Preserving Privacy in Caller ID Applications

Tamara Stefanović(✉) and Silvia Ghilezan

University of Novi Sad, Novi Sad, Serbia
{tstefanovic,gsilvia}@uns.ac.rs

Abstract. Caller identification (or Caller ID) is a telephone service that transmits a caller's phone number to a receiving party's telephony equipment when the call is being set up. Besides the telephone number, the Caller ID service may transmit a name associated with the calling telephone number. The appearance of the first Caller ID devices caused suspicion among the users and the public because of the potential security and privacy issues that the caller identification may cause. Privacy issues apply to users of the Caller ID applications, but also to non-users whose phone numbers are stored in the database of some Caller ID application. The emergence of the data privacy laws has led discussions on the privacy policies of Caller ID applications and their compliance with the law. In this paper we investigate two Caller ID applications, Truecaller and Everybody, and compliance of their privacy policies with the data privacy laws, especially the GDPR, the ePrivacy Directive and the ePrivacy Regulation. Further, we deal in more detail with the data privacy problem of non-users and we give the connection between those problems, and the inverse privacy problem. In order to solve the privacy problem of non-users, we develop the mathematical model based on the notions of privacy variables and sensitivity function. Finally, we discussed open questions related to the identity protection of Caller ID app users and non-users, and their trust in Caller ID apps.

Keywords: Caller ID applications · Privacy policy · GDPR · Inverse privacy · Name sensitivity · Privacy variables

1 Introduction

Caller identification (or Caller ID) is a telephone service, available in analog and digital phone systems and in most Voice over Internet Protocol (VoIP) applications, that transmits a caller's phone number to a receiving party's telephony equipment when the call is being set up [17]. Besides the telephone number, the Caller ID service may transmit a name associated with the calling telephone number. Even before the proliferation of mobile phones, the appearance of the

Published by Springer Nature Switzerland AG 2021
M. Friedewald et al. (Eds.): Privacy and Identity 2020, IFIP AICT 619, pp. 151–168, 2021.
https://doi.org/10.1007/978-3-030-72465-8_9

first Caller ID devices caused suspicion among the users and the public. The New York Times asked the question whether the Caller ID was a "friend" or a "foe" back in 1992 [18]. The reason for the suspicion lies in the potential security and privacy issues that the caller identification may cause.

Since then, the invention of smartphones and the success of the first Caller ID applications have further developed the problem. Despite this, the new functionalities of Caller ID applications have attracted a large number of users and in recent years, the use of Caller ID applications is constantly increasing. Currently, some of the most commonly used Caller ID applications are Truecaller[1], Hiya[2], Everybody[3], etc. These kind of applications provide their users with the following features: identification of incoming calls from the unknown numbers, blocking calls and messages, call recording, search function and so on. There are a lot of benefits associated with caller identification. From the business perspective, it allows companies to better understand where the majority of their calls are coming from, and in that manner, they can find where most of their interested clients come from. From the perspective of an average person, caller identification is useful because it can protect from harassment. In addition to the good sides of caller identification, there are also some concerns that Caller ID leads to problems with confidentiality, privacy and individual's safety. Users and non-users of Caller ID applications are facing those problems equally. The question is whether it is possible to find a way to use these applications, but also to ensure the privacy and safety of others.

Results. In this paper, the focus is on the privacy problem of non-users in Caller ID applications. We first explore the privacy policies of two Caller ID applications, Truecaller and Everybody, and their compliance with data privacy laws such as GDPR, ePrivacy Directive and ePrivacy Regulation. We further argue that the problem of non-user privacy falls into the scope of inverse privacy. Building on this connection, we introduce the notion of name sensitivity and privacy variables and solve the problem of non-user privacy. Finally, we discuss open questions related to the identity protection of Caller ID app users and non-users, the place for storing privacy preferences about a phone number and linked data in Caller ID apps.

Paper Outline. The rest of the paper is structured as follows. Section 2 presents the main Caller ID app features and describes how the Caller ID applications violate individual's privacy. Section 3 investigates privacy policies of two Caller ID applications and their compliance with the data privacy laws. Section 4 gives a connection between the privacy problem of non-users and the inverse privacy problem. Section 5 introduces the name sensitivity and the privacy variables. Section 6 and Sect. 7 discuss and conclude the paper.

[1] https://www.truecaller.com/.
[2] https://hiya.com/.
[3] http://www.evrbd.com/.

2 Background and Motivation

In this section, we recall the main features of Caller ID applications and describe how they may violate individual's privacy.

2.1 Caller ID App Features

After the installation of a Caller ID app, the app creates and stores some information (user's profile) about the user. The user's profile includes his/her name, phone number, and/or e-mail address. Information from user's address book (contact names, phone numbers, e-mail addresses, etc.), call lists and messages is also included in the user's profile. The information provided by the user can change at any time (e.g., the user can change his/her name, the user can change names and phone numbers in his/her address book, and also add or delete some contacts) so the database storing this information is dynamic. Besides information provided by the user, the user's profile may include information gathered by the app itself relating to activities of the user. Such activities includes interactions and connections with other users and non-users (e.g., call frequency). A non-user is a person whose phone number is stored in the address book of a user of the Caller ID application, but he/she is not using the Caller ID app by himself/herself. The relationship between individuals is modeled by a social graph. A node of the social graph corresponds to Caller ID app user, or a non-user, while edges connecting two nodes correspond to a relationship between two individuals.

The main functionality of these applications is caller identification. To be specific, when a user receives an incoming call, the phone can display the caller's phone number and the name associated with this number if available. If the caller's phone number and a corresponding name are available in the user's address book, the phone displays the corresponding name in addition to the phone number. If the caller's phone number and a corresponding name are not available in the user's address book, the app searches its database and in case the name associated with the given number exists in the database, it forwards the name to the user. The search is based on the user's interaction with other users and non-users.

The Caller ID apps also provide a search function. Particularly, a user can access an address book stored in the Caller ID app's database, look up a contact in the address book and connect with the contact. The search can be based on the name or on the phone number.

For better functionality, Caller ID applications may require location access. Also, various applications have the additional option for interacting with a third party services, like Facebook, Twitter and so on, in which case those third party services may provide the app with additional information.

2.2 Violating Individual's Privacy

The definition of data privacy itself is notoriously complex, so we have to put it in a particular context. For further analysis, we will think of privacy in terms of transfer

of personal information. The term "personal information" refers to any information related to an identified or identifiable natural person. As names and telephone contacts actually enable the adequate identification of natural individuals, there is no doubt that such information actually constitutes personal data.

Once you install a Caller ID application, it will have access to your address book, call list, messages and so on. Additional features require access to your microphone, camera, and location. During the installation, you confirm that all the persons from your address book have given their consent to share their phone number and the name under which their number is stored in your address book. The problem is that the required consent is not strictly defined, and the users are not required to provide any proof of given consent. This leads to the conclusion that non-users do not have any control over their personal data, and they are not even aware of the collection and publication of their personal data.

In addition to the fact that Caller ID apps violate individual's privacy, participating in a Caller ID app database can greatly endanger individual's safety.

Danger for Users. Nowadays using just Caller ID is not enough to prove who is the real caller since there are several ways to manipulate the caller identity. As Caller IDs are vulnerable to spoofing attacks, they have been used in a variety of misuse and fraud incidents:

- credit card frauds, when credit card companies mistakenly authenticate newly issued cards or stolen cards by phone calls of a fake credit card holder, like in [19], or when a person gets a false call from a bank in order to give personal information, like credit card number [15];
- swatting, frauds representing an attempt to trick an emergency service with false reporting of an incident, which can even end tragically [21].

There have been several attempts to detect Caller ID spoofing attacks like [13, 22].

Danger for Non-users. The applications we cherish can endanger not only us but the people around us. They are in even more dangerous situation, because they are not aware that their data is collected. The example of a journalist betrayed by an application she was not even using, can be found in [1]. Namely, she was working on sensitive TV show stories like human trafficking, drug cartels and so on. That is why she was not using social media and she did not appear on screen. It was very important for her to protect her anonymity. She traveled to a foreign country for a story. She bought a local SIM card and she was using it only for communication with her sources. The people she was investigating were just regular citizens and she didn't have to worry about state surveillance of her communications. It allowed her to be open with her sources about her name, affiliation and intentions. At some point she used her local number to order a cab, and when she entered the cab, the driver already knew the name of the TV show she worked. She could see her name and the name of the TV show she worked next to her phone number on the driver's phone. What happened to her is that one of her sources was using the same Caller ID app as the cab driver. This source tagged her phone number using her name and affiliation, and this

tag became available for all the users of this particular Caller ID app. She didn't use any of Caller ID applications, but her safety was endangered by one of them. Although this example is taken from a web article, the scenario of the example is easily applicable in everyday life.

It should be mentioned that there are some practical methods for individuals to suppress their identity. One of them is prefixing a phone number with a certain code, which is usually characteristic for each country. This action suppresses the caller's number for almost all cell phones. The same can be done in the settings of the phone. Although this is one way individuals can protect themselves, it is not a global solution to the problem. The global solution should be provided by the Caller ID app.

Unlike the problems faced by Caller ID application users, the problem of non-users is insufficiently investigated. Therefore, in this paper, the emphasis will be on the problem of non-users.

3 Compliance with the Data Privacy Laws

Below we will analyze the privacy policies of Caller ID applications, as well as their compliance with the data privacy laws. Given that countries are creating their own data privacy laws, we will focus on the General Data Protection Regulation (GDPR) [5], the ePrivacy Directive [6], and the ePrivacy Regulation [4] because of their territorial scope. As examples for privacy policies of Caller ID applications, we will use Truecaller and Everybody privacy policies [7,24].

3.1 Truecaller and Everybody Privacy Policies

Truecaller. First of all, Truecaller privacy policy states that if you provide them with personal information about someone else, you confirm that they consent to use their information. Further, by enabling the Truecaller Enhanced Search Functionality, you can share with the app contact information from your address book, but only phone numbers, attached names, Google ID's and e-mail addresses. Other information from your address book is filtered away. Sharing the contact information with the Truecaller is optional, and you can make your contact information unavailable for search in the Truecaller database at any time. If you are in the Truecaller database and someone search your name, you will get a message and you may choose whether or not to share your phone number with that person. Also, if you are a non-user of Truecaller and you do not want your personal information to be in their database, you can contact them to unlist you.

Everybody. Everybody privacy policy states that the user, immediately after installation, is asked if he/she wants to share his/hers address book with the Everybody app or not. If accepted, contact information is stored on their server. Contact information includes name, surname, and phone number. The consent for sharing someone else's personal information is not mentioned. Further, contradictory with the first one, they state that non-user information will not be

stored on their server. When it comes to the transfer of personal information of non-users, the Everybody app works as follows. When the non-user first calls a user who does not have his/her phone number in the address book, the application notifies the non-user via SMS and asks him/her to state whether he/she allows further processing and forwarding of his/her name. It is possible at any time to withdraw the consent for sharing information from your address book with the application, and they respect the right of deletion of your personal data from the Everybody's database.

3.2 Compliance with the GDPR, ePrivacy Directive and ePrivacy Regulation

First of all, according to the definition of personal data in the Art.4 of GDPR, there is no doubt that Caller ID applications process the personal data of both their users and non-users. When it comes to users, Caller ID applications indeed protect personal information of their users.

When it comes to users, sharing and processing of user's personal data is lawful because is based on the given consent (according to the Art.6 of GDPR), users have the right to withdraw the consent at any time (according to the Art.7 of GDPR), and also have the right of erasure of their personal data (according to the Art.17 of GDPR).

When it comes to non-users, the situation is a little more complicated. Namely, Caller ID applications obtain data of non-users from a third party, that is, the data is not obtained directly from the data subject. In that case the data subject has the right (according to the Art.14 of GDPR): to be provided with categories of personal data concerned, the purpose of processing, the recipients of the personal data, the information from which source the personal data originates and the right of erasure or restriction of processing of his/her personal data. The information referred above shall be provided at the latest when the personal data are first disclosed. The privacy policies of the mentioned applications are compliant with those requirements. They are also compliant with Art.12 of the e-Privacy Directive 2002/58/EC which requires that data controllers obtain additional consent from data subjects before including their data in directories searchable solely on the basis of telephone number. In recent years these applications have faced numerous criticisms and in the meantime have updated their privacy policies to comply with privacy laws.

The sharing of non-users' personal data by the Caller ID application is legal, but the question is whether they have the right to store this data in their database at first. Although the non-users are entitled to exclude themselves from a Caller ID app database, in order to do so, they must be aware of the existence of that particular Caller ID app. In case they want to be excluded from the databases of all Caller ID applications, they would have to be excluded from each database individually, which is practically impossible. This brings us back to the question of the consent that users should obtain from their contacts for sharing personal data. According to the Art.2 of GDPR, this regulation does not apply to the processing of personal data by a natural person in the course of a purely personal

activity. Processing also includes disclosure by transmission (according to the Art.2 of GDPR). People use the Caller ID app for personal purposes, but sharing contact information that will be stored in a public database goes beyond personal purposes. In that sense, either the GDPR is not applicable in this situation or the Caller ID applications force their users to break the law while they remain clean. Then, in a legal sense, the opponent of the non-user would be the user who forwarded the non-user's personal information without his/her consent, and not the application itself. However, a problem that non-users might face when legally claiming their rights is that they would not have information on exactly which person forwarded their data to the Caller ID app. Caller ID app providers certainly have that information and according to the Art.14 of GDPR they should pass it to non-users, but this action would harm their users.

The ePrivacy Regulation [4] was proposed in 2017 and is still waiting to be adopted (up to December 2020). It would repeal ePrivacy Directive [6] and would be lex specialis to the GDPR [5]. The purpose of this regulation is protection of personal data in electronic communications. It could resolve many questions, because it could apply to all sorts of tech firms providing electronic communication services, and also protect people from unsolicited communications like marketing phone calls. Marketing callers will need to display their phone number or use a special pre-fix that indicates a marketing call. This could raise a question of the primary purposes for using Caller ID applications.

4 The Inverse Privacy Problem

In this section, we recall the notion of inverse privacy and give a connection between the inverse privacy problem and the Caller ID app privacy problem.

4.1 The Inverse Privacy

People have shared their personal data with many public and private institutions for a very long time, but over the past decades these institutions have been given a significant advantage over regular persons in terms of data collection capacity. For this reason there is more and more personal data that is not available to us. This kind of data is called **inversely private data**. Access to this data could benefit individuals and we should be entitled to access our inversely private data.

To be more specific, an item of your personal information is inversely private if some party has access to it but you do not have [10]. The inaccessibility to you of your personal data is the inverse privacy problem. A good solution of the problem should provide you with convenient accessibility to your inversely private data. In order to ensure convenient access, the following three items should be considered:

– *Large database.* On a daily basis we interact with various institutions leaving and creating new inversely private information. The database that would store our inversely private data would be large and it would be growing daily.

- *Sensitive information.* Some of our inversely private information might be too sensitive to be shared with us, and among our inversely private data there might be information that we do not want and should not see. Although it may be hard to believe, even in addition to intelligence data, there are certain scenarios in which our inversely private information should not be available to us. For example, some of our personal information could be a part of the research that examines the risks of getting a certain type of disease. This information should be available to us in many scenarios (if our doctor suspects the possibility of the disease), but not on a daily basis. Another example comes from the perspective of Caller ID apps. To be specific, Caller ID app database can contain some sexist and offensive names associated with our phone number, and it is not convenient to provide us the access to such names.
- *Privacy and Security.* Sharing back our inversely private information has to be secure, ensuring certainty provided by user identity credentials [2], differential privacy [3], multi-party computation [20], and other relevant means.

The question is where to look for a solution to the inversely private problem. The inverse privacy problem is a product of technological progress, and it can be solved only with better technology.

The Biggish Platform. Microsoft Research proposed a solution for the inverse privacy problem in [9]. The solution includes the Biggish platform, a trusted platform for storing your personal data, driven by the idea of securely storing on the cloud. Your data about finance, health, location, etc. would be consolidated on the Biggish platform, and you would access them through various analytics apps. In order to ensure customer trust, these apps should preserve user's privacy and they should not leak data. In that sense, Biggish would be useful not only to individuals, but also to software developers. One more beneficiary of Biggish would be a company or set of companies that host Biggish. The ownership of all that data would require the development of new privacy policies. Biggish is a solution proposal for the inverse privacy problem, and based on that proposal, Microsoft started testing new Bali project in 2019. The test results are not yet published [16].

4.2 Caller ID App Privacy Problem and the Inverse Privacy Problem

We claim that there is a link between the Caller ID app privacy problem and the inverse privacy problem. Our phone number is our personal information, the name under which our phone number is stored in someone's address book is also our personal information but to which we generally do not have access. This leads to the conclusion that the Caller ID app privacy problem falls into the scope of the inverse privacy problem. Nevertheless, it must be stated that in this situation the inverse privacy problem is only relevant from the perspective of the provider of the Caller ID app and it is not been caused by an individual user of the Caller ID app.

To explain the main similarities and differences between our problem and the problem of inverse privacy, we will present *infons*. The notion of infons has been introduced in [11] - an infon is defined as an item of information.

First of all, previous research on the topic of inverse privacy has been based on the assumption that certain infon is personal to only one individual. Our problem is slightly different. The name, under which the telephone number of the person A is stored in the address book of the person B, is also personal to the person B. If the person A were given access to that information, then the privacy of the person B would be also violated. Even if one did not know whose address book a particular name belonged to, it might still be possible to infer from some previous experience.

Also, as we stated before, some of our inversely private information might be too sensitive to be shared with us. Caller ID applications generally contain mechanisms that exclude offensive names, but that doesn't mean that if people have access to a database of names under which their phone number is stored somewhere, they won't see the names they don't like.

Given the link between the Caller ID app privacy problem and the inverse privacy problem, the solution for the inverse privacy problem can be used to solve the Caller ID app privacy problem. Our proposition on how to use the Biggish platform to solve privacy issues in Caller ID applications is as follows. Some part of the Biggish platform could contain a database with names under which our phone number is stored in other people's address books. Given the size of the database and the presence of sensitive data, it would be necessary to provide adequate access.

Access to the database will be two-fold. On the one hand, we will have access to the database, and on the other hand, the Caller ID applications will have access as well (Fig. 1). Also, we should not see which name belongs to which address book, in order to preserve the privacy of the owner of the particular address book. Hiding this information would protect the privacy of the address book owner in many scenarios, but not in general. There are certain scenarios in which, with the help of some additional information, it is possible to find out whose directory a certain name belongs to (for example if any characteristic nickname is in question). This leads to the conclusion that our access to the database does not have to be through the display of the list of names, but through the possibility of certain manipulations over the database. Particularly, we could set our preferences regarding the names. In the first place, we could set our phone number to be private, partially private and non-private. In that way, the information about privacy of a phone number would be stored in one place to which each Caller ID application would have access. Setting our phone number to be private at only one place is much more simple than excluding ourselves from the databases of every Caller ID application separately. It's the same thing with setting the number to be non-private. In that case, the Caller ID applications according to their mechanisms could forward any name associated with our number. The most interesting case is the partially private phone number. In order to explain partial privacy, in the next section, we introduce the notion of privacy variable.

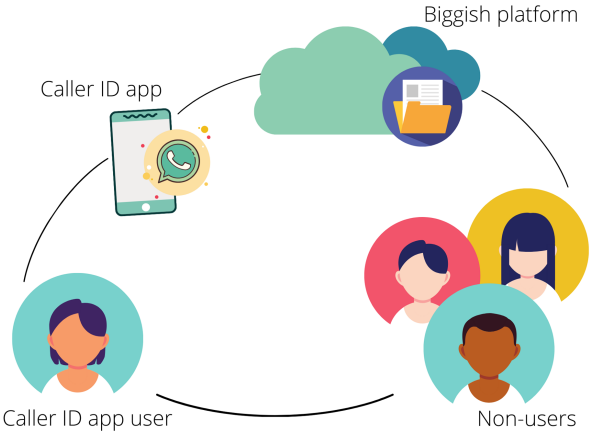

Fig. 1. The connection between users, non-users and the Biggish platform

5 Preserving Privacy in Caller ID Apps

In this section, we introduce privacy variables and name sensitivity function. We further design an algorithm for determining caller identification and providing an adequate privacy preserving name. Finally, we discuss the position of the name sensitivity function and it's use aside the use in Caller ID apps.

5.1 Name Sensitivity and Privacy Variables

As noted before, individuals and the connections between them can be represented by a social graph. Each individual is associated with a set of attributes: name, school, profession, etc. Since the Caller ID applications, in addition to the name, phone number and e-mail, filter all other data from the phonebook and pass only the name to users, we will focus below on the attribute that indicates the name. A Caller ID app social graph is then defined as follows.

Definition 1 (Caller ID App Social Graph). The Caller ID application data is modeled by a graph $G = (V, E, A, P)$, where V is a set of nodes representing users and non-users, $E \subseteq V \times V$ is a set of edges representing relations between users and non-users, A is a set of attributes associated with every node, and P is a set of privacy variables.

A *privacy variable* is a variable associated with the name attribute: name, surname, hometown, school, profession and so on. Those variables can have only two values, 0 for non-private information and 1 for private information. When setting his phone number to be partially private, the person defines values for each of the privacy variables. The motivation for introducing privacy variables is the greater sensitivity of certain names compared to other names. For the

journalist mentioned before, the information about her profession is more sensitive than her name and surname. She could define that the names that include her profession are not shared with the users of Caller ID applications. Below we define a function that represents the sensitivity of a particular name, or the sensitivity of a particular value for a name attribute.

Definition 2 (Name Sensitivity Function). Let X be a name attribute and x any value of the name attribute, sensitivity of x denoted by $S(x)$ is defined by:

$$S(x) = \begin{cases} 0, & \text{if } 1^\circ \text{or } 2^\circ \\ \frac{n}{m}, & \text{if } 3^\circ \\ 1, & \text{if } 4^\circ \text{or } 5^\circ \end{cases} \tag{1}$$

where n is a number of variables that are marked as private, and the value x matches them, m is a total number of variables that are marked as private, and

1° The phone number associated with the name x is non-private.
2° The phone number associated with the name x is partially private, but x does not match any variable that is marked as private.
3° The phone number associated with the name x is partially private and x matches some of the variables that are marked as private.
4° The phone number associated with the name x is private.
5° The phone number associated with the name x is partially private and x matches all the variables that are marked as private.

Example 1. Let the set of privacy variables be $P = \{P_1, P_2, P_3, P_4\}$ where P_1 defines name, P_2 defines town, P_3 defines profession, P_4 defines affiliation and let Alice set P_2, P_3 and P_4 to be private. The sensitivity of the following names associated with her phone number is:

- if $x_1 =$"Alice", $S(x_1) = \frac{0}{3} = 0$;
- if $x_2 =$"Bob", $S(x_2) = \frac{0}{3} = 0$;
- if $x_3 =$"Alice the professor", $S(x_3) = \frac{1}{3}$;
- if $x_4 =$"Bob Novi Sad", $S(x_4) = \frac{1}{3}$;
- if $x_5 =$"Alice the professor Novi Sad", $S(x_5) = \frac{2}{3}$;
- if $x_6 =$"Bob the professor Faculty of Technical Sciences", $S(x_6) = \frac{2}{3}$;
- if $x_7 =$"Alice the professor Faculty of Technical Sciences Novi Sad", $S(x_7) = \frac{3}{3} = 1$;

Example 2. For the journalist, mentioned in Sect. 2, information about her/his profession and affiliation is more sensitive than her/his name. Given the set of privacy variables $P = \{P_1, P_2, P_3\}$ where P_1 defines name, P_2 defines profession, P_3 defines affiliation, and setting P_2 and P_3 to be private, the sensitivity of the following names associated with her/his phone number is:

- if $x_1 =$"Alice", $S(x_1) = \frac{0}{2} = 0$;
- if $x_2 =$"Bob", $S(x_2) = \frac{0}{2} = 0$;

– if x_3 ="Alice the journalist", $S(x_3) = \frac{1}{2}$;
– if x_4 ="Bob the journalist New York Times", $S(x_4) = \frac{2}{2} = 1$.

Each person has his/her own sensitivity function that reflects his/her preferences regarding the privacy of his/her phone number and the names associated with his/her phone number. A finer model could be obtained by defining privacy variables to take values from an interval $[0, 1]$. In this way, each person could even define a certain sensitivity threshold for each variable.

5.2 APPN Algorithm

In order to give the answer to the question when the sensitivity of the name is checked, we design the algorithm that provides adequate privacy preserving name (APPN algorithm) in Fig. 2.
Below we give a detailed description of the algorithm:

Step 1. Obtain a phone number of an incoming call.
Step 2. Check whether the phone number is stored in the user's address book. If the phone number is not stored in the user's address book, skip to the Step 3, either way skip to the Step 6.
Step 3. Check the privacy of the phone number. If the phone number is set to be non-private skip to the Step 4. If the phone number is set to be partially private, check which of the privacy variables are set to be private and skip to the Step 4. If the phone number is set to be private, skip to the Step 6.
Step 4. Single out the names associated with the given phone number from the database, and skip to the Step 5.
Step 5. Check the sensitivity of every name that has been singled out in the Step 4. Select one of the names with zero sensitivity and skip to the Step 6.
Step 6. Present the adequate name of the calling party, or inform the user that it was not possible to find an adequate name.

A problem that has not been considered is the matching of the name with a certain privacy variable. Names, professions, towns can be represented by their abbreviations. They are also characteristic of each speaking area. Further, people tend to use personal indications for labeling contacts in their address books (names like "mom", "dad", etc.). A mitigating circumstance is that Caller ID apps already have their own mechanisms that exclude names based on personal indication and such names are not passed on to their users. However, the mechanism for verifying the matching of names with privacy variables should be further investigated.

5.3 The Position of the Name Sensitivity Function

We will here discuss the position of the name sensitivity function from legal and practical aspects.

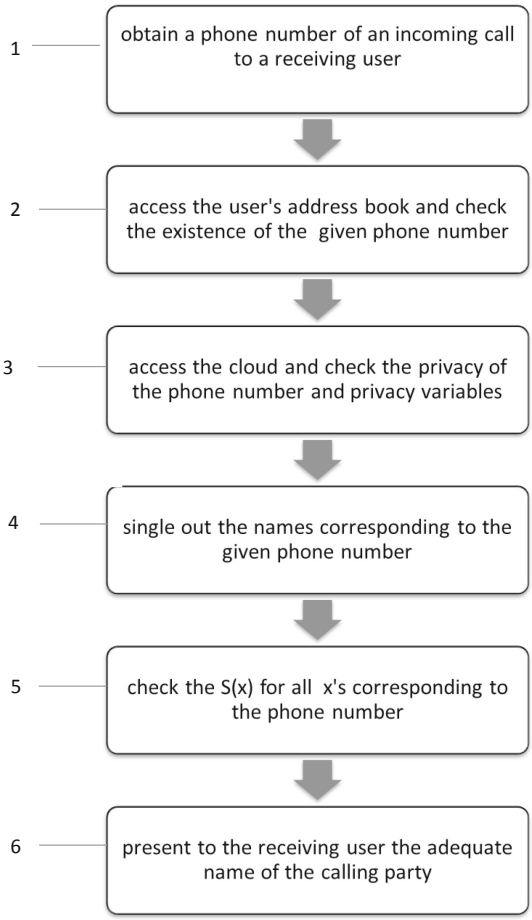

Fig. 2. An algorithm for determining caller identification and providing adequate privacy preserving name.

Legal Side. With respect to the inclusion of the Biggish platform in the solution for Caller ID app privacy problem, the question of the position of the name sensitivity function arises. There are two possibilities. The first one is to include the approach of the sensitivity function in the business model of the Caller ID app provider. The second one is to include the approach of the sensitivity function in the business model of the provider of the Biggish platform. Both business models should comply with the requirements imposed by the GDPR, and data subjects should be aware of their data being processed. In the second case, the Biggish platform becomes the controller of the personal data. The Biggish platform is then obliged to inform data subjects about the processing and to check whether providing access to this data by the Caller ID app provider is according to the data privacy laws. However, when it comes to the way in which the Biggish

platform obtained name information from the Caller ID apps, the responsibility regarding the forwarding of that information is still on each individual Caller ID app.

Practical Side. The algorithm presented above is designed on the basis of the name sensitivity function and privacy variables. The idea is that Caller ID apps can use only names with zero sensitivity (cases where $S(x) = 0$). The remaining two cases of the name sensitivity function do not play an essential role in the algorithm. This leads to the conclusion that the algorithm could be designed only on the basis of privacy variables, and the step 5 of the algorithm could become "choose a name for which no attributes have been set to be private". However, the remaining two cases of the name sensitivity function are applicable in other areas related to data privacy. The social graph that represents Caller ID app data has significant application value. With the help of the name sensitivity function, the graph model could be used to make some statistical analysis of different levels of privacy preferences. The analysis is then usable in tracking down those with the highest privacy settings in a world of privacy nihilism. Companies for data analysis as well as marketing companies are very interested in this type of data.

6 Discussion

The first issue we dealt with is whether the storage and processing of data by the Caller ID apps is lawful. In order to find the answer, we analyzed privacy policies of two Caller ID apps and their compliance with data privacy laws. We concluded that the privacy policies of the analyzed applications respected the privacy of their users and non-users at a large degree, but some questions still remain unresolved.

Caller ID app users decided not to share their address books. What will happen? One of the unresolved questions is whether Caller ID app users have the right to share personal data of their contacts with the Caller ID app they use, and what would happen if users stopped sharing data from their address books. The data from the Caller ID app database mainly comes from the address books of its users (some of the data comes from public databases, from interaction with third party services like Facebook, Twitter, LinkedIn, etc.). When users would not share the data from their address books with Caller ID apps, the way these apps fundamentally work would have to be changed. That would conflict with the purpose of Caller ID apps and the interests of the Caller ID providers.

The right of deletion of your personal data. How to be sure that your personal data is actually removed from the database? The next unresolved question concerns the non-users. The privacy policies of both investigated apps respect the right of non-users to unlist themselves from the app's database and to delete their personal data. Both applications have their own websites. Assuming that you are aware of the existence of these applications, you can contact the app providers via their website in order to unlist you from their database. The questions that remain open are:

– how to be sure that your personal data is actually removed from the Caller ID app database,
– how this action prevents adding your personal data again to the app's database once another user having you as a contact on his/her address book actually allows for such access.

The Biggish platform - one more technology product to trust? Further in the paper we discuss the connection between the inverse privacy problem and the Caller ID app privacy problem. We argue that the solution for the inverse privacy problem, the Biggish platform, can resolve the privacy problem of Caller ID apps. This solution is based on a mediator between users, non-users and caller id apps. In this sense, individuals should trust not only the Caller ID applications, but also the Biggish platform, in terms of storing and processing their data and in terms of respecting their preferences (about privacy of their phone number and the names associated with their phone number). In order to achieve high level of trust, the Biggish platform should function in accordance with all GDPR requirements. In this sense, the main principles of privacy by design and privacy by default should be respected (according to the Art.25 of GDPR), which requires a deeper analysis of this problem.

The Biggish platform-failure for the Caller ID apps? The next question regarding the Biggish platform is would the existence of such platform that stores all data of all users and non-users would lead to the failure of all Caller ID apps. If the Caller ID applications had the access on the Biggish platform only to preferences and not to the complete data related to a certain phone number, they would continue to function as before. Competition between them would still be based on the fact who has a larger database and better search mechanisms. Of course, since it is not yet known how the platform will work, it is not possible to say with certainty how its existence will affect the Caller ID applications and also many other applications that operate on personal data. Caller ID was invented in the 70's and since then many of its functions have evolved due to the development of technology. There is a possibility that the Caller ID application will change their functions over time and with the discovery of new technologies.

A safe place for storing privacy preferences of a phone number. Where to find it? Another question related to the Biggish platform is what would happen if it was suddenly removed and what would it mean for the operation of the Caller ID applications. We argue that is much more simple to set preferences about privacy of a phone number at only one place than to exclude ourselves from the databases of every Caller ID application separately. That place must be protected from sudden deletion and illegal access. The solution can be found in distributed databases [23] or blockchain technology [8].

Identification of individuals. Is it possible by combining public and semi-private information? Finally in the paper we introduce the notion of privacy variables and name sensitivity. There are certain open questions on this topic. We argue that certain names have greater sensitivity compared to other names, and that allowing someone to set his/her phone number to be partially private (to set some privacy variables to be 1) can greatly increase person's security. On

the other hand, we state that the Caller ID applications collect data from public databases and through the interaction with social media, in addition to the address books of their users. The question that arises is whether it is still possible to uniquely identify individuals by combination of public and semi-private information. The answer is yes, but individuals would be more aware of that possibility than they are now. At the moment the majority of the population is not sufficiently familiar with the ways in which identification of individuals can be done and in general with their rights based on the data privacy laws. Having the ability to set phone number privacy preferences would increase individuals' awareness of their rights at least in some way.

Informations from multiple sources. How should Caller ID apps treat them? The last question we will discuss relates to linked data. To be more specific, the question is how should Caller ID apps treat combined information from multiple sources while ensuring privacy protection. On the one hand, the Caller ID app has the information about privacy of our phone number collected from the Biggish platform, and on the other hand, it has many other informations about our phone number collected from their users, public databases and social media. In order to provide privacy protection in this case, a logic-based model can be implemented like in [12].

7 Conclusion and Further Work

Our aim is to give a comprehensive view on data privacy problems caused by the Caller ID applications. To this end, we first explored the privacy policies of two Caller ID applications, Truecaller and Everybody, and their compliance with data privacy laws such as GDPR, ePrivacy Directive and ePrivacy Regulation. It can be concluded that the creators of the Caller ID applications have updated their privacy policies to comply with privacy laws, but also that some issues related to the privacy of non-users have not yet been resolved. We further argue that the problem of non-user privacy falls into the scope of inverse privacy. Building on this connection, we introduce the notion of name sensitivity and privacy variables and solve the problem of non-user privacy. Finally, we discussed open questions related to the trust of the Caller ID app users, the place for storing privacy preferences about a phone number and linked data in Caller ID apps.

The above discussion motivated us to continue our research on the topic of blockchain with the intention of finding a safe way for storing privacy preferences about a phone number. We also plan to extend the mathematical model presented in [12], in order to answer the question how should Caller ID apps treat combined information from multiple sources while ensuring privacy protection.

Another direction of the further research will be the privacy problem on the server side, where the Caller ID app provider learns significant information about the social graph. Due to the fact that social graph data has significant application value, this data is increasingly being made public or sold to third parties. The data is then used for commercial and research purposes. In this paper we have seen that the data of the Caller ID applications contains private

information of individuals, so the question is how to prevent individual privacy disclosure while publishing this data. The ongoing and future work will consider anonymization and privacy preserving techniques along the lines [14] and [25].

Acknowledgement. We would like to thank the referees for their careful reading of our submission. Their comments and suggestions have guided us to revise and improve the paper. Furthermore, we would like to thank the organisers and session chairs of the IFIP Summer School for the feedback of the discussions which highly broadened our research. Last but not least, we would like to thank the IFIP Summer School participants for valuable discussions during the summer school.

This work has been partially supported by the Science Fund of the Republic of Serbia under grant AI4TrustBC (6526707).

References

1. Betrayed by an app she had never heard of https://ifex.org/serious-privacy-concerns-raised-about-the-app-truecaller/. Accessed 3 Sept 2020
2. Bosworth, K., Gonzalez Lee, M.G., Jaweed, S., Wright, T.: Entities, identities, identifiers and credentials - what does it all mean? BT Technol. J. **23**(4), 25–36 (2005). https://doi.org/10.1007/s10550-006-0004-2
3. Dwork, C.: Differential privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) ICALP 2006. LNCS, vol. 4052, pp. 1–12. Springer, Heidelberg (2006). https://doi.org/10.1007/11787006_1
4. European Commission: Proposal for a Regulation on Privacy and Electronic Communications (ePrivacy Regulation). https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications
5. European Parliament and Council of the European Union: Regulations (EU) 2016/679 of the European Parliament and of the Council - general data protection regulation (GDPR) (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj
6. European Parliament, Council of the European Union: Directive 2002/58/EC of the European parliament and of the council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (directive on privacy and electronic communications) (2002). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2002:201:TOC
7. Everybody privacy policy. http://www.evrbd.com/policy.html#. Accessed 5 Sept 2020
8. Ghosh, J.: The blockchain: opportunities for research in information systems and information technology. J. Global Inf. Technol. Manage. **22**(4), 235–242 (2019). https://doi.org/10.1080/1097198X.2019.1679954
9. Gurevich, Y., Haiby, N., Hudis, E., Wing, J., Ziklik, E.: Biggish: a solution for the inverse privacy problem. Technical report. MSR-TR-2016-24, Microsoft, May 2016. https://www.microsoft.com/en-us/research/publication/230-biggish-solution-inverse-privacy-problem/
10. Gurevich, Y., Hudis, E., Wing, J.M.: Inverse privacy. CoRR abs/1510.03311 (2015). http://arxiv.org/abs/1510.03311
11. Gurevich, Y., Neeman, I.: Logic of infons: the propositional case. ACM Trans. Comput. Log. **12**(2), 9:1–9:8 (2011). https://doi.org/10.1145/1877714.1877715
12. Jaksic, S., Pantovic, J., Ghilezan, S.: Linked data privacy. Math. Struct. Comput. Sci. **27**(1), 33–53 (2017). https://doi.org/10.1017/S096012951500002X

13. Mustafa, H., Xu, W., Sadeghi, A.R., Schulz, S.: You can call but you can't hide: detecting caller id spoofing attacks. In: Proceedings of the International Conference on Dependable Systems and Networks, June 2014. https://doi.org/10.1109/DSN.2014.102
14. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. CoRR abs/0903.3276 (2009). http://arxiv.org/abs/0903.3276
15. News, A.: https://abcnews.go.com/GMA/Consumer/story?id=3305916. Accessed 2 Sept 2020
16. Otokiti, E.: Microsoft tests new personal data bank, project Bali. https://channels.theinnovationenterprise.com/articles/microsoft-tests-new-personal-data-bank-project-bali. Accessed 14 Oct 2020
17. Papakipos, M.N., Walkin, B.M.: Caller identification using social network information, August 2012
18. Ramirez, A.: Caller id: Consumer's friend or foe? (1992). https://www.nytimes.com/1992/04/04/news/caller-id-consumer-s-friend-or-foe.html
19. Schneier, B.: https://www.schneier.com/blog/archives/2006/03/caller_id_spoof.html. Accessed 2 Sept 2020
20. Schoenmakers, B.: Multiparty Computation, pp. 812–815. Springer, Boston (2011). https://doi.org/10.1007/978-1-4419-5906-5_7
21. Stevens, M., Chow, A.R.: https://www.nytimes.com/2018/11/13/us/barriss-swatting-wichita.html. Accessed 2 Sept 2020
22. Sukma, N., Chokngamwong, R.: One time key issuing for verification and detecting caller id spoofing attacks. In: 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1–4 (2017)
23. Tan, K.L.: Distributed Database Systems, pp. 894–896. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-39940-9_701
24. Truecaller privacy policy. https://www.truecaller.com/privacy-policy. Accessed 5 Sept 2020
25. Xie, Y., Zheng, M.: A differentiated anonymity algorithm for social network privacy preservation. Algorithms **9**(4), 85 (2016). https://doi.org/10.3390/a9040085