Analysis of Identity Separation Against a Passive Clique-Based De-anonymization Attack

Gábor Gy. Gulyás and Sándor Imre

***Abstract*—Most of today's online social networking services have a flat structure, i.e., these services only allow a single choice of connection type (usually called "friends") for their users, and lack the functionality of identity separation. However, identity partitioning allows users to group their contacts, to share different or even diverse information, and therefore offer privacy protection against third parties looking to re-identify users in sanitized social graph data. In this paper, we analyze the protective strength of identity separation against these types of structural de-anonymization attacks by introducing a statistical user behavior model and defining attack failure probability formally. It turns out from simulations and the parameter analysis of the model that in case of even a relatively small number of users applying identity separation, an attacker is likely to fail.**

***Index Terms*—De-anonymization; Identity Separation; Social Network; Seed Identification.**

I. Introduction

A social network (SN) is a web of connections between certain individuals (or organizations) in the society, who are tied together by one or more specific relations or attributes, e.g., the trails of e-mails or phone calls altogether. The group of social networks includes social networking services (SNS), which allow individuals and other entities such as organizations to form links. These services have an underlying social network graph, where vertices represent individuals or registered users, and edges indicate relationships, connections or other kinds of links. Content on most SNSes is based on user-generated information, e.g., status messages, family photos and videos, and hyperlinks to other websites. Such services may also include the functionality of creating and joining groups of interest, and rating other users' content – a simple concept for which being Facebook's likes.

A social network is a rich information base for many branches of science. Sociologists, for instance, may find out valuable information about the structure of the society, group behavior, etc., by analyzing the *anonymized export* of the database, which can be considered a graph with labels on the vertices and edges, where vertices represent members of the network, and edges connection between them. Such an export may either be obtained through a request to the operator of the SNS, or by manually collecting the information through the use of a so-called web crawler.

Anonymization is not a trivial task; merely stripping the names from the database has proven to be insufficient [19]. An attacker may embark on restoring the deleted identifiers for various reasons, e.g., for obtaining previously unknown or unconfirmed information about a user for improving the efficiency of illegal or otherwise malicious practices, such as phishing. Research in the field has made such attacks against social networks readily available, and has proven that they do not have a prohibitively high complexity [20].

As a means of thwarting de-anonymization attempts, we propose to use identity separation in social networks [12], a concept for selectively concealing and revealing certain pieces of information in specific contexts, which is called the technique of Partial Identities [6] or Role-Based Privacy [12]. For instance, one may want to separate her colleagues from her friends (in the stricter sense of the word) on a SNS, by effectively managing separate contact lists for separate identities [11]. This concept, since it harmonizes with our real-life information sharing habits [2], is now available as built-in function, in a novel SNS, called Google+[2].

In this paper, we analyze the effectiveness of identity separation against de-anonymization attacks in case of a cooperative service provider (i.e., who leaves identities separated in the export; the analysis of the uncooperative service provider is considered as future work). These attacks can be categorized as active, semi-passive and passive methods. Active attacks allow creating new nodes in the social network, and adding edges with the rest before obtaining the anonymized export, while semi-passive attacks only allow creating additional edges without adding vertices [4]. Passive attacks rely on the unmodified content in the database, but use auxiliary data sources to de-anonymize users [20], [21].

Our work focuses on the state-of-the-art passive attack in [20]. Although the work described in [21] is more recent, but that attack is not proven to be generic: while the attack in [20] is executed on two totally different networks, namely Flickr and Twitter[3], the attack in [21] is executed on two snapshots of the same (Flickr) network. The latter attack has two significant disadvantages, since it tries to match the top l nodes in the two networks: first, due to its characteristics, it

G. Gy. Gulyás is an adjunct assistant professor at the Department of Telecommunications, Budapest University of Technology and Economics. Magyar tudósok körútja 2., H-1117, Budapest, Hungary. (gulyasg@hit.bme.hu).

S. Imre is a professor, and the head of department at the Department of Telecommunications, Budapest University of Technology and Economics. Magyar tudósok körútja 2., H-1117, Budapest, Hungary. (imre@hit.bme.hu).

[2] http://plus.google.com

[3] http://flickr.com and http://twitter.com

could not work for another set of nodes with lower degrees, and second, matching the top l nodes in different networks may involve some difficulties (see Section II.A).

Therefore, as our main contribution in this paper, we have analyzed the success rate of the state-of-the-art attack from a user's point of view by calculating failure probability for a single vertex of the anonymized export. To assure the generality of our analysis – and since user behavior is yet unknown – we have defined a generic attack-independent statistical model for user behavior regarding identity separation and edge anonymization.

The paper is structured as follows. In Section II, we review the related literature on the main areas involved by our research: re-identification in anonymized social network graphs, identity separation and anonymity in social networks. In Section III, a novel user behavior model is introduced for identity separation, which incorporates multiple behavior types dependent on the possibilities of the user. The effects of identity separation on active attacks are discussed in Section IV, and in Section V, the analysis of passive de-anonymization attacks is presented. Finally, in Section VI, we conclude our work.

II. Literature Survey

In this section, we briefly discuss the literature most related to our work. First by including the most relevant de-anonymization attacks, and also discuss the literature of identity separation for social networks. Finally, we present a method for applying anonymity in social networks, and for exporting such data.

A. De-anonymization Attacks for Social Networks

The first de-anonymization algorithms were active (and semi-passive) attacks [4]. As discussed in the introductory section, attacks of this type intend to insert a hidden but unique pattern into the graph by systematically adding vertices to the social network graph, and trying to create edges between some targeted users, before obtaining the anonymized export. The adversary may then attempt to recognize this hidden structure, and infer information about the selected users under attack (i.e., which are connected to the structure), including their contacts and (if provided along with the graph) their profiles.

However, since active attacks intrinsically assume that the adversary is able to modify the network before creating the export, they inherently have some weak spots. For instance, in case the modification is possible, the operator of a major social networking website is likely to attempt to find the fake user accounts, and delete them. Since real users of the social network do not have a meaningful motive to link back to the malicious nodes (i.e., confirm friend requests), the service provider will find that edges linked to these vertices are mostly going outwards from, and seldom coming inwards to them. The computational complexity of active attacks is likely to be small [4].

