UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA
Departamento de Ingeniería del Software e Inteligencia Artificial

TESIS DOCTORAL

Nuevos ataques estadísticos de revelación de identidades en redes de comunicaciones anónimas

New statistical disclosure attacks on anonymous communications networks

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Alejandra Guadalupe Silva Trujillo

Directores

Luis Javier García Villalba
Javier Portela García-Miguel

Madrid, 2016

Nuevos Ataques Estadísticos de Revelación de Identidades en Redes de Comunicaciones Anónimas

– – – – – – – – – – – – – – – – – – –

New Statistical Disclosure Attacks on Anonymous Communications Networks

Thesis by

Alejandra Guadalupe Silva Trujillo

In Partial Fulfillment of the Requirements for the Degree of
Doctor por la Universidad Complutense de Madrid en el
Programa de Doctorado en Ingeniería Informática

Advisors

Luis Javier García Villalba
Javier Portela García-Miguel

Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense de Madrid

Madrid, November 2015

New Statistical Disclosure Attacks on Anonymous Communications Networks

Thesis by

Alejandra Guadalupe Silva Trujillo

In Partial Fulfillment of the Requirements for the Degree of
Doctor por la Universidad Complutense de Madrid en el
Programa de Doctorado en Ingeniería Informática

Advisors

Luis Javier García Villalba
Javier Portela García-Miguel

Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense de Madrid

Madrid, November 2015

Nuevos Ataques Estadísticos de Revelación de Identidades en Redes de Comunicaciones Anónimas

TESIS DOCTORAL

Memoria presentada para obtener el título de
Doctor por la Universidad Complutense de Madrid
en el Programa de Doctorado en Ingeniería Informática

Alejandra Guadalupe Silva Trujillo

Dirigida por:

Luis Javier García Villalba
Javier Portela García-Miguel

Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense de Madrid

Madrid, Noviembre de 2015

Dissertation submitted by Alejandra Guadalupe Silva Trujillo to the *Departamento de Ingeniería del Software e Inteligencia Artificial* of the *Universidad Complutense de Madrid* in Partial Fulfillment of the Requirements for the Degree of *Doctor por la Universidad Complutense de Madrid en el Programa de Doctorado en Ingeniería Informática.*

Madrid, 2015.

(Submitted November 1, 2015)

Title:

New Statistical Disclosure Attacks on Anonymous Communications Networks

PhD Student:

Alejandra Guadalupe Silva Trujillo (asilva@fdi.ucm.es)
Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense de Madrid
28040 Madrid, Spain

Advisors:

Luis Javier García Villalba (javiergv@fdi.ucm.es)
Javier Portela García-Miguel (jportela@estad.ucm.es)

This work has been done within the Group of Analysis, Security and Systems (GASS, `http://gass.ucm.es`), Research Group 910623 from the Universidad Complutense de Madrid (UCM) as part of the activities of different research projects. This research has been supported by the Ministerio de Defensa (MDE, Spain) through project UCM 321/2011, by the Agencia Española de Cooperación Internacional para el Desarrollo (AECID) of the Ministerio de Asuntos Exteriores y de Cooperación (MAEC, Spain) through project A1/037528/11 and by Safelayer Secure Communications S. A. through project UCM 307/2013, thanks to which part of this work was done during my stay in Computer Security and Industrial Cryptography (COSIC) Research Group of the Department of Electrical Enginnering (ESAT) at the Katholieke Universiteit Leuven (K. U. Leuven), Belgium.

This thesis is dedicated to my Family.

Acknowledgments

I would like to thank my supervisors, Javier García and Javier Portela, for everything I have learned from them and for all their support since the first day. Their fingerprints are present on all aspects of this research. This work would not have been possible without them.

The fruitful collaboration with the all the members of the GASS research group has also been crucially positive during the development of the ideas for this work. I feel really thankful towards them.

I would also like to thank my parents and friends for all the support I have received during this period.

Contents

List of Figures

List of Tables

Abstract

Anonymity is a privacy dimension related to people's interest in preserving their identity in social relationships. In network communications, anonymity makes it possible to hide information that could compromise the identity of parties involved in transactions. Nowadays, anonymity preservation in network information transactions represents a crucial research field.

In order to address this issue, a number of Privacy Enhancing Technologies have been developed. Low latency communications systems based on networks of mixes are very popular and simple measures to protect anonymity in users communications. These systems are exposed to a series of attacks based on traffic analysis that compromise the privacy of relationships between user participating in communications, leading to determine the identity of sender and receiver in a particular information transaction. Some of the leading attacks types are attacks based on sending dummy traffic to the network, attacks based on time control, attacks that take into account the textual information within the messages, and intersections attacks, that pretend to derive patterns of communications between users using probabilistic reasoning or optimization algorithms. This last type of attack is the subject of the present work.

Intersection attacks lead to derive statistical estimations of the communications patterns (mean number of sent messages between a pair of users, probability of relationship between users, etc). These models were named Statistical Disclosure Attacks, and were soon considered able to compromise seriously the anonymity of networks based on mixes. Nevertheless, the hypotheses assumed in the first publications for the concrete development of the attacks were excessively demanding and unreal. It was common to suppose that messages were sent with uniform probability to the receivers, to assume the knowledge of the number of friends an user has or the knowledge a priori of some network parameters, supposing similar behavior between users, etc.

This work proposes in the first place a framework to apply an universal Statistical Disclosure Attack in the sense that there are not special restrictions or knowledge assumed about user's behavior. The proposal includes a novel model schema using contingency tables, generation of feasible tables through a simulation algorithm and estimations of a measure for the ordering and final classification of each pair of users from highest to lowest relationship probability. Sensitivity analysis in a simulation framework is developed with

respect to factors such as the number of users, the mean rate of messages per unit of time, the number of relationships or the range of information retrieved by the attacker. The excellent results obtained about the classification of users relationships validate the present modeling approach.

The attack is then refined using EM algorithm under two types of probabilistic modeling of the rate of sent messages: the Poisson distribution and a discrete tabulated distribution. In the last case significant improvements are made lowering the error rate in the final cell classification and with respect to the estimation of the mean number of messages between each pair of users. These models have been checked with real email data obtained from the Calculus Center of the Complutense University of Madrid, leading to the first real application performed on real data of a Statistical Disclosure Attack.

One of the last works about Statistical Disclosure Attacks and first order reference in the state-of-the-art, where the general hypotheses settings are similar to our research framework, presents a modeling setting based on a least squares approach. In this work marginal totals of sent and received messages are used to estimate the conditional probabilities that a message obtained by receiver has been sent by each of the possible senders in the system. The comparison between this method en the method presented in our work is very positive favoring the last one. Results obtained are superior in our method on the application over real email data with respect to every metric analyzed: estimation of the mean number of messages between pairs of users and classification decision about the existence or not of relationships between each pair of users.

The attack method employed in this work is related to communications network with senders and receivers. These networks are present in contexts such as email or social networks. The email network data used for the application and performance study of the disclosure attack presented here is a particular case of social networks, where there can be studied different measures of individuals behavior (centrality, betweenness, etc.) or about the network itself (degree distribution, cluster coefficient, etc). Although this work principal aim was to classify relationship between users in existent or non existent, the information rerieved can be used to estimate the social network characteristic measures, relatives to individuals or to the whole network. This idea is performed obtaining accurate results when estimating the metrics involved.

Keywords: Anonymity, Anonymous Communication, Communication Patterns, Contingency Tables with Fixed Marginals, EM Algorithm, Email Data, Intersection Attack, Identity, Mixes, Privacy, Social Networks, Privacy Enhancing Technologies, Statistical Disclosure Attack, Traffic Analysis.

Resumen

El anonimato es una dimensión de la privacidad en la que una persona se reserva su identidad en las relaciones sociales que mantiene. Desde el punto de vista del área de las comunicaciones electrónicas, el anonimato posibilita mantener oculta la información que pueda conducir a la identificación de las partes involucradas en una transacción. Actualmente, conservar el anonimato en las transacciones de información en red representa uno de los aspectos más importantes.

Con este fin se han desarrollado diversas tecnologías, comúnmente denominadas tecnologías para la mejora de la privacidad. Una de las formas más populares y sencillas de proteger el anonimato en las comunicaciones entre usuarios son los sistemas de comunicación anónima de baja latencia basados en redes de mezcladores. Estos sistemas están expuestos a una serie de ataques basados en análisis de tráfico que comprometen la privacidad de las relaciones entre los usuarios participantes en la comunicación, esto es, que determinan, en mayor o menor medida, las identidades de emisores y receptores. Entre los diferentes tipos de ataques destacan los basados en la inundación de la red con información falsa para obtener patrones en la red de mezcladores, los basados en el control del tiempo, los basados en el contenido de los mensajes, y los conocidos como ataques de intersección, que pretenden inferir, a través de razonamientos probabilísticos o de optimización, patrones de relaciones entre usuarios a partir de la información recabada en lotes o durante un período de tiempo por parte del atacante. Este último tipo de ataque es el objeto de la presente tesis.

Los ataques de intersección pronto derivaron en el establecimiento de estimaciones estadísticas de los patrones de comunicación (número promedio de mensajes enviados, probabilidad de relación, etc.). Estos modelos comenzaron a denominarse Ataques Estadísticos de Revelación de Identidades, y pronto se demostró que eran capaces de comprometer seriamente el anonimato de los sistemas de redes basados en mezcladores. Las hipótesis planteadas en las primeras publicaciones para el desarrollo de estos ataques eran, sin embargo, excesivamente exigentes y poco realistas. Así, presuponían un envío de mensajes con probabilidad uniforme por parte de todos los usuarios, un conocimiento previo del número de amigos de un usuario o de algunos parámetros de red, comportamientos similares para todos los usuarios, etc.

En primer lugar, este trabajo propone un marco para aplicar un ataque estadístico de revelación de identidades universal, en el que la información se obtiene por el atacante en lotes y no hay requisitos previos sobre el comportamiento de los usuarios. La propuesta incluye una modelización novedosa (no utilizada previamente en la literatura) en forma de tablas, la generación de tablas factