly in a centralized authentication system since false matches will occur with a high probability[4]. Techniques for biometric data protection such as biometric transforms also introduces some performance limitations that makes it troublesome when applied to a huge user population [23]. It can, therefore, be concluded that the proposed system needs to adapt to the fact that fingerprint templates will only contain partial[5] fingerprint information that has been transformed. The system should nevertheless be able to operate in such a setting. The proposed solution in this paper is to use metadata filters. Even if metadata filtering can degrade the overall user privacy of the system, it does not influence the biometric privacy if a biometric transform is used in combination with the metadata filtering technique. An alternative approach with biometrics transforms, such as the biometric-based indexing techniques presented in [23] has much worse performance than the filtering principle we present in this paper. Furthermore, it does not solve the large user group FAR issue but is a tool for faster identification through usage of biometric indices without leaking biometric features information. To solve this issue, we in this paper are investigating an approach where end-user metadata is collected from the user and used as a first *identification* procedure to find a candidate set of users. This limited set is then used to authenticate the end-user using traditional biometric matching based on the end-user biometric input. In order to provide a truly user-friendly experience, the metadata collection should require minimal end-user effort while still allow high accuracy identification performance.

The proposed approach, in this paper, relies on the fact that it is possible to collect metadata to perform an initial candidate reduction prior to biometric matching. The proposed system, will, therefore, gather and store metadata during both enrollments, as well as, identifications to enable efficient reduction. The motivation for this study is two-fold. Firstly, we want to cast light on the problem and its requirement. Secondly, we want to propose a solution to the stated problem and show the efficiency of the solution. We have carried out our investigation by considering a selected, realistic and generally available on mobile devices, set of metadata types for the Swedish population. The reason for limiting the study to the Swedish population is that it is hard to find sufficiently good statistics on a large scale. However, the results can easily be extended to other populations or metadata types. The proposed method can be decomposed into two components, where the first component is the metadata-based candidate reduction and the second component is the biometric-based matcher. The paper main focus is on the former component.

The main contributions in the paper are the following:

– We investigate metadata filters for a set of metadata types generally available or easy collectible on wide range of devices.
– We demonstrate through simulations that is feasible to identify users using previously unused devices with high confidence using advanced metadata filters.
– We show that it is possible to achieve a user-friendly and reliable identification with acceptable FAR and false rejection rate(FRR) using automatically collectible metadata types in many of the cases, only requiring manual metadata collection in a limited number of cases.

The organization of this paper is as follows. Section 2 provides a presentation of the proposed architecture. Section 3 defines an attack model and states assumed conditions. Section 4 presents the developed metadata filters. Section 5 discusses the performed metadata generation. Section 6 gives performance figures for the proposed filters, as well as, discussing the security, provided by the system. Section 7 discusses potential privacy concerns associated with location information, as well as, potential mitigation strategies. Section 8 presents an overview of related work. Finally, Section 9 summarizes the paper and suggests potential future work.

2 System overview

In this paper, we study a biometric identification system with the following three main components:

- A trusted, centralized identification service, which we refer to as the biometrics trusted service (BTS). This component is responsible for providing secure storage of both biometrics and metadata, as well as, performing secure identification of end-users on the behalf of different applications.
- An end-user device (C) equipped with a fingerprint sensor, as well as, means to provide relevant metadata. A user can enroll in the system using any compatible device, and can thereafter, use any compatible device to identify him or herself.
- An application (A) delegating the burden of performing end-user identification and authentication to the BTS. Potential applications include for an example mobile apps, as well as, a payment method for various services.

Figure 1 illustrates the enrollment procedure, where a user enrolls in the system by using a compatible device. The device assembles an enrollment request consisting of biometrics and metadata which is relayed to the BTS. The BTS then stores this information in a way such that it is possible to later perform the required candidate reduction.

Figure 2 shows the proposed identification procedure followed during end-user identification and final authentication. The device initiates the procedure by sending an identification request to the application, which then relays the request to the BTS. The BTS then produces a reduced candidate set before applying the biometric-based matcher to reach an authentication decision, which is finally relayed to the application, as well as, the device. The BTS also updates its internal data structures based on the combined identification and authentication result to offer more accurate and efficient future authentication. The core focus of this paper is on steps 3 and 5 in Fig. 2.

3 Security assumptions and attack model

It is important to note, that the proposed metadata-based pre-processing approach is an integral part of a system. The security requirements, imposed on the reduction strategy, is therefore only a part of the requirements imposed on the entire system. In particular, we assume that the encapsulating system fulfills a number of conditions. We assume that the following conditions are met by a concrete implementation:

- The system supports secure enrollment of legitimate end-user biometrics and metadata, without any interference from any third party.
- Out-of-band authentication is performed during enrollment to access the validity of an enrolling user.
- The communication channels between devices, applications, and the BTS is confidentiality- and integrity-protected.
- The devices provide a secure execution environment, directly coupled with the biometric sensor. This environment is responsible for providing biometric data protection before the data is submitted to the BTS over a confidentiality and integrity-protected channel.
- The system provides authentication of the trusted environment on the compatible devices. In more concrete terms, the system ensures that only data from trusted, compatible units are accepted for further processing on the BTS.
- The trusted execution environment on the device applies a non-invertible biometric transform before transmission to minimize the impacts associated with any data leak.
- The device provides a uniquely, cryptographically verifiable device ID which can be verified by the BTS.

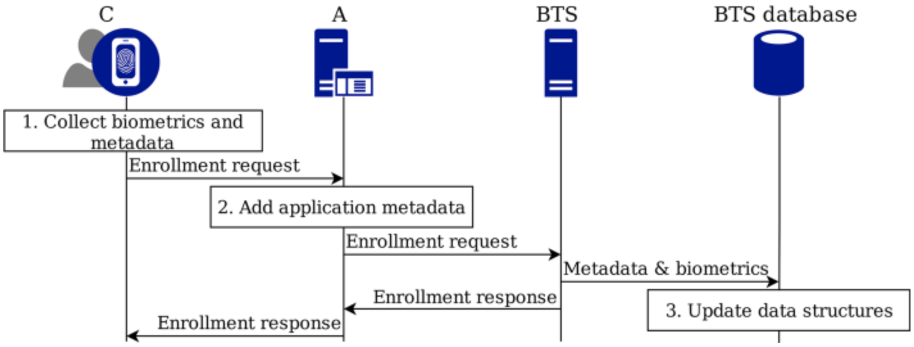

Fig. 1 Illustration of the enrollment procedure

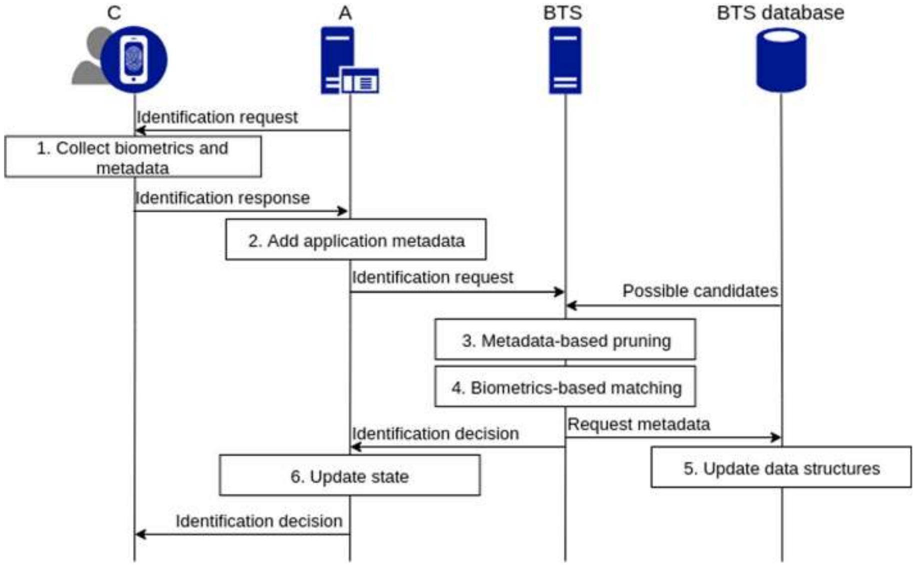

Fig. 2 Illustration of the identification procedure

– The BTS protects the stored data and provides secure environments for data storage, pre-filtering, and biometric matching.