Passive attacks, while computationally more expensive, do not require the modification of the social network graph, and therefore the service provider cannot proceed with reactive measures, only with proactive ones. This makes the attack more difficult to counter, and also more versatile in terms of area of application (i.e., the same concept can be applied to multiple kinds of social networks). Furthermore, these attacks are capable of extending to the entire network, or at least a significantly large part thereof.

The state-of-the-art passive attack described in [20] maps the corresponding nodes of two graphs (i.e., accounts of the same person) solely based on structural information. The adversary defines an error parameter ε to control the acceptance of a mapping, i.e., if the algorithm should be more lenient and possibly accept erroneous mappings, or be stricter and be prone to rejecting correct ones.

The algorithm executes in two phases: seed identification and propagation. In the first phase, the attacker tries to find in the anonymized graph the counterpart of a unique k-clique present in the source graph (the algorithm in [20] uses 4-cliques, to be exact). First, for a unique k-clique in the source graph, the attacker computes the degree of each vertex and the number of common neighbors for each pair of nodes, then looks for similar k-cliques with similar values (within a factor of $1 \pm \varepsilon$) in the target graph. The error factor is considered for mapping each vertex (in the case of degrees) and each pair of vertices (in the case of common neighbor counts). Structural modifications within the cliques are disallowed; identification fails if one or more edges are erased from the clique.

In the second phase, the algorithm iteratively adds nodes to the mapping until there are unmapped vertices that have reasonably good mappings. If the attacker fails in the first phase, the second one is never run; therefore, we focused on analyzing the success rate of the attacker, but plan to analyze the effects of identity separation on this phase as future work. However, we expect that if the seed identification is not successful in general, then the second phase should fail, too.

Narayanan et al. in [21] introduce a similar attack with a less rigid, non-pattern-based seed identification phase (the propagation phase is essentially the same). Instead of looking for several seeds, this attack tries to find matches of node pairs in the top l nodes of the two networks, and then starts the propagation phase from there. Matches are based on node degrees and common neighbor counts by applying cosine similarity for the pairs.

However, this attacker algorithm does not seem to be generic. First, since degrees of nodes in the top l set differ the most in the whole network, this technique can not be applied to other set of l nodes with lower degrees: there would be too many similar nodes in the compared sets (e.g., see Fig. 1-2 in [21]). Second, for social networks with a similar purpose (e.g., Facebook[4] and Google+), it may be right to assume that the top l nodes overlap, but in general, that should not be true. For instance, it is not very likely that accounts belonging to the same owners are the most popular on Flickr and on LiveJournal.

Therefore, to the best of our knowledge, the [20] passive

[4] http://facebook.com

attack is still the state-of-the-art attack that can be found in the literature. Comparison of attack types are summarized in Table 1.

TABLE 1. ATTACK TYPES AND ATTACKER CAPABILITIES.

Data sources	Passive	Semi-passive	Active
External data	Use public data as auxiliary source		
Internal data	-	Modify profiles, connections	
	-	-	Create new registrations

B. Identity Separation in Social Networks: a Desired and Privacy-Enhancing Feature

We do not classify our social contacts on SNSes by default, as there is only one category: "friends". However, this is normally not the way we, humans, classify our acquaintances [2]. We keep track of multiple groups of people we know from different "stages" of our life, e.g., school, workplace, and family, and interact with them in a disjoint fashion in terms of place and time [8]. If our offline disclosure of information works in a different way than an SNS, we will act in a different way online: we will likely self-censor ourselves, and, at least sometimes, disclose some content to unwanted audiences.

This is a clear indication for the need of identity separation within social networks [11]; SNSes allowing users to share diverse information with different user groups (e.g., sharing different availability status with colleagues and friends) or to commit identity separation in some contexts (e.g., making political and private identities totally unlinkable). Such methods exist in the literature as the technique of Partial Identities [3], [5], [6], [10], [9], and Role-Based Privacy [24], [14], [15], [16], [18], [12]. Both allow users to publish diverse attributes under different pseudonyms.

For the case where the consent of the SNS provider cannot be assumed, one can use cryptography to enforce identity separation on an existing SNS [22], [1] or implement the social network on a distributed, cryptographically secured architecture [7]. However, our current work focuses on the possibilities on the model issues, and not the cryptographic side of the problem.

Google+ is the first to implement identity separation as a tool for privacy protection; the goal of this feature (namely Google Circles) is to allow proper audience selection for sharing content. Currently, this is not yet a complete solution, as it only works for content sharing, but there is only a single profile (flat structure). One important attribute of the Circles feature that it is mandatory.

Similar, optionally available features exist in many services. For instance, in the Windows Live Messenger[5], one can set invisible mode for each contact group separately, or in Facebook, friends can be sorted into groups, and content sharing can be done accordingly. Compared to the Google Circles, the biggest disadvantage of these features is that they are optional, and therefore probably fewer users know about and use them.

C. Anonymity, Pseudonymous Identifiers and Data Export Sanitization

There are two types of identity separation. The simpler identity separation function is where you can sort your contacts into lists, and, for instance, post content for them separately. This is called internal identity separation [11]. The other means of separating identities – external separation – is either managing multiple profiles on the same SNS, or using multiple networks for different audiences (e.g., Facebook as our means of informal online presence, and LinkedIn[6] as our formal one).

Obviously, the user's pseudonyms on different sites must not disclose that they belong to a single user [5]. It must be noted, too, that the internal separation functionality found on many SNSes is insufficient. For instance, Facebook does not allow hiding group memberships, implicitly exposing attributes of the user that she may have wanted to conceal [25].

Therefore, a social network supporting identity separation allows three levels of anonymity [12]:

- The weakest is pseudonymous identification (i.e., internal separation): the user is identified by a globally unique identifier (a pseudonym), and her identities can be linked through it.
- Unlinkable pseudonymity is a stronger level of anonymity (i.e., external separation), where the user may separate her identities by the means of multiple pseudonymous identifiers. These cannot be linked together, since content corresponding to an identity has its respective pseudonym as the originator.
- The strongest level is total anonymity, which allows the user to post content without any identifier linked to it, and therefore making it very hard to trace the information back to her.

In our work, we assume a service provider that honors the above mentioned separation methods when creating the network export. The reasoning behind this is that if the attacker cannot reverse the separation, she cannot know that multiple nodes in an anonymized graph belong to one person, and therefore they should be mapped together to a vertex in the known graph.

Transformation of a social network where the users can use identity separation and edge anonymization into a traditional social network graph is simple. Linkable nodes (i.e., those that use linkable pseudonymous identifiers) are merged, unlinkable nodes (i.e., those that use separate pseudonymous identifiers) are preserved as separate vertices, and anonymized edges are simply deleted. This way, all the aforementioned levels of anonymity in the privacy-enhancing social network model are reflected in the transformed graph. To sum it up, it can be seen that user behavior can greatly affect the structure of the exported social network graph.

[5] http://explore.live.com/windows-live-messenger

[6] http://linkedin.com

The original articles on passive attacks analyze scenarios where the source (or auxiliary) and target graphs were both regular social networks. In this paper we analyze a mixed scenario, where the source graph is a regular social network, and the target is one which allows identity separation. In this case, modeling user behavior is easy, since only the separation process needs to be approximated with a statistical model.

We leave the analysis of the third type of attacks, namely that uses networks with identity separation for both the source and the target, as future work. In this case pattern-based methods may also work; however, if there is no reference data on user awareness, it needs to be modeled. The reason for this is simple: for instance, a user that has a higher level of awareness may use different identities in the source and target networks to make such de-anonymization attacks more difficult to execute. Therefore, the success of these attacks is not as trivial as for the first type of attacks.

III. Modeling User Behavior for Identity Separation

In this section, we describe a model (a set of sub-models) of the user behavior for identity separation that allows all the three levels of anonymity mentioned in Section II.C. We have mapped identity separation to the graph model as "splitting" a vertex in a graph and probabilistically sorting the edges (represented connections with her contacts) between the new nodes, in some cases allowing duplication of edges with a certain probability. As mentioned before, anonymization of an edge is reflected by deleting it from the graph with a certain probability.

There might be other approaches for modeling user behavior; however, in our opinion this approach is the closest to how people manage their acquaintances in their lives [2]. Furthermore, just to mention a real-life example, our approach is quite similar to functionality in Google+, namely how contacts are managed in the circles feature.

Another important property of our model is that it is attack independent. This allows analyzing multiple attacks with this model, even pattern-based and other, non-pattern-based ones.

A. Modeling Identity Partitioning

Let us define a regular social network graph as $G_{SN} = G(V, E)$. Node $v \in V$ is a user who has $n = \deg(v)$ neighbors in G. While performing identity separation, node v introduces y new vertices (i.e., new identities), and sorts edges with probabilities $p_1, p_2, \dots, p_y$ to each of the new identities.

We can categorize the model parameters as:

- Context-dependent parameters: the user has little influence over such parameters. The only context-dependent item in our model is the neighborhood size of the user (n).
- User-dependent parameters: these are the statistical descriptors of user behavior. In our model, the number of new identities denoted as y and the probabilities of sorting edges ($p_1 \dots p_y$) can be considered as user-dependent parameters.
- Attacker-dependent parameters: the adversary is free to choose these before executing the de-anonymization attack. Currently, there are no such parameters (as the model is attack independent), but for instance, in Section V. such new parameters will be introduced.

The number of new identities (y) is modeled with a random variable Y. The distribution of the edge sorting is $P(X_1 = x_1, \dots, X_y = x_y)$, where X_i is a random variable describing the number of edges between the ith new identity and the neighbors of the original node. We do not assume any distribution for Y, and the distribution for X_i is defined with the chosen user-behavior model.

The model and the parameters could be fine-tuned with quality reference data; however, there are some obstacles in the way. As mentioned before, Google+ is the first service that compels its users to sort their contacts, and although the profiles are public, circles of users are not yet. Furthermore, in other services where sorting contacts is available but not mandatory, we experienced that this feature is rarely used, and often the contact groups are not publicly available – making social network data harvesting a futile task. Therefore, verifying our model with reference data is still an open research task that stays future work.

B. General Assertions

Our model is based on some assertions about the structure of the network before and after applying identity separation. These assertions are assumed to be true in all sub-models.

Assertion 1. *A new identity can have even zero of the original contacts in the export (i.e., due to edge anonymization).*

Assertion 2. *A user v_i may create a maximum of* $\deg(v_i)$ *new identities.* While it is possible to create an unlimited number of identities, and assign duplicate edges to them, we believe that this does not match with the user's expected behavior and this is an acceptable rational limitation.

Assertion 3. *A user may create even 0 new identities (i.e., perform self-deletion from the graph).* This happens when all the connections are anonymized.

Assertion 4. *The only contacts existing in the source network are modeled in the identity partitioning.* This simplifies the behavioral model, but does not necessarily make the results more favorable: including new contacts would add noise to the model, which would increase failure rates.

Assertion 5. *All actions of the nodes in the network are assumed to be using identity separation independently.* Our analysis does not cover collaborating users, even though collaboration would mean stronger resistance and higher failure rates.

Assertion 6. *Edges are not sorted independently.* This is a rational consideration, since all new identities belong to the same user, who sorts the edges (in an intelligent way).

C. Sub-Models for User Behavior

Dependent on the chosen user behavior, there are further aspects to be considered in the sub-models:

- Can different identities of the same user have overlapping neighborhood (i.e., duplicated edges)? Overlapping allows the overall number of connections to increase, formally, $\exists P(X_1 = x_1, \dots, X_y = x_y) > 0$, that $\sum x_i > n$ with $(0 \le x_i \le n)$.
- Is edge anonymization permitted? Deleting edges allows the overall number of connections to decrease, as $\exists P(X_1 = x_1, \dots, X_y = x_y) > 0$, that $\sum x_i < n$.

Based on these aspects, new sub-models can be introduced that we have summarized in Table 2. The names of the sub-models require some explanation. We have named the model with no edge anonymization, and no overlaps the basic model, since this allows the least privacy enhancing functionality for the user (only identity separation itself). Conversely, the realistic model is just the opposite: it implies the fewest limitations in her possibilities. We believe that most users of a social network would use anonymization or duplication for their connections; hence the notation "realistic".

TABLE 2. CATEGORIES OF MODELS OF USER BEHAVIOR.

	Overlap	**No overlap**
Edge deletion	Realistic model	Best model
No edge deletion	Worst model	Basic model

Besides, a worst and a best model also exist, which are named from the algorithm's point of view. The best model allows a user to only decrease the number of her contacts, and therefore causing more information loss (i.e., structural damage). The worst model is the opposite: it only allows creating duplications, and therefore making "backups" of structural information, and helping the attacker that way.

IV. IDENTITY SEPARATION AND ACTIVE ATTACKS

Backstrom et al. describe two attacks, a semi-passive and an active attack, in which both the attackers are able to modify the network prior to the sanitization [4]. In both attacks the attackers' goal is to insert a specific structure (a subgraph) into the SN graph that can be revealed later only by the attackers but no one else – this is what they call structural steganography. This subgraph is connected to the SN graph by creating new edges to a small number of targeted users. This is one the disadvantages of this attack: they only allow revealing the identity of a small number of users. However, for some networks active and semi-passive attacks can not be executed for one of the following reasons:

- The modification of the network structure may be expensive (e.g., phone calls).
- The modification may not be executable (e.g., network created from observed e-mails).
- To insert the structure too many modifications would be required (e.g., a valid e-mail address must be providing for the registration).
- The attacker is not always able to influence connections (e.g., connections require two-way confirmation).

All these problems inspired the research of passive attacks [20]. However, from the viewpoint of identity separation, the active attacks are better than passive attacks: the inserted structure is under the exclusive control of the attacker, and therefore its structure is always known, and can be found by the malicious collaborators. On the other hand, even if such attacks may not be prevented, one can use identity separation to separate herself from suspicious users, neighborhoods (i.e., structures) to prevent re-identification. This kind of self-defense works against passive attacks, too.

V. ANALYSIS OF FAILURE PROBABILITY FOR THE CLIQUE-BASED PASSIVE ATTACK

In this section, we discuss our results of the analyses based on the user behavior model in case of an attacker using the clique-based algorithm (discussed in Section II.A). We included an assumption from the original attack: there are some unique 4-cliques that exist in both networks and have similar neighborhoods in both, i.e., the cliques contain vertices with similar degrees [20].

Seed identification is considered to be successful for a clique if it remains a clique, and retains its degree values within an error factor after applying the identity separation. While the original algorithm compares common neighbor counts as well, our analysis concludes that even these two criteria can be violated effectively with identity separation, as shown later (i.e., we analyze the lower bound for the failure probability).

Here, we analyze the failure probability of the attacker on statistical basis; however, it should be noted that individual protection against attacks is still possible, even if statistically an attack seems to be feasible, i.e., a user can intentionally create different neighborhood structures in different networks.

In our opinion, the basic and the realistic models are the closest to real user behavior: we expect users to have roughly the same number of contacts before and after the identity separation (not including new contacts). Therefore, in our research, we focused on these models: the basic model is an analytically simpler model allowing identity separation only, and the realistic model allowing more functionality for users (with more mathematical complexity).

A. Naïve Analysis on 4-cliques

By using real-life data harvested from different social networks, we simulated identity separation to analyze its effects on the network structure from the attacker's clique-oriented point of view. Cliques can be easily destroyed via identity separation:

- One of the users separates herself totally from the clique. This is equivalent to the removal of the representing node.
- One of the users removes at least one edge from the clique.
- At least one of the users uses identity separation and separates at least an edge from the clique.

In these cases, the clique no longer remains connected, and the attacker will fail in finding it. We executed simulation experiments to determine how effectively identity separation removes 4-cliques from the network. For our experiments we used structural information crawled from two real-life networks: the Slashdot friend or foe links that were crawled in February 2009 [17], and the Epinions who-trust-who network data that were crawled in 2003 [23]. From the Slashdot network 10,000 nodes were selected containing 1,816,110 4-cliques, and 1,000 nodes were selected containing 2,102,842 4-cliques from the Epinions network. For comparison, a full graph of 100 nodes (with 3,921,225 4-cliques) was also included.

For simulating identity separation, random nodes were selected to be split into two new nodes, and edges were assigned to each with equal probability (i.e., accordingly to the basic model). We have also defined a theoretical limit to show the expected number of cliques affected by identity separation. Adding more privacy-enhancing functionality to the simulation, such as edge and node removal, the number of cliques would be furthermore decreased, closer to the theoretical limits.

If random variable Y denotes the number of identities belonging to the same user (without assigning a distribution to it), the probability if there are any identity separations in a k-clique $C_k = \{v_1, \dots, v_k\}$ can be calculated as in

$$p_{ids} = P(\exists v_i : Y_i > 1 \mid i = 1..k) = 1 - P(\forall v_i : Y_i = 1 \mid i = 1..k) = 1 - P^k(Y = 1). \quad (1.)$$

Therefore, the expected number of cliques remaining intact can be calculated as the expected value of the binomial distribution $Z \sim B(N_{kcls}, 1 - p_{ids})$, where N_{kcls} denotes the number of 4-cliques in the original graph. The expected value of Z is

$$E[Z] = N_{kcls} \cdot (1 - p_{ids}) = N_{kcls} \cdot P^k(Y = 1). \quad (2.)$$

The relative values of $E[Z]$ with $k = 4$ are denoted on Fig. 1 as the expected number of cliques remaining intact by identity separation (denoted as "Theoretical"). It is possible that the clique remains a clique, but the probability of recovery depends on further errors regarding the compared degree and common neighbor count values. An analysis of these issues is discussed in the next sections.

We found that as the number of users who use identity separation increases, the number of 4-cliques decreases fast and almost similarly for all networks (see Fig. 1). For instance, in both test networks for $P(Y = 2) = 0.2$ the number of remaining cliques was almost halved: the percentage of intact 4-cliques was 52.26% for the Slashdot network, 51.27% for the Epinions network, and 55.22% for the full graph. It is also visible on Fig 1. that graphs having more 4-cliques degrade faster. The reason behind this phenomenon is simple: usually several 4-cliques overlap in a single node, and therefore splitting it causes the deletion of multiple 4-cliques.

Our conclusion is for the naïve analysis: identity separation erodes network structure effectively, thus it offers strong protection against structural attacks, and therefore it needs to be furthermore analyzed.

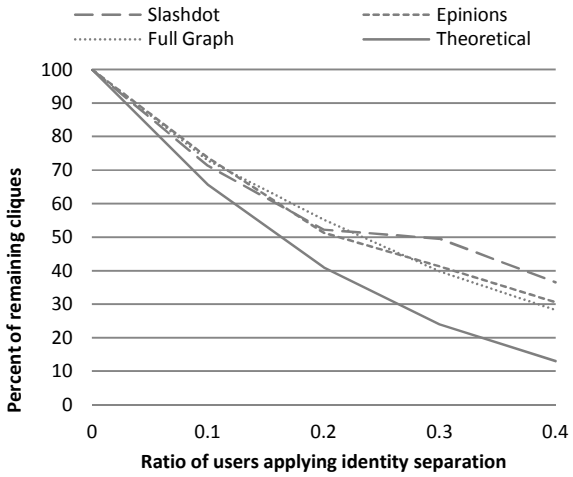


Fig. 1. Simulation results (including the theoretical limit) show the degradation in clique numbers in case of allowing identity separation.

B. User Behavior Model with Attack Related Parameters

In this case, node $v \in V$ is a user who is part of a k-clique, and has $n = \deg(v)$ neighbors in G, and therefore node v has $k - 1$ inner and $n - k + 1$ outer edges, as seen from the viewpoint of the clique.

For the inner edges, the distribution of the edge sorting is described as $P(X_1 = x_1, \dots, X_y = x_y)$, with no predefined distribution included; the distributions are defined with the chosen model. For the outer edges, the distribution is described similarly as $P(X'_1 = x'_1, \dots, X'_y = x'_y)$. X_i and X'_i are random variables describing the number of edges between the ith identity and the members of the original clique, and those between the ith identity and the neighbors of the original node, respectively.

The original algorithm defines an error parameter ε for the seed identification, and an error measure based on it: the matched node degree values need to match within an error factor of $1 \pm \varepsilon$.

Based on this, we define an error measure function that will be used in the calculation of the failure probability, given by the function of

$$g(x, y) = \begin{cases} 1, \left(\frac{x+y}{n} < 1 - \varepsilon \wedge y = k - 1\right) \vee y < k - 1 \\ 0, \text{otherwise} \end{cases}, \quad (3.)$$

where x denotes the number of outer, and y denotes the number of inner edges.

The node degree value n, the clique size k, and the error parameter ε are assumed to be known constants. We note, that the clique size (k) and the error parameter (ε) are new attacker-dependent parameters introduced to the model (see Section III.A). It should be noted that the attacker, to achieve better results, can choose to execute several attacks with

different values for these parameters, without any limitations.

C. Calculation of Failure Probability

It must be noted that different actors have different views on the measure of failure probability. The adversary is interested in discovering the correct mapping for several cliques. As such, she is likely to be interested in the probability of failure in identifying a k-clique. Here, we only define the failure probability for a single node, but for a clique it can be calculated simply by giving the probability of the union of failure events, where members of the clique damage the clique or change node degree values and causing errors.

The point of view of a user is, on the other hand, that she herself should not be vulnerable to the attack; other users are more or less irrelevant to her. This is why we have focused on calculating failure probability of single users. It must be noted that this probability is clique-independent, and therefore the same regardless of the number of cliques the user is member of.

Furthermore, the calculation does not take actions of other users in the clique into account, i.e., it is assumed that they neither perform identity separation, nor anonymize any of their edges. If we took these effects into account, the failure probability would be higher in most cases, and at least equal in theory, since other users could also destroy the clique or change the degrees of the vertices thereof, making identification less probable.

The probability of failure for a node v_a, based on the variables, assumptions and assertions introduced previously is

$$P(\text{"fail for } v_a\text{"}) = P(Y=0) + \sum_{y=1}^{\deg(v_a)} P(\text{"fail"}|Y=y)\cdot P(Y=y). \quad (4.)$$

The first member of the sum is the probability of the case where the user has 0 identities in the exported graph, i.e., all her edges are anonymized. The other part of the sum incorporates Assertion 2, namely that the user creates at most as many identities as many contacts she has. The results for the different sub-models of user behavior mainly deviate in the definition of the conditional probability $P(\text{"fail"}|Y=y)$. Note that the formula for the sum may slightly differ in some cases, e.g., in that of the basic model, where it does not include probabilities for $y=1$.

In the general case, the conditional failure probability in (4) can be unfolded as

$$P(\text{"fail"}|Y=y) = \sum_{\forall l_i} P\big(\text{"fail"}|X_1=l_1,\dots,X_y=l_y\big)\cdot P\big(X_1=l_1,\dots,X_y=l_y\big). \quad (5.)$$

Furthermore, probability $P\big(\text{"fail"}|X_1=l_1,\dots,X_y=l_y\big)$ can be calculated differently for two cases. If $\forall l_i < k-1$, i.e., the clique is always destroyed, since all edges are sorted in groups having less than $k-1$ edges, then $P\big(\text{"fail"}|X_1=l_1,\dots,X_y=l_y\big)=1$ always.

In the other case, where $\exists l_i = k-1$, the conditional failure probability in (5) is calculated as

$$P\big(\text{"fail"}|X_1=l_1,\dots,X_y=l_y\big) = P\left(\bigcup_{\forall m_i}\big(X'_1=m_1,\dots,X'_y=m_y\big|g(m_1,l_1)=1,\dots,g\big(m_y,l_y\big)=1\big)\right) \quad (6.)$$

By knowing that these events are mutually exclusive, (6) equals to

$$\sum_{\forall m_i} P\big(X'_1=m_1,\dots,X'_y=m_y\big)\cdot g(m_1,l_1)\cdot\dots\cdot g\big(m_y,l_y\big). \quad (7.)$$

Therefore, the failure probability for a node with y identities is

$$P(\text{"fail"}|Y=y) = \sum_{\nexists l_i=k-1} P\big(X_1=l_1,\dots,X_y=l_y\big) + \sum_{\exists l_i=k-1} P\big(X_1=l_1,\dots,X_y=l_y\big)\cdot\left(\sum_{\forall m_i} P\big(X'_1=m_1,\dots,X'_y=m_y\big)\cdot g(m_1,l_1)\cdot\dots\cdot g\big(m_y,l_y\big)\right) \quad (8.)$$

This is applicable for any $y\neq 0$ number of identities, and by using this formula the overall failure probability can be described accordingly.

D. Failure Probability in the Basic Model

The basic sub-model is the analytically the simplest one, and the results obtained with this restricted model are quite satisfactory. The basic model introduces additional assertions.