Given these conditions, we then consider an adversary with the following capabilities:

– The adversary is capable of manipulating the metadata sent to the BTS during his/her own enrollment procedure. One motivation for assuming such a capability is that the device will use various sensors for metadata collection, that may not be entirely trusted.
– An adversary is able to adjust the metadata so that he or she can control the metadata seen by the BTS. An attack is successful if the adversary can successfully authenticate him or herself through a match against another user.

Section 4 will describe a protection mechanism against these capabilities and Section 6 will provide a security evaluation of the system.

4 Metadata filtering

In this section, we present our general filtering approach and filter designs for an elected set of metadata types, generally available or efficiently collectible on mobile devices.

4.1 Generic filter procedure

Figure 3 shows the identification procedure followed by the BTS for each incoming identification request. The BTS begins with verifying the authentication request to ensure that the request has been sent by a trusted device. The BTS then performs a check to protect against an adversary attempting to identify as a random user. Cnt_{did} is a sliding-window counter, recorded by the BTS to keep track of how many times each device ID has failed to authenticate correctly the last year and R is a rate limit. Requests that are deemed as legitimate are then passed along to the metadata-based filtering stage, where the most likely candidates' templates are extracted. These templates are then used to reach an identification verdict. Cnt_{did} is then either increased or left unchanged depending on the outcome. The metadata records are also updated in the case of a successful identification. The counters, as well as, limits, introduced in this section, are aimed at preventing illegitimate access attempts as discussed in Section 3. These numbers, as well as the number of selected top candidates, k, can be tuned to meet different security requirements, as will be discussed in Section 6. Section 6 will also slightly revise the proposed procedure in Fig. 3 to achieve a more user-friendly solution. The security mechanisms will, however, be unaffected.

4.2 Selection of metadata types

The investigated metadata types were elected taking the following metadata properties into account, chiefly inspired by the properties defined by Jain et al. [17] to evaluate different biometric traits.

1. Universality: the selected metadata types should have high availability, implying that most users possess and can supply the metadata type.
2. Distinctiveness: metadata types of higher entropy are more desirable than metadata types of lower entropy.

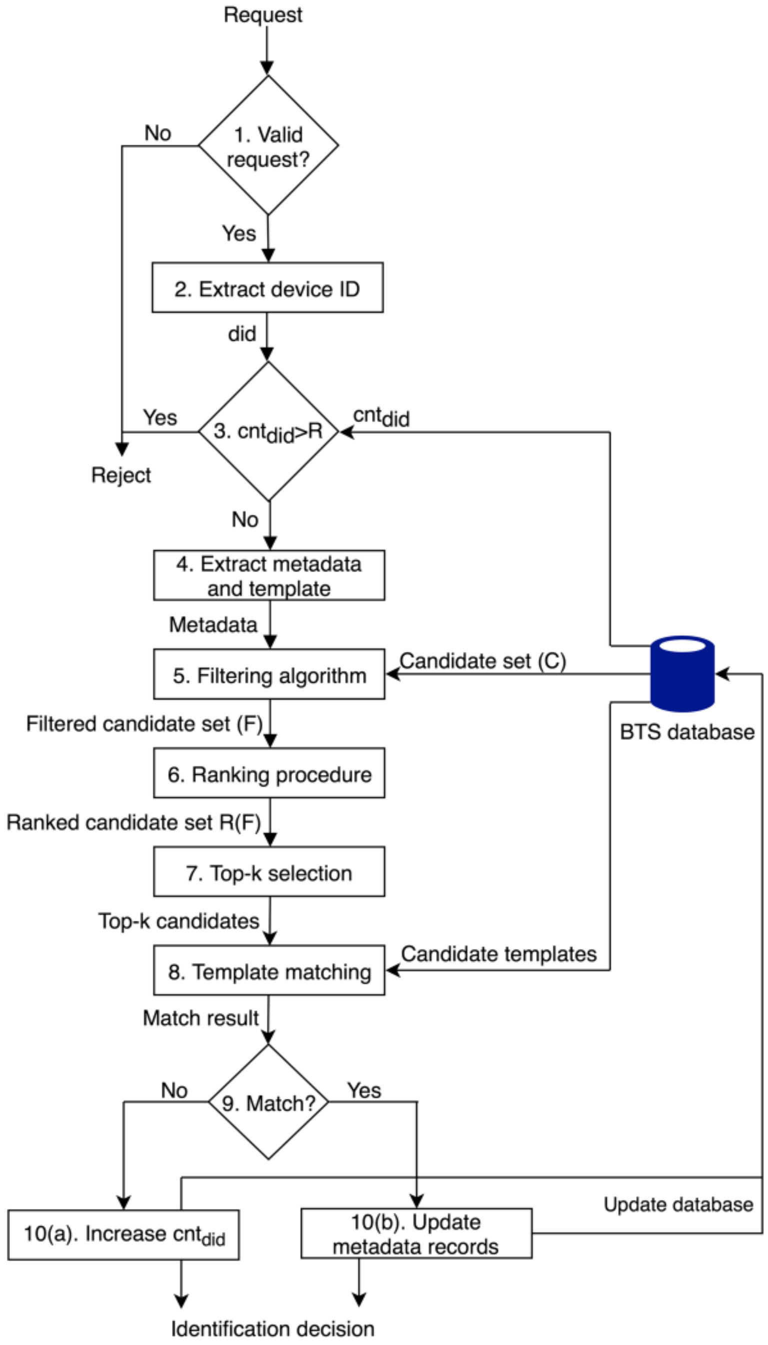

Fig. 3 Illustration of generic metadata filtering

3. Permanence: the selected metadata types should be relatively stable over time.
4. Collectability: the metadata types should be as effortless as possible to collect to ensure a high level of user-friendliness. The metadata types can be divided into automatically and manually collectible metadata types. The automatically collectible metadata types are superior from a user-friendliness perspective.
5. Acceptability: the privacy concerns associated with meta collection varies between types [13] where less sensitive metadata types are preferred.

The goal of the conducted metadata election was to select a set of both automatically and manually collectible metadata of different distinctiveness, permanence, and acceptability believed to be widely collectible using the targeted mobile devices. However, the elected metadata types are by no means a complete set and other potential metadata candidates such as gait exist, discovered to be unique to some extent [30]. Furthermore, it is also the case that there exist other potential algorithms for the selected metadata types than the algorithms investigated in this paper.

4.3 Filter implementations

We have investigated unimodal filters for the chosen metadata types, i.e., device ID, location, name, and age. Below, the different unimodal filters for all these four categories are defined. We have also investigated different combinations of the unimodal filters. The general principle we used to combine unimodal filters is described in Section 4.3.4.

4.3.1 Filtering based on device ID or age

The usage of certified readers with uniquely assigned device IDs permits reliable and user-friendly filtering. The principle for filtering on age is the very same, and Algorithm 1 describes the device ID-based (age-based) filter.

Algorithm 1: Device ID-based (age-based) filtering

```
Data: candidate set ($C$) and query device ID (q_id) (age (q_age))
Result: Filtered candidate set $R(F)$
cands ← candidates in $C$ known to have used q_id (q_age); return cands;
```

The privacy issues associated with usage of fixed device IDs can be mitigated using communication channel encryption to ensure safe transmission between the certified readers and the BTS.

4.3.2 Filtering based on location

Prior studies, [6, 22] has discovered that user movement is deterministic to some degree, indicating that location information is a suitable candidate to include in the filter procedure. Smartphones also typically come with an integrated GPS module thus enabling user-friendly automatic position information collection. Algorithm 3 describes the location-based filter approach using an R-tree [15] together with user-dependant kernels to model each user's movement patterns, as suggested by Zhao et al. [47]. Each user will, therefore, have a Gaussian kernel updated after each successful identification attempt to facilitate accurate prediction of each user's probability of occurring at the query position. The used kernel bandwidth was obtained through a grid search strategy, optimizing the global identification performance.

Algorithm 2: Location-based pre-processing

```
Data: candidate set ($C$) and query location (q_lat, q_lng)
Result: Filtered candidate set $R(F)$
lat_low, lng_low, lat_high, lng_high ← rectangle covering all positions within 30km of (q_lat, q_lng);
cands ← Obtain users with a location record within (lat_low, lng_low, lat_high, lng_high) using an R-tree;
return cands;
```

Algorithm 3: Location-based filtering

```
Data: candidate set ($C$) and query location (q_lat, q_lng)
Result: Filtered candidate set $R(F)$
cands ← Algorithm-2(C, q_lat, q_lng);
cand_scores ← [];
for cand in cands do
    cand_kernel ← location kernel associated with cand;
    append (cand, cand_kernel(q_lat, q_lng)) to cand_scores;
end
return sorted cand_scores;
```

The major obstacle that may perhaps prevent the usage of location information is that it is privacy sensitive [22, 24], implying that some users may have concerns about sharing such information. Investigated mitigation strategies will be introduced in Section 7.

4.3.3 Filtering based on name

Asking the user to enter his or her name, cause some extra burden on the end-user. One way to alleviate the user burden is to construct a system able to function even if the users provide clumsily entered names. The proposed filter as seen in Algorithm 4 utilizes a n-gram-based lookup [46] as well as Jaro-Winkler distances [44] to provide stability against spelling mistakes. Another way as reviewed in Section 6 is to postpone name collection until it is necessary, thus providing a more user-friendly identification on average.

4.3.4 Combination filters

We are also using combination of the different filters. A combined age filter with name or location is obtained by first running the sorted candidate list obtain from the name or location filter and next applying the age filter on the obtained list. A combination filter of location and name is obtained by first applying the location filter and then *fusing* the normalized scores (direct addition) from the location and name filter respectively to get a sorted

Algorithm 4: Name-based filtering

```
Data: candidate set ($C$) and query name (q_fname,
      q_lname)
Result: Filtered candidate set $R(F)$
grams ← 2-grams of q_fname ∪ 2-grams of q_lname;
cands ← all candidates in $C$ sharing at least one
2-gram with grams;
cand_scores ← [];
for cand in cands do
    cand_score = Jaro-Winkler(first name of cand,
    q_fname) + Jaro-Winkler(surname of cand,
    q_lname);
    append (cand, cand_score) to cand_scores;
end
return sorted cand_scores;
```

candidate list. Furthermore, we obtained a combined filter on location, age, and name by first applying the combined location and age filter and finally using the fusing score method (as used for combining location and name as described above) to combine the output of the age and location filter with the naming filter.

Direct combinations between the device ID filter and the other filters are not good approaches as the ID filter is binary. Either, the correct candidate is within the set with moderate probability or it is missing. If it is missing, one should not exclude any of the remaining candidates in the database. Hence, the device ID filter is better used as filter to make a first try for the biometric matcher and if not match is obtained, continue with any of the other filter options as we describe in Section 6.

5 Simulation framework

We have evaluated the suggested approach using statistical databases and a simulator framework. The actual filtering is following the procedures shown in Figs. 1 and 2. We have also tested the approach with fingerprint matching with the sourceAFIS open source fingerprint matching algorithm[6] and the FVC2006 fingerprint database [8]. This has been done to show a complete matching behavior rather than evaluating the meta data filtering as such. Below, we describe the different parts of the metadata simulation framework as well as the fingerprint matching used in some of the experiments.

5.1 Enrollment metadata

We distinguish between data generated during enrollment and later during an identification session. We start each simulation by enrolling users to the database. During this phase, each user is assigned a set of metadata attributes. We discuss these different attributes and how they are assigned to users in more details below.

5.1.1 Device ID

Each user enrolling to the system is assigned an initial number of associated devices which will be stored in the simulation database. The number of devices initially assigned to a specific user is regulated according to a configurable distribution. The only device known by the BTS is the device used during the enrollment, all other devices are unknown until they are used.

5.1.2 Age

The age of an enrolling user is generated using a gender-dependant age distributions extracted from SCB [37]. The SCB is governmental service providing highly reliable statistics for the Swedish population. Figure 4 shows the truncated age distribution of males and females in Sweden between 0 and 100 years old. This distribution has been extracted using statistics from SCB and is used to generate the age of an enrolling user, where it is assumed that users will enter their correct age during enrollment.

5.1.3 Name

The name of an enrolling user is generated using a gender-dependant name distributions also extracted from Swedish SCB [35, 36]. Each enrolling user is also assigned a user-dependant typing-error probability, incorporating the fact that different users tend to be different good at entering their names correctly. The typing-error probability is only stored in the simulation database, meaning that the BTS is completely unaware of this probability. More specifically, we have extracted truncated probability distributions for the 100 most common first and surnames, where we are also taking into account that females and males typically have different names. Figure 5 shows the extracted distributions that have been extracted using statistics from SCB. As with age, we assume that users will enter the correct name during enrollment.

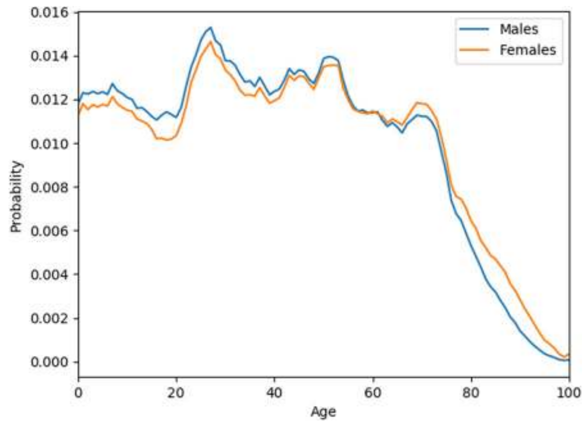

Fig. 4 Truncated age distribution of males and females in Sweden between 0 and 100 years old

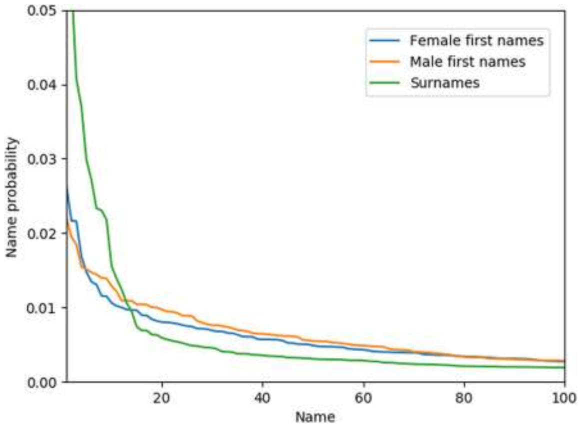

Fig. 5 Truncated most common name distribution for males and females in Sweden

5.1.4 Location

Firstly, we assign each enrolling user a home town which is based on statistics gathered from SCB [34]. We then assign each enrolling user a specific location within this town using statistics about the size of the town. This means that the enrolling user now has a given enrollment location. We then associate each enrolling user which a given number of significant locations, with support from various studies [16, 48]. The significant locations are stored in the simulator database to only allow the BTS to know the enrollment location. In this manner, we can simulate scenarios where the BTS does not know anything about significant locations which allows us to test how the different filtering algorithms behave when they have a varying amount of previous login locations. Figure 6 shows a heat map illustrating the enrollment position of 200,000 enrolled users. As can be seen in the figure, northern Sweden is less densely populated than southern Sweden.

5.1.5 Fingerprint data

We do *not* assign biometric data to the user during enrollment. The main reason to this is that there is no biometric data available for such large user groups which we are interested to investigate. Instead, in the simulations we *randomly* assign a fingerprint enrollment data to users from a smaller, available, fingerprint database [8], after the metadata filtering has taken place. Hence, we have divided the available fingerprint data into two subcategories, enrollment dataset and candidate dataset, and the matching data is only selected during the actual final identification step.

5.2 Identification metadata

Identification metadata is simulated using the simulation database together with statistics. The goal is to mimic a user following the procedure in Fig. 2.

5.2.1 Device ID

The device ID generation is performed using three probabilities. The first probability (20%) corresponds to the likelihood of reusing a previously used device. The second probability is the likelihood of borrowing another users device (10%). The third and final probability is the likelihood of using an unenrolled device (10%). The results shown is Section 6 are used with these probability selections. However, we have run simulation tests with several other different probabilities as well.

5.2.2 Age

The age of an identifying user is simulated by first extracting the correct age from the simulation database. An error-injection procedure is then invoked to account for the fact that it may happen that some users may mistype their age. More concretely, with a probability of 25% we assign a to low age by one year and by a probability of 25% we assign a to high age by 1 year. With a probability of 50%, we simulate a login with the correct age of the user[7]

We do only apply this random typing error procedure during identification session, i.e., we do not consider typing error during enrollment but only during identification.

5.2.3 Name

The simulation of names associated with identifying users is divided into several steps. The first step of the simulation is to extract the correct name from the simulation database together with the user's typing-error probability. The typing-error probability is then used to generate the number of typos to inject, according to the distribution extracted from the simulation database, into the name. This number is then passed along together with the correct name to a transformation, responsible for generating realistic typos according to a model found in an empirical study by Baba and Suzuki [3]. The approach in [3] is used to generate which kind of typos that should be applied sequentially to the extracted name from the simulation database. The number of applied typos is then governed by the number of typos that should be applied to this particular login attempt, which is sampled from the extracted typing-error probability for the particular user which, basically, represents how sloppy the user is when typing his/her name during identification. This distribution is generated during enrollment and is saved in the simulation database, i.e., the BTS does not have any knowledge of this distribution and can thus not use if for identification purposes.

5.2.4 Location

The location simulation is accomplished by first extracting the user's significant places from the simulation database. A time-dependent location model [26] is then used together with a model, modeling user-activity during

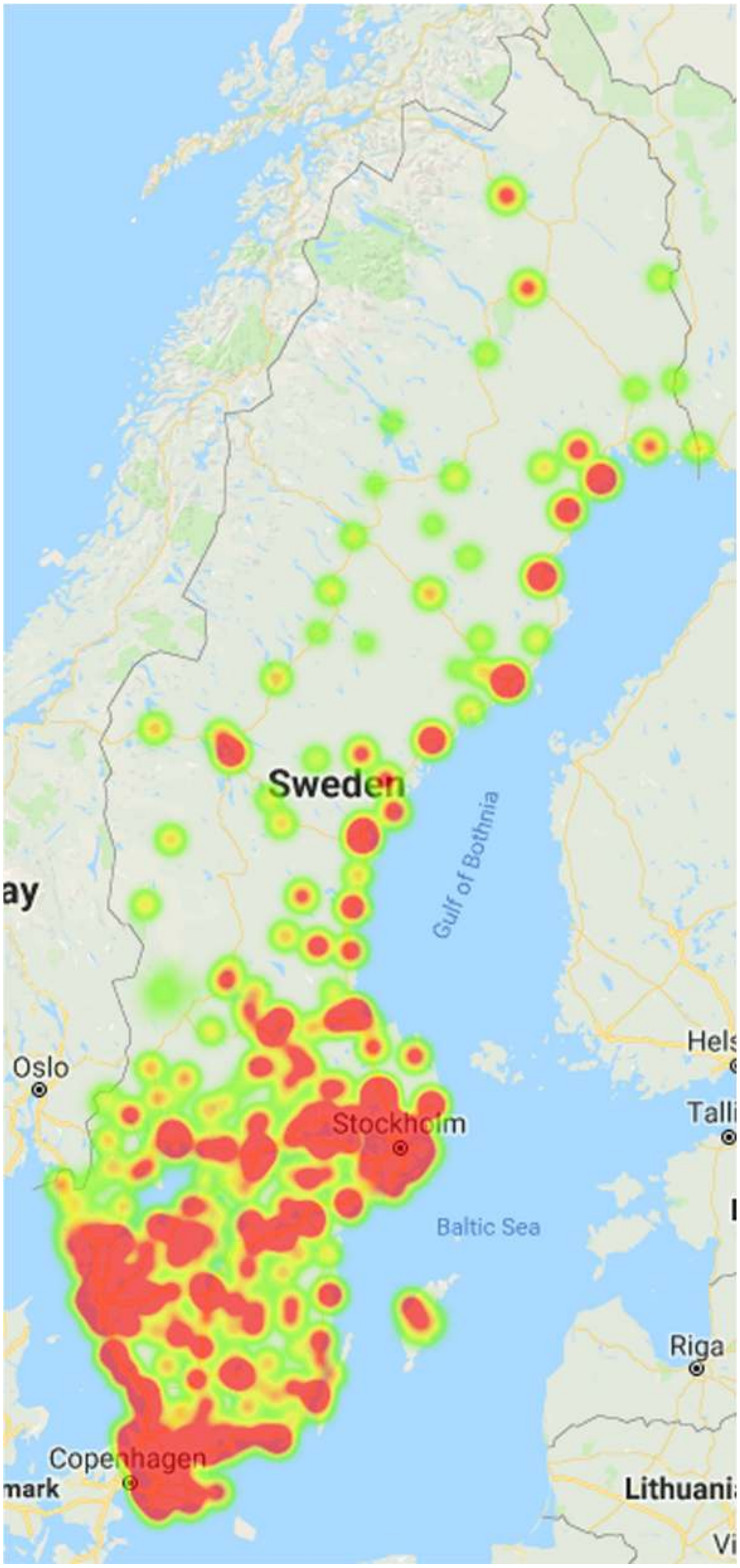

Fig. 6 Heap map showing enrollment locations of 200,000 users

the day [18], to generate the query position. More particularly, the first step gives us a number of significant places associated with the user that tries to authenticate which are unknown to the BTS. We then select the particular significant place to emulate according to the two distributions in [26] and [18]. Given the extracted significant place we then move away from the center of the significant place following a uniform disc distribution where the radius of the disc is varying depending on the type of significant place. This strategy allows us to incorporate the fact that different users will follow different patterns.

5.2.5 Fingerprint candidates

In the simulations, first an enrollment set is picked from the enrollment fingerprint dataset. We have used 4 templates from DB2_A in FVC2006 [8] as enrollment set and the remaining 8 templates as matching candidates. A random candidate not part of the enrollment set is then selected for the actual matching test. If the correct user is caught by the meta data filter, the candidate is selected from the non-enrollment set of the same user. If the correct user is not caught, a random candidate from the non-enrollment set for a *different* user is selected.

The identification test is then performed against enrollment data of users corresponding to the chosen filter set size, k ($k = 4, 7, 10, 17, 22,$ and 50) for the different considering filtering cases. The enrollment set of the chosen size is picked at random from FVC 2006 DB2_A with the exception that if the correct candidate is filtered out by the metadata filter, the enrollment sets is selected from the remaining $k - 1$ enrollment templates in the database (different from the candidate).

6 Results

In this section, we present performance evaluation of the introduced filters using the simulation framework. We also present a security analysis of the filtering approach.

6.1 Performance evaluation

We divide the performance evaluation into two categories:

- Top k recall rate evaluations
- Incremental filtering

The top k recall rate applies the previous introduced filters and sort out the top k candidates from the filter. The top k recall rate is, in this context, defined according to formula (1) below. $N_{\text{successful reductions}}$ corresponds to the number of successful candidate reductions, meaning that the correct candidate is within the reduced set, consisting of the k most likely candidates. N_{requests} corresponds to the number of identification requests.

$$\text{Recall@}k = \frac{N_{\text{successful reductions}}}{N_{\text{requests}}} \qquad (1)$$

The recall rate gives the statistical correctness of the filter, i.e., the rate of correctly includes the true user candidate in the filtered set. The incremental filtering is a heuristic and more user-friendly approach where we instead of choosing a fixed filtering size, connect different filters with different choices of k after each other in a cascade. We show that this approach indeed outperforms the straight forward top-k approach.

6.1.1 Top k recall rate

All measurements, concerning filer accuracy and execution time, have been carried out on a machine running Ubuntu 18.04.1 LTS with an Intel Core i7-3770 CPU clocked at 3.40 Ghz with 16 GB RAM. The simulations have been done using implementations mainly in Python and Java.

Figure 7 shows the obtained top 50 as well as the top 10 recall rates for the reviewed filter strategies. The advanced multimodal filters are superior to the unimodal filters, especially for large end-user populations. The figure also clearly shows a presence of a strong positive correlation between length of location history and performance, therefore, indicating that a deployed system should strive to utilize all available location information. It is, furthermore, concluded that unimodal filters are more perceptive to changes in k than advanced multimodal filters. The figure also shows that the device ID-based filter can provide unaffected performance for large end-user populations since the number of devices grows with population size.

We have also measured FAR and FRR for a complete identification case for different thresholds and a user database sizes of 100,000 and 200,000 people respectively. Table 1 summarizes the main results for a couple of different filter cases and a fixed, matching score threshold equal to 50 for the source AFIS-matching algorithm. For this matcher and these parameter choices, we have a FAR $= 0.0016$ and FRR $= 0.004$ for the given matcher and an identification case with just 50 users, i.e., without any filtering. As expected, when we then increase the population size, if the filter is good, the FRR will be determined by the FRR for the given k value. While if the filter is not accurate, the FRR will be determined by the recall rate. Hence, it is important to adjust the filters such that the FRR is kept at the level well below the FRR of the biometric matching algorithm. The FAR is determined by the chosen k and threshold together with the performance of the biometric-matching algorithm. If k grows, the FAR grows as well. As our database is limited to 140 users, it is not possible to get good figures for large user populations, but Fig. 8 shows how the FAR grows with k and matching threshold 50 for the user in the FVC2006 DB2_A database. As can be concluded from the figure, FAR grows approximately linear with the user population size. Hence, for the given

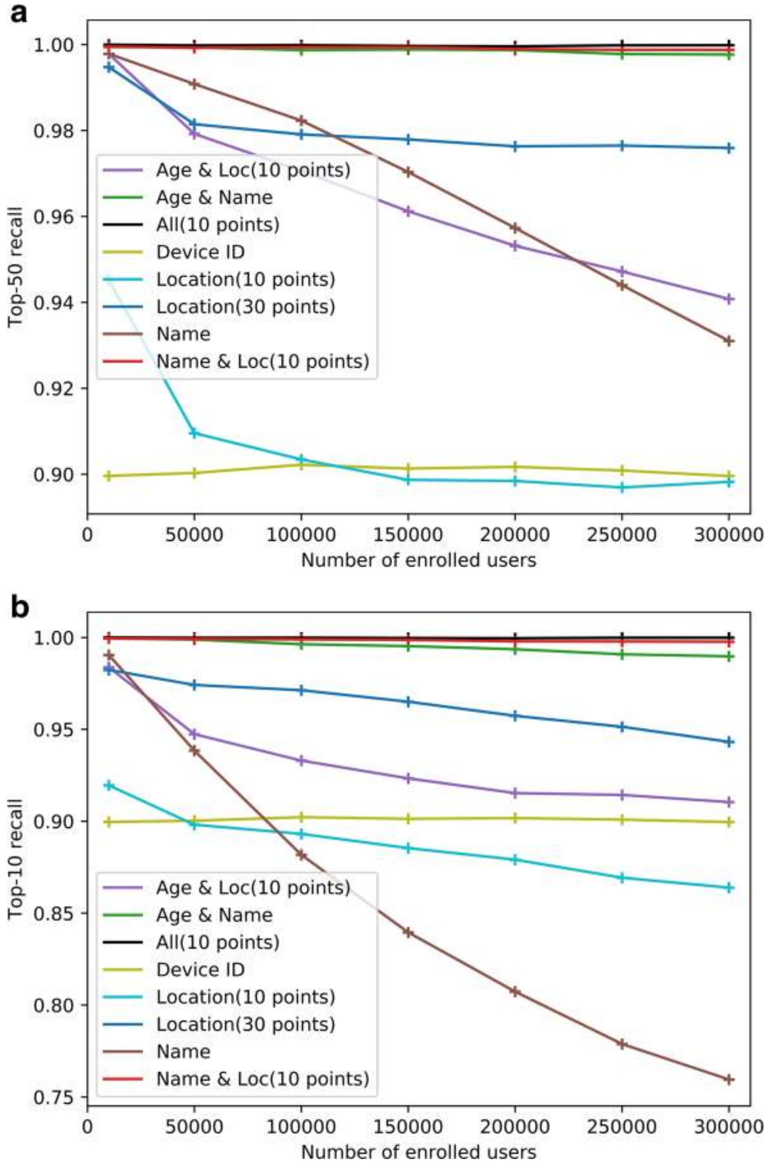

Fig. 7 Top-k recall for investigated strategies

matcher and threshold, we expect to reach a FAR of 1 at around 30,000 users in the matching set without the proposed filtering. Figure 9 shows the computation time for the filters presented in Section 4. Here, we only measure the computation time for the filter and not the matching as the actual matching contains reading from the fingerprint database and not from a real fingerprint reader. This is justified by the fact that we are mainly interested in the filtering performance and not the biometric matching performance as such (any biometric matcher can be used together with the suggested filtering techniques). One can notice, that the only algorithm not increasing in a traditional linearly fashion is the device ID-based algorithm. This is to be expected since this algorithm is indexed-based, meanwhile, the other algorithms are score-based. It is, furthermore, seen that the multimodal algorithms based on age, as well as, name and location are quite fast. The algorithms incorporating age information gains speed through efficient age lookup and the algorithm using name and location information gains speed because of the pre-filtering steps consisting of location-based indexing and n-gram lookup.

6.1.2 Incremental filtering approach

Figure 10 describes an alternative approach building upon incremental application of the previously presented filters.

Table 1 Recall and FRR at FAR = 0.00164, k = 50 for different filters and population sizes

	100.000 users			200.000 users		
	Name	Location (30 p.)	Combined	Name	Location (30 p.)	Combined
1-Recall	0.018	0.022	0.00025	0.0042	0.024	0.00031
FRR	0.022	0.028	0.004	0.046	0.026	0.004

The main advantage of this approach is that, compared to the earlier introduced filters, this approach constitutes a more user-friendly alternative since it avoids excessive metadata collection. A grid search was performed, optimizing global identification performance to determine the top-k configuration shown in the figure. Table 2 presents the obtained recall rates when the device ID-based filter fails for a varying number of enrolled users, as well as the metadata collection frequency measuring how often an identifying user is prompted by automatic or manual means for a specific metadata type. The table clearly shows that the proposed approach is user-friendly since manual metadata collection is limited to a few cases. It is also, straightforward to observe the adaptability of the proposed system, requiring more metadata as the population size grows.

Figure 11 shows the computation time for the improved algorithm. The figure clearly shows that the improved algorithm is still able to operate efficiently, compared to the unimodal filters. This is because of its incremental nature, only collecting metadata when necessary.

It should be reasonably clear that the provided execution time measurements do not include the time consumed by a biometric matcher, given that it has been mentioned already. It is, therefore, of great interest to quantify the total time consumed during an identification. Let us, for an example, assume that the enrolled database contains 50,000 enrolled users. We can then see that the average identification time is less than 0.2 s. It can also be inferred that a biometric-matcher, for an example the biometric matcher used in [23], needs less than 0.1 s to compute the exact score for 50 candidates. The conclusion is, hence, that the proposed solution is able to perform the complete identification (both initial filtering and final biometric matching) in less than 0.3 s when 50,000 users are enrolled. This can for instance be compared against the result of 1 s with 32,000 enrolled users presented in [23].

6.2 Security evaluation

This section provides a security analysis of the proposed service. It should first be noted that filtering on metadata is very different from security point of view than just making biometric score fusion between metadata matching and biometric matching. While metadata filtering is very powerful to *reduce* the population size against which a biometric identification is done, from security

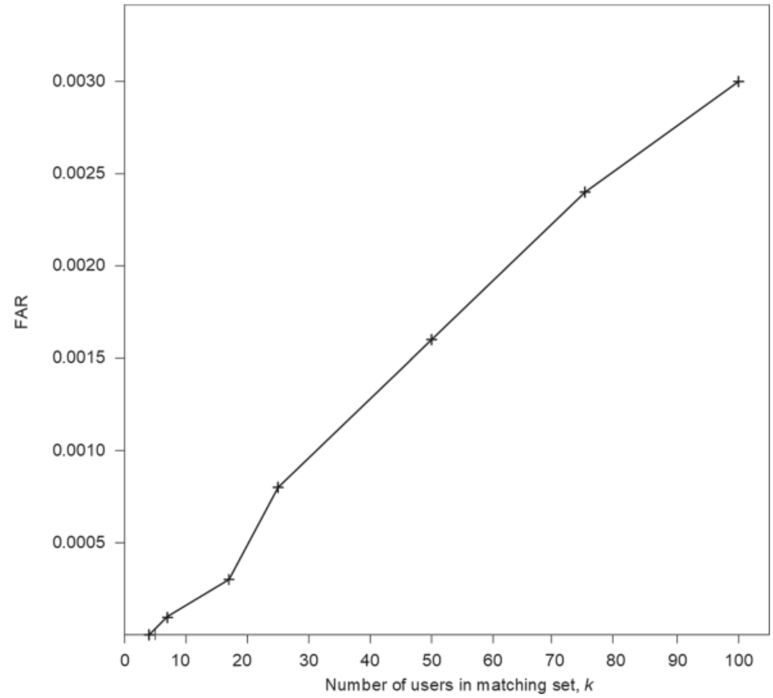

Fig. 8 FAR as a function of the size of the matching set

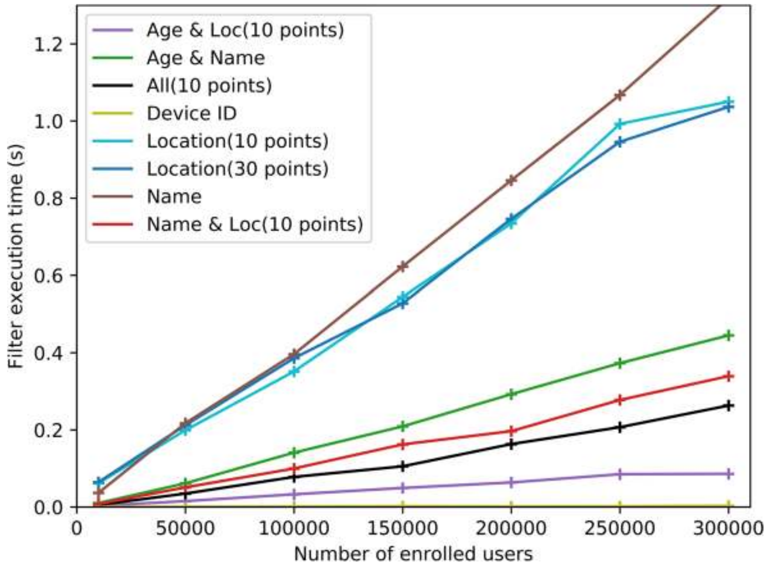

Fig. 9 Average computation times for the investigated filters

perspective, one can typically not rely on the metadata and make fusing scores between biometric matching score, and metadata filter score is not possible as that would allow an attacker to choose the meta data parameters in a way that would increase the FAR. Instead, as previously explained, we work with a two step approach: *first*, we reduce the matching set using metadata filtering and *second*, we perform a traditional biometric matching on a largely reduced population.

The assumptions stated in Section 3 is assumed to be satisfied by the encapsulating system implementation. Under those circumstances, we then discuss the following three kinds of attacks:

- Malicious enrollment
- Identification as a random user
- Identification as a specific user

6.2.1 Malicious enrollment

An adversary may have three different goals in the first attack. One goal could be to provide false metadata together with genuine biometric data to identify successfully as another person. Another goal could be to provide false biometric data together with genuine metadata to attempt to get another user identified as the adversary. A third, and final goal, could be to register false metadata together with false biometric data to generate confusion.

The proposed system is well-protected against the first attack because of two reasons. The first protection mechanism is that a user profile consists of both metadata and biometric data. This implies that an adversary will still need to bypass the biometric matcher to be identified as another user. It is, furthermore, assumed according to the security assumptions in Section 3, that an out-of-band authentication is performed during enrollment. These two mechanisms make it hard for an adversary to be successful with this approach.

The system is in the same way protected against the second attack scenario since it is assumed that the devices provide a trusted execution environment, thus protection against tampering of biometric data. One could,

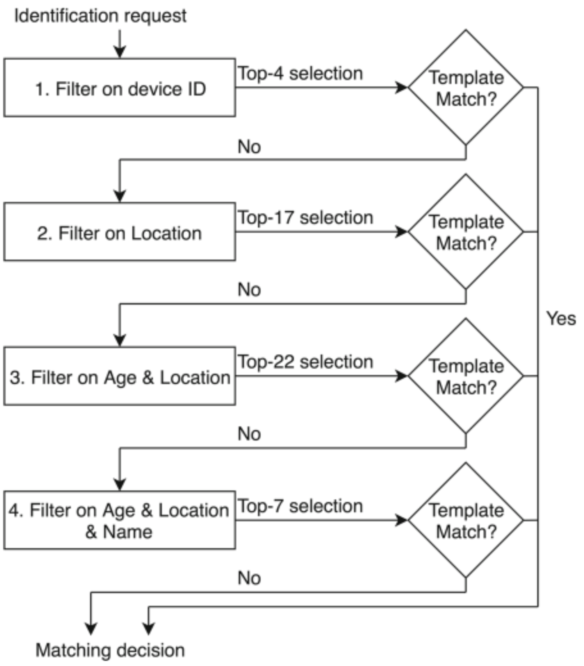

Fig. 10 Incremental filter strategy

Table 2 Performance of the incremental approach

Number of users	Recall	Location collection rate	Age collection rate	Name collection rate
10,000	0.9999	1.000	0.0609	0.0076
100,000	0.9994	1.000	0.0980	0.0430
200,000	0.9986	1.000	0.1076	0.0663
300,000	0.9974	1.000	0.1157	0.0809

however, imagine a scenario where an adversary simply asks another person to put their finger on the physical sensor. However, this will not increase the adversary's probability of being identified as another user later on.

The system is protected against the third scenario because of the previously presented mechanisms for the other two scenarios.

6.2.2 *Identification as a random user*

An adversary's goal in the second attack is to identify as another random user. Given the stated conditions in Section 3, the adversary then needs to attempt to exploit the fact that the biometric matcher has a probability FAR of falsely stating that two templates come from the same finger. It is, therefore, the case that a false match will occur on average if $\frac{1}{FAR}$ comparisons are carried out. Let us now assume the worst-case scenario that each identification request retrieves k completely new candidates. Let R equal the rate limit set by the system per device, i.e., the maximum number of allowed non-successful identification requests from a particular device per year. An adversary will then need D devices on average, where D is given by formula (2), to be successful within T years.

$$D = \left\lfloor \frac{\frac{1}{FAR}}{R * T * k} \right\rfloor \qquad (2)$$

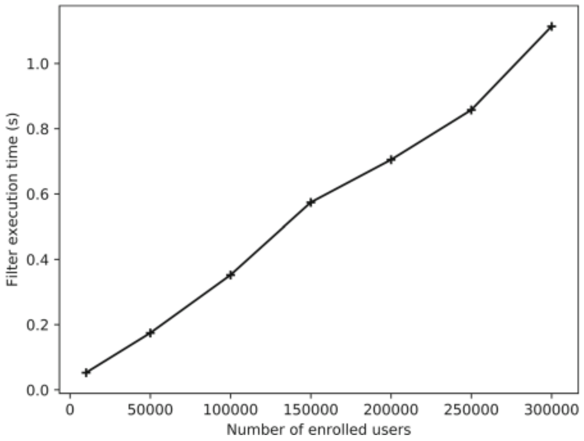

Fig. 11 Average computation time of the improved algorithm

One should, therefore, tune the parameters R and k to achieve the concrete desired security requirements of the deployed solution. It should also be mentioned, that the introduced counter introduces an average number of possible genuine identification requests L that an honest user can perform each year per device. This number is given by formula (3), where RECALL corresponds to the top k recall of the filters. One should therefore also consider the concrete user-friendliness requirements when selecting R.

$$L = \left\lfloor \frac{1}{1 - \text{RECALL}} * R \right\rfloor \qquad (3)$$

6.2.3 *Identification as a specific user*

An adversary's goal in the third attack is to be identified as a specific user, implying that the adversary needs to manipulate the metadata associated with the identification request to have the targeted candidate included in the reduced candidate set. This is, however, only a practical problem. The major challenge is then to defeat the biometric matcher since the template provided by the adversary needs to be close enough to the target's template to be accepted, implying that the adversary only has a probability of FAR to succeed in one attempt. The probability P of succeeding within T years using D number of devices is then given by Formula (4). The parameter R can be tuned to reach different security requirements. However, this attack assumes that the adversary is capable of providing a large number of different templates generated from different fingerprints. Given the security assumptions in Section 3, the adversary then needs to rely on manual methods since it should be unfeasible to manipulate the trusted execution environment on the devices.

$$P = 1 - (1 - \text{FAR})^{D*R*T} \qquad (4)$$

7 Location privacy

As concluded in Section 4, some users may have concerns about sharing location information, therefore, motivating the conducted study into privacy-preserving techniques. An optimal method shall provide sufficient privacy, meanwhile, maintaining adequate discrimination power to permit efficient pre-filtering. The investigated techniques in this paper include a noise-based method [2] as well as a transformation-based method inspired by Gutscher [14].

The noise-based method builds on sequential application of operators aimed at increasing uncertainty associated with location information. Simulations revealed, however, that a noise-based approach building on operators suggested by Ardagna et al. [2] had a severe performance impact. The reason behind the performance issue is that the method inflicts loss of spatial information used by the filtering algorithm.

An investigation of the latter method was partly carried out to assess the feasibility of constructing a privacy-enhancing technique for location information, under the hypothesis that it is possible to derive a stable key from a fingerprint [5, 27]. The key derivation shall only demand the user's biometrics to provide a stable key in a user-friendly manner, implying that a feature extraction aimed at extracting invariant features should be carried out. The derived key can then be used in a key-dependant location transformation, thus providing a privacy-enhancing technique since an adversary needs to be in possession of the used key to deobfuscate a position. The conducted investigation assumed that such a key could be derived and investigated the performance impact of a key-dependant transformation. Simulations revealed maintained performance under the assumption that the key derivation method is completely stable. Utilization of a fingerprint-based key derivation technique is, however, anticipated to have a negative performance impact motivated by reported results [5, 27], thus also affecting the provided user-friendliness.

8 Related work

Related work includes several studies [7, 23, 39] investigating the feasibility of biometric-based indexing schemes to accomplish significant candidate reductions, resembling some of the fundamental idea behind the proposed metadata-based pre-filtering. The state-of-the-art indexing scheme proposed by Cao et al. [7] produces impressive results even for quite large populations. However, different from the solution in [23], it does not protect the biometrics index. We suggest a fundamental different approach to the indexing problem using metadata such as name, location and device ID. Such data reveals personal information but not the biometric data as such. Furthermore, it allows identification in larger user groups when the biometric information available is very limited.

Another flavor of previous work includes studies [1, 11] reviewing the feasibility of using multimodal biometrics to achieve enhanced identification performance. The obtained results from these studies indicate that, for small end-user populations, it is indeed possible. In this paper, we evaluated our system using large populations.

Another branch of related work includes studies [9, 42, 43] using soft biometrics or what is sometimes referred to as pruning to reduce the search space in a pre-processing step. This paper complements these studies by using additional information.

A significant number of previous studies have investigated the feasibility of using location information to perform user identification. One approach investigated by Rossi et al. [32] builds on the existence of trajectories. Another approach reviewed by Naini et al. [25] uses histograms to match a query histogram against a set of histograms. These approaches are not suitable for our application since only a single position will be available when using a new device. A promising approach, as investigated by several studies [31, 47] is to use recorded location information to construct a frequency-based model. This approach is adopted and evaluated using simulations of the Swedish population.

Location privacy is extensively studied in the literature[21]. Numerous different attack models and solutions have been proposed. Similar to several previous approaches, we suggest using location obfuscation [2, 14] to protect the user location privacy. In particular, similar to the solutions in [14], we suggest a coordination transform but with the help of a key derived from the biometric data.

9 Conclusions and future work

We have presented a biometric-based service capable of providing secure identification in large end-user populations. The method works for an identification scenario where a user uses a new device or a device temporarily. The proposed solution utilized metadata, as discussed in the paper, to reduce the risk of false matches, which is a critical step in the journey toward a more user-friendly central identification service. The main advantage with this approach is that the user is, in most, not required to present any data except for the biometric data during the identification session. Furthermore, in those cases, the user indeed is requested to enter additional data; this data is such data that the user can easily remember and quickly enter such as name and/or age. In addition, our proposed filtering mechanism is robust against typing errors. This is a major advantage compared to a system that will require a user to enter a unique user ID without any errors at all.

Different metadata types, as well as different filters, were developed and evaluated to investigate the performance of the different filters. We showed that our incremental filtering approach is possible to use for very large user populations with good performance.

One thing that must be taken into consideration, when selecting a simulation-based evaluation approach, is that the simulation must be grounded in statistics to allow realistic evaluations. In this paper, we used statistics available from scientific articles, as well as, governmental institutes to ensure realistic simulations. It should also be noted, one last time, that we selected to evaluate the adopted approach for the Swedish population. Nevertheless, we believe that this approach should be applicable to a wide range of different populations and future work could, therefore, be carried out to validate such a belief.

Another interesting insight is that future improvement could be carried out in two directions. The biometric matcher could be improved to gain performance. It is also the case that the filters can be further improved to gain performance from another direction. The performance can both include better computation times, as well as, increased security.

Endnotes

[1] https://www.fingerprints.com/technology/hardware/sensors/fpc-touch-sensor-series/

[2] http://www.goodix.com/news/detail1199.html

[3] https://www.fingerprints.com/asset/assets/downloads/fingerprints-biometric-technologies-whitepaper-2017-revb.pdf

[4] This is obvious only the case if we not require the user to enter a unique user ID like in the Adhaar system.

[5] Due to the fact that a small fingerprint sensor is used, only a part of the finger will be possible to catch during identification.

[6] https://sourceafis.machinezoo.com/

[7] 25% typing error is much more than one actually would expect for a normal user, but we have used these high figure in the simulation to represent a worst case scenario.

Acknowledgements
We would like to thank Fingerprint Cards AB in Malmö for providing the prerequisite for the experimental work as well as access to biometric matching data for mobile fingerprint sensors.

Funding
This work has partly been conducted at Fingerprint Cards AB in Malmö.
This research work has been carried out with funding from the Strategic Swedish Foundations (SSF) in Stockholm under grant number RIT17-0032.
This research work has been carried out with funding from the EU 2020 project CloudiFacturing under grant number 768892.

Availability of data and materials
All the data used in the study has been gathered from public Swedish statics. The material is publicly available in the references given in the article, i.e., references: [34–36] and [37].

Authors' contributions
The main author, CG, came up with the basic idea behind the problem and approach presented in this article. He also was the main responsible for the scientific methodology and quality. Second author, MR, has made most of the work with the practical implementation and evaluation of the approach. The simulator framework was jointly developed by MR and NJ. All authors read and approved the final manuscript.

Authors' information
Christian Gehrmann is adjunct professor in computer security at Lund University. He received his PhD in 1997, also at Lund University with a thesis on information theoretical secure authentication codes. He has been active in research and development of secure computer and communication systems for more than 20 years. He has numerous scientific publications and patents in these areas.
Marcus Rodan and Niklas Jönsson did there master thesis project at Lund University in 2018 under supervision of the main author, Christian Gehrmann.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Received: 10 December 2018 Accepted: 9 May 2019
Published online: 06 June 2019

References

1. K. Aizi, M. Ouslim, A. Sabri, in *2015 4th International Conference on Electrical Engineering (ICEE)*. Remote multimodal biometric identification based on the fusion of the iris and the fingerprint, (2015)
2. C. A. Ardagna, M. Cremonini, E. Damiani, S. D. C. di Vimercati, P. Samarati, in *Proceedings of the 21st Annual IFIP WG 11.3 Working Conference on Data and Applications Security*. Location privacy protection through obfuscation-based techniques, (2007), pp. 47–60
3. Y. Baba, H. Suzuki, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, vol. 2*. How are spelling errors generated and corrected?: A study of corrected and uncorrected spelling errors using keystroke logs, (2012), pp. 373–377
4. G. Bae, H. Lee, S. Son, D. Hwang, J. Kim, in *2018 IEEE International Conference on Consumer Electronics (ICCE)*. Secure and robust user authentication using partial fingerprint matching (IEEE, 2018)
5. S. Barman, S. Chattopadhyay, D. Samanta, in *2014 International Conference on High Performance Computing and Applications (ICHPCA)*. Fingerprint based symmetric cryptography (IEEE, 2014)
6. A. Bhattacharya, S. K. Das, Lezi-update: An information-theoretic framework for personal mobility tracking in pcs networks. Wirel. Netw. **8**(2/3), 121–135 (2002)
7. K. Cao, A. K. Jain, in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. Fingerprint indexing and matching: An integrated approach, (2017)
8. R. Cappelli, M. Ferrara, A. Franco, D. Maltoni, Fingerprint verification competition 2006. Biom. Technol. Today. **15**, 7–9 (2007). https://doi.org/10.1016/S0969-4765(07)70140-6
9. A. Dantcheva, C. Velardo, A. D'Angelo, J. L. Dugelay, Bag of soft biometrics for person identification. Multimed. Tools Appl. **51**(2), 739–777 (2010)
10. A. Das, J. Bonneau, M. Caesar, N. Borisov, X. Wang, in *21st Annual Network and Distributed System Security Symposium*. The tangled web of password reuse, (2014)
11. M. O. Derawi, D. Gafurov, R. Larsen, C. Busch, P. Bours, in *2010 2nd International Workshop on Security and Communication Networks (IWSCN)*. Fusion of gait and fingerprint for user authentication on mobile devices, (2010)
12. K. Dharavath, F. A. Talukdar, R. H. Laskar, in *2013 IEEE International Conference on Computational Intelligence and Computing Research*. Study on biometric authentication systems, challenges and future trends: A review (IEEE, 2013)
13. A. P. Felt, S. Egelman, D. Wagner, in *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices - SPSM '12*. I've got 99 problems, but vibration ain't one (ACM Press, 2012)
14. A. Gutscher, in *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium*. Coordinate transformation - a solution for the privacy problem of location based services? (2006)
15. A. Guttman, R-trees: A dynamic index structure for spatial searching. SIGMOD Rec. **14**(2), 47–57 (1984)
16. S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, A. Varshavsky, in *Pervasive Computing, Pervasive'11*, ed. by K. Lyons, J. Hightower, and E. M. Huang. Identifying important places in people's lives from cellular network data, (2011), pp. 133–151
17. A. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition. IEEE Trans. Circ. Syst. Video Technol. **14**(1), 4–20 (2004)
18. M. Kihl, P. Ödling, C. Lagerstedt, A. Aurelius, in *International Congress on Ultra Modern Telecommunications and Control Systems*. Traffic analysis and characterization of internet user behavior, (2010), pp. 224–231
19. V. Kotwal, S. Parsheera, A. Kak, in *2017 ITU Kaleidoscope: challenges for a data-driven society (ITU K)*. OPEN data & digital identity: Lessons for aadhaar, (2017)
20. Y. Li, H. Wang, K. Sun, in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. A study of personal information in human-chosen passwords and its security implications (IEEE, 2016)
21. B. Liu, W. Zhou, T. Zhu, L. Gao, Y. Xiang, Location privacy and its applications: A systematic study. IEEE Access. **6**, 17606–17624 (2018)
22. Y. A. de Montjoye, C. A. Hidalgo, M. Verleysen, V. D. Blondel, Unique in the crowd: The privacy bounds of human mobility. Sci. Rep. **3**(1) (2013)

23. T. Murakami, T. Ohki, Y. Kaga, M. Fujio, K. Takahashi, Cancelable indexing based on low-rank approximation of correlation-invariant random filtering for fast and secure biometric identification. IEEE Access. **7**, 45563–45582 (2018)
24. F. M. Naini, J. Unnikrishnan, P. Thiran, M. Vetterli, Where you are is who you are: User identification by matching statistics. IEEE Trans. Inf. Forensic. Secur. **11**(2), 358–372 (2016)
25. F. M. Naini, J. Unnikrishnan, P. Thiran, M. Vetterli, Where you are is who you are: User identification by matching statistics. IEEE Trans. Inf. Forensic. Secur. **11**(2), 358–372 (2016)
26. A. Noulas, S. Scellato, C. Mascolo, M. Pontil, in *Fifth international AAAI conference on weblogs and social media*. An empirical study of geographic user activity patterns in foursquare (AAAI Press, Palo Alto, 2011), pp. 70–573. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2831
27. G. Panchal, D. Samanta, S. Barman, Biometric-based cryptography for digital content protection without any key storage. Multimedia Tools Appl., 1–22 (2017). https://doi.org/10.1007/s11042-017-4528-x
28. S. Prabhakar, S. Pankanti, A. Jain, Biometric recognition: security and privacy concerns. IEEE Secur. Priv. Mag. **1**(2), 33–42 (2003)
29. C. Rathgeb, F. Breitinger, C. Busch, in *2013 International Conference on Biometrics (ICB)*. Alignment-free cancelable iris biometric templates based on adaptive bloom filters, (2013)
30. A. R. A. Raziff, M. N. Sulaiman, N. Mustapha, T. Perumal, in *2016 IEEE Conference on Open Systems (ICOS)*. Gait identification using one-vs-one classifier model, (2016)
31. L. Rossi, M. Musolesi, in *Proceedings of the second edition of the ACM conference on Online social networks - COSN '14*. It's the way you check-in, (2014)
32. L. Rossi, J. Walker, M. Musolesi, Spatio-temporal techniques for user identification by means of GPS mobility data. EPJ Data Sci. **4**(1), 1–16 (2015)
33. M. Sandhya, M. V. Prasad, R. R. Chillarige, Generating cancellable fingerprint templates based on delaunay triangle feature set construction. IET Biom. **5**(2), 131–139 (2016)
34. SCB, Befolkningstäthet (invånare per kvadratkilometer) per tätort 2017 (2017). http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101C/BeftathetkvkmT/?rxid=78383e5a-efb2-407c-817f-37939848a29d. Accessed Apr 2018
35. SCB, Efternamn med minst 10 bärare bland folkbokförda 31 december 2017 (2017). http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0001__BE0001G/BE0001ENamn10/?rxid=f45f90b6-7345-4877-ba25-9b43e6c6e299. Accessed Apr 2018
36. SCB, Förnamn med minst 10 bärare bland folkbokförda 31 december 2017 (2017). http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0001__BE0001G/BE0001FNamn10/?rxid=f45f90b6-7345-4877-ba25-9b43e6c6e299. Accessed Apr 2018
37. SCB, Medelfolkmängd (efter födelseår) efter region, ålder och kön. År 2017 (2017). http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101D/MedelfolkFodelsear/?rxid=059feb7c-7c33-4103-88e0-a91d9db9d207. Accessed Apr 2018
38. C. Su, Y. Zhu, in *2017 IEEE International Conference on Communications (ICC)*. Using personal information to aid in guessing passwords of chinese webs (IEEE, 2017)
39. Y. Su, J. Feng, J. Zhou, Fingerprint indexing with pose constraint. Pattern Recog. **54**, 1–13 (2016)
40. K. Takahashi, K. Naganuma, Unconditionally provably secure cancellable biometrics based on a quotient polynomial ring. IET Biom. **1**(1), 63 (2012)
41. V. Taneski, M. Hericko, B. Brumen, in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Password security - no change in 35 years? (IEEE, 2014)
42. C. Velardo, J. Dugelay, in *2012 13th International Workshop on Image Analysis for Multimedia Interactive Services*. Improving identification by pruning: A case study on face recognition and body soft biometric, (2012), pp. 1–4. https://doi.org/10.1109/WIAMIS.2012.6226747
43. C. Velardo, J. L. Dugelay, Improving identification by pruning: a case study on face recognition and body soft biometric. Tech. Rep. EURECOM+3593, Eurecom (2012)
44. W. W. Cohen, P. Ravikumar, S. E. Fienberg, A comparison of string distance metrics for name-matching tasks. **2003** (2003)
45. W. J. Wong, A. B. Teoh, M. D. Wong, Y. H. Kho, Enhanced multi-line code for minutiae-based fingerprint template protection. Pattern Recogn. Lett. **34**(11), 1221–1229 (2013)
46. Z. Yang, J. Yu, M. Kitsuregawa, in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, vol. 3*. Fast algorithms for top-k approximate string matching, (2010), pp. 1467–1473
47. D. Zhao, J. Ma, X. Wang, X. Tian, in *Web-Age Information Management*. Personalized location anonymity - a kernel density estimation approach, (2016), pp. 52–64
48. C. Zhou, N. Bhatnagar, S. Shekhar, L. Terveen, in *2007 IEEE 23rd International Conference on Data Engineering Workshop*. Mining personally important places from GPS tracks (IEEE, 2007)