Assertion 7. *Contacts of the separated identities do not overlap.*

Assertion 8. *Edges cannot be anonymized.*

In this model, the user sorts n edges among y identities. The multinomial distribution is a natural choice for describing such a case, since it describes n trials when the outcomes can be sorted into one of y groups. Additionally, group probabilities can be adjusted, and therefore this model allows fine-tuning the distribution in a way for describing user behavior in the desired way. Multinomial distribution is used as

$$P\big(X_1=x_1,\dots,X_y=x_y\big)\sim Mu\big(k-1,p_1,\dots,p_y\big), \text{ and}$$
$$P\big(X'_1=x'_1,\dots,X'_y=x'_y\big)\sim Mu\big(n-k+1,p_1,\dots,p_y\big),$$

where $\sum p_i = 1$.

The formula for failure probability can then be derived as:

$$P(\text{"fail"}|Y=y) \tag{9.}$$
$$= \sum_{\substack{\sum l_j=k-1 \\ \exists l_j=k-1}} \Bigg(P\big(X_1=l_1, X_2=l_2, \dots, X_j=l_j, \dots, X_y=l_y\big) \cdot \Bigg(\sum_{m_1=0}^{n-k+1} \cdots \sum_{m_{y-1}=0}^{n-k+1-\left(\sum_{h=1}^{h=y-2} m_h\right)} P\Big(X_1'=m_1, X_2'=m_2, \dots, X_j' = m_j, \dots, X_y' = n-k+1-\Big(\sum_{h=1}^{h=y-1} m_h\Big)\Big) \cdot g(m_j, l_j)\Bigg)\Bigg)$$
$$+ \sum_{\substack{\sum l_j=k-1 \\ \nexists l_j=k-1}} P\big(X_1=l_1, X_2=l_2, \dots, X_y=l_y\big).$$

Then, the overall failure probability can then be derived easily. We know that for $P(\text{"fail"}|Y=1)$ the neighborhood of the node would remain the same, and therefore would not introduce any error in the seed identification. Consequently, the result can be deduced from the general formula, with the exclusion of the case for $y = 1$.

During our numerical analysis, we have found that in this model for a fixed p_1, the failure probability for two identities is the lower bound for all failure probabilities with a higher number of identities that include p_1. In other words, for $\forall p_1, \dots, p_k$ with a fixed p_1:

$$P(\text{"fail"}|Y>2) \geq P(\text{"fail"}|Y=2). \tag{10.}$$

This is an important finding for two reasons. On the one hand, this is a lower bound for failure probability, and therefore it is enough to continue analysis in the case of $P(\text{"fail"}|Y=2)$. On the other hand, it facilitates the estimation of the overall failure probability as well:

$$P(\text{"fail"}) \geq \sum_{m=2}^{\deg(v_i)} P(\text{"fail"}|Y=m) \cdot P(Y=m) \tag{11.}$$
$$\geq \sum_{m=2}^{\deg(v_i)} P(\text{"fail"}|Y=2) \cdot P(Y=m)$$
$$= P(\text{"fail"}|Y=2) \cdot \big(1 - P(Y=0) - P(Y=1)\big).$$

Fig. 2 describes how failure probability changes with different values for parameter n, while parameters $k = 4$ and $\varepsilon = 0.05$ are fixed.

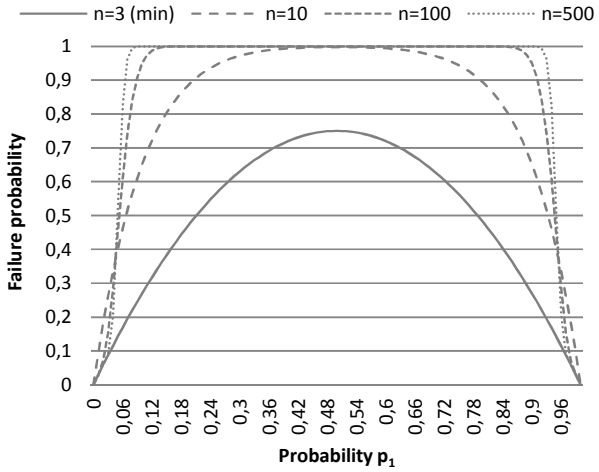


Fig. 2. Parameter analysis of n: $P(\text{"fail"}|Y=2)$ as a function of p_1, with fixed $c = 4$ and $\varepsilon = 0.05$ with different values for n.

The analysis has several interesting consequences. First of all, it can be seen that the failure probability is conveniently high even for a small n (e.g., $n \geq 10$). Secondly, users are given a relatively wide range of options for making their identification fail. Even if they use identity separation for just two identities, and the probability of using the second identity is small, the failure probability still remains high (e.g., for $p_1 = 0.1, n = 50$: $P(\text{"fail"}|Y=2) = 0.899$). It can be seen that the curve has inflection points. These are functions of the error parameter ε.

Fig. 3 describes how the failure probability changes in the function of ε while parameters $n = 100$ and $k = 4$ are fixed. The curves do not deviate significantly for other n values either.

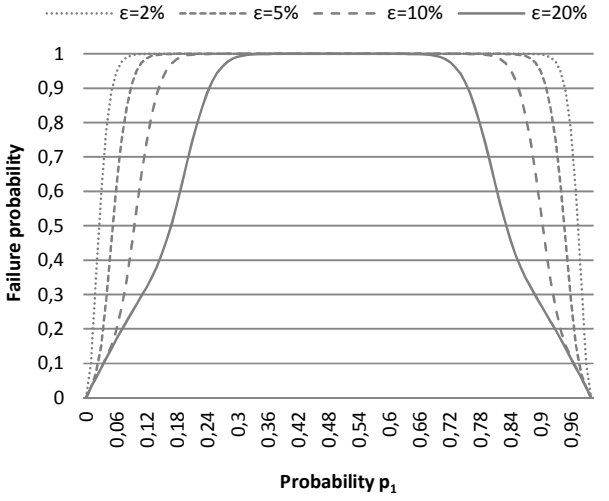


Fig. 3. Parameter analysis of ε: $P(\text{"fail"}|Y=2)$ as a function of p_1, with fixed $c = 4$ and $n = 100$ with different values for ε.

This shape of the curve practically concludes that a user making use of identity separation in a meaningful way, the adversary cannot influence the success of the attack. It is demonstrated in the original article that the value of ε should

be around 0.05, and that a practical limitation of $0 < \varepsilon \leq 0.1$ applies. For these values, users should choose p_1 and p_2 such that $0.1 \leq p_1, p_2 \leq 1$ $(p_1 + p_2 = 1)$, because this marks a point (i.e., a failure probability) beyond the inflection point of the curve. Finally, the analysis of parameter k has shown that there is no deviation in the failure probability for different clique sizes with different neighborhood sizes (n with $\varepsilon = 0.05$).

To sum it up, we can conclude that if the users use identity separation wisely, considering the influencing power of different parameters as mentioned above, the attacker has a low probability of identifying the nodes. This means that users need to separate their contacts into larger, but not necessarily equally sized groups. Therefore, this user behavior model can be suggested for users as a practical way to use identity separation, since it offers powerful protection if applied widely throughout the network.

E. Analysis of the Realistic Model

In this section, we discuss the analysis of the realistic model, which deviates from the basic model in regard of the additional Assertions 7 and 8.

Assertion 9. *Contacts of separated identities can overlap.*

Assertion 10. *Edges may be anonymized by users.*

Selecting the proper distribution is not an easy choice, therefore it should be defined by its probability matrix, denoting the probability of a possible outcome in a cell. Deciding which distribution to choose in such a model is an interesting question. In our opinion, the distribution should reflect that the most likely case is that the number of all contacts after the identity separation is similar to that before, i.e., a few deletions and duplications are likely, but major deviations are not (see Fig. 4).

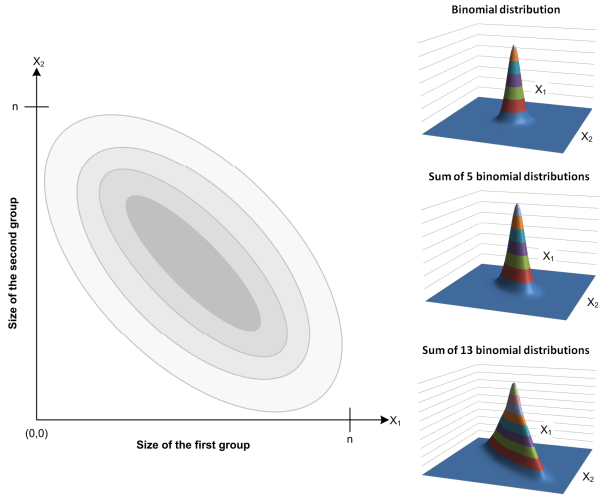


Fig. 4. Concept for the distribution of realistic models with $y = 2$, and some examples. On the left part of the figure, the darker areas have higher probabilities (these values are outstanding on the right part).

Accordingly to the given distributions and the generic formulae for failure probability, we have done the parameter analysis numerically. Its characteristics are similar to that of the basic model, and the preliminary results are satisfactory for this model, too (see Fig. 5).

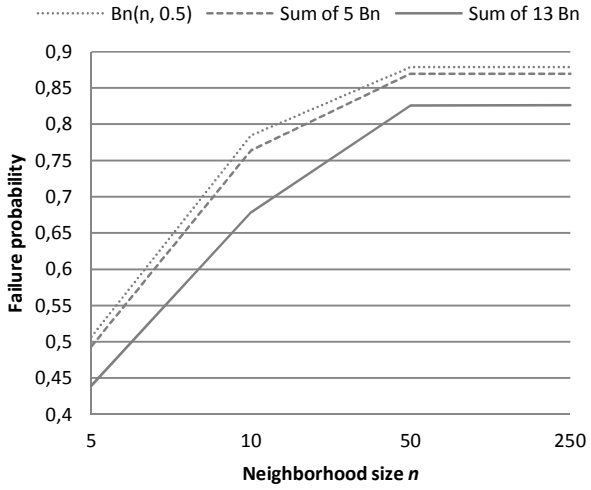


Fig. 5. Failure probabilities for different distributions with $y = 2$, for different sizes of n.

We can conclude that the results are satisfactory even for small ns in all distributions under examination; however, these models deserve further research dependent on reference data, which we assign as future work.

VI. Conclusion and Future Work

Our analysis has shown that our proposed models make seed identification fail with high probability. Therefore, we can consider identity separation as an effective countermeasure against de-anonymization attacks if the user chooses the parameters wisely.

However, besides the answered questions, new ones arise. In the future, we would like to extend our analysis to the best and worst models, and discuss further results with the realistic model including new distributions compared with reference data if possible.

As it is mentioned in this paper, the analysis focused on the seed identification phase in the state-of-the-art passive attack, but the propagation phase should be analyzed in the future, since it is incorporated in two passive attacks [20] and [21].

It also seems to be desirable to extend the user behavior and the attacker model with new parameters to make it open for new attacks yet unknown. For example, the model can be extended to allow the analysis of the attack in [21].

Additionally, there are other types of third party attacks in the literature, such as attribute based ones [13], for which the effects of privacy-enhancing identity management should be analyzed. Instead of standalone use for re-identification, attributes can also be used to strengthen structural attacks: de-anonymization results can be easily verified and corrected by inspecting the available attributes of nodes. Perhaps identity separation has also a viable effect on these attacks – it would be interesting to see this in the future.

Acknowledgment

We would like to thank András Telcs for his valuable

comments and suggestions on our work, and Ádám Máté Földes for inspiring discussions on the topic, and his suggestions during writing this paper.

Our work presented in this paper was supported from the KMOP-1.1.2-08/1-2008-0001 project by the BME-Infokom Innovátor Nonprofit Kft.

REFERENCES

[1] J. Anderson, C. Diaz, J. Bonneau, and F. Stajano, "Privacy-enabling social networking over untrusted networks," in *Proc. of the 2nd ACM workshop on Online social networks*, Barcelona, Spain, 2009, pp. 1-6.

[2] P. Adams, "The Real Life Social Network", presented at the Voices that Matter Web Design Conference, June 28-29, 2010, San Francisco, USA. Available at: http://www.slideshare.net/padday/the-real-life-social-network-v2

[3] K. Borcea, H. Donker, E. Franz, K. Liesebach, A. Pfitzmann, and H. Wahrig, "Intra-Application Partitioning of Personal Data," in *Proc. of Workshop on Privacy-Enhanced Personalization*, Edinburgh, UK, 2005.

[4] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography," in *Proc. of the 16th international conference on World Wide Web*, Banff, Alberta, Canada, 2007, pp. 181-190.

[5] K. Borcea-Pfitzmann, E. Franz, and A. Pfitzmann, "Usable presentation of secure pseudonyms," in *Proc. of the Workshop on Digital identity management*, Fairfax, USA, 2005, pp. 70-76.

[6] S. Clauß, D. Kesgodan, and T. Kölsch, "Privacy enhancing identity management: protection against re-identification and profiling," in *Proc. of the Workshop on Digital identity management*, Fairfax, USA, 2005, pp. 84-93.

[7] L. A. Cutillo, R. Molva, and T. Strufe, "Safebook: A Privacy-Preserving Online Social Network Leveraging on Real-Life Trust," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 94-101, Dec. 2009.

[8] J. M. DiMicco, and D. R. Millen, "Identity management: multiple presentations of self in facebook," in *Proc. of the ACM 2007 International Conference on Supporting Group Work*, Sanibel Island, Florida, USA, 2007, pp. 383-386.

[9] E. Franz, and K. Liesebach, "Supporting Local Aliases as Usable Presentation of Secure Pseudonyms," in *Proc. of the 6th International Conference on Trust, Privacy and Security in Digital Business TrustBus*, Linz, Austria, 2009, pp. 22-31.

[10] E. Franz, C. Groba, T. Springer, and M. Bergmann, "A Comprehensive Approach for Context-dependent Privacy Management," in *Proc. of the 2008 Third International Conference on Availability, Reliability and Security*, Barcelone, Spain, 2008, pp. 903-910.

[11] S. Gürses, R. Rizk, and O. Günther, "Privacy Design in Online Social Networks: Learning from Privacy Breaches and Community Feedback," in *Proc. of 29th International Conference on Information Systems*, Paris, France, 2008.

[12] G. Gy. Gulyás, R. Schulcz, and S. Imre, "Modeling Role-Based Privacy in Social Networking Services," in *Proc. of Third International Conference on Emerging Security Information, Systems and Technologies*, Athens, Greece, 2009, pp. 173-178.

[13] D. Irani, S. Webb, K. Li, and C. Pu, "Large Online Social Footprints – An Emerging Threat," in *Proc. of the 2009 International Conference on Computational Science and Engineering*, Washington, USA, 2009, Volume (3), pp. 271-276.

[14] U. Jendricke and D. Gerd tom Markotten, "Usability meets security - The Identity-Manager as your Personal Security Assistant for the Internet," in *Proc. of the 16th Annual Computer Security Applications Conference*, New Orleans, USA, 2000, pp. 334-344.

[15] J. Hakkila, and I. Kansala, "Role based privacy applied to context-aware mobile applications," in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Hague, Netherlands, 2004, Volume (6), pp. 5467-5472.

[16] M. Hansen, A. Schwartz, and A. Cooper, "Privacy and Identity Management", *IEEE Security and Privacy*, vol. 6, no. 2, pp. 38-45, Mar/Apr. 2008.

[17] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters,". Tech. Rep., Preprint: arXiv:0810.1355, 2008.

[18] R. Leenes, J. Schallaböck, and M. Hansen, "PRIME white paper (V3)", May 2008. Available: https://www.prime-project.eu/prime_products/whitepaper/

[19] A. Narayanan, and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets," in *Proc. of the 29th IEEE Symposium on Security and Privacy,* Oakland, California, USA, 2008, pp. 111-125.

[20] A. Narayanan, and V. Shmatikov, "De-anonymizing social networks," in *Proc. of the 30th IEEE Symposium on Security and Privacy,* Oakland, California, USA, 2009, pp. 173-187.

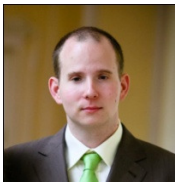
[21] A. Narayanan, E. Shi, and B. I. P. Rubinstein, "Link Prediction by De-anonymization: How We Won the Kaggle Social Network Challenge," in *Proc. of the 2011 International Joint Conference on Neural Networks*, San Jose, California, USA, 2011.

[22] T. Paulik, Á. M. Földes, and G. Gy. Gulyás, "BlogCrypt: Private Content Publishing on the Web," in *Proc. of the Fourth International Conference on Emerging Security Information, Systems and Technologies*, Venice, Italy, 2010, pp. 123-128.

[23] M. Richardson, R. Agrawal, and P. Domingos, "Trust Management for the Semantic Web," in *Proc. of the 2nd International Semantic Web Conference,* Sanibel Island, Florida, USA, 2003, pp. 351-368.

[24] M. Reichenbach, H. Damker, H. Federrath, and K. Rannenberg, "Individual Management of Personal Reachability in Mobile Communication," in *Proc. of the 13th International Information Security Conference*, Copenhagen, Denmark, May 1997, pp. 164-174.

[25] E. Zheleva, and L. Getoor, "To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles," in *Proc. of the 18th international conference on World Wide Web*, Madrid, Spain, 2009, pp. 531-540.



Gábor György Gulyás received his M.Sc. degree in Computer Engineering in 2007, and now currently an adjunct assistant professor at the University of Technology and Economics (BME), Department of Telecommunications. He is a student member of the IEEE, and the Mobile Communications and Computing Laboratory (MC2L). He is one of the founders, and the moderator of the only Hungarian portal on Privacy Enhancing Technologies, the PET Portal & Blog. The focus of his research interests is on applying privacy-enhancing identity management to social networks, but he is also interested in the following topics: privacy and security issues of social networks, web privacy, data protection issues, and using steganography for enhancing privacy.

Sándor Imre was born in Budapest in 1969. He received the M.Sc. degree in Electrical Engineering from the Budapest University of Technology (BUTE) in 1993. Next he started his Ph. D. studies at BUTE and obtained dr. univ. degree in 1996, Ph.D. degree in 1999 and DSc degree in 2007. Currently he is carrying his activities as Professor and Head of Dept. of Telecommunications. He is Chair of Telecommunication Scientific Committee of the Hungarian Academy of Sciences. He participates the Editorial Board of two journals: Infocommunications Journal and Hungarian Telecommunications. He was invited to join the Mobile Innovation Centre as R&D director in 2005. His research interest includes mobile and wireless systems. Especially he has contributions on different wireless access technologies, mobility protocols, security and privacy and reconfigurable systems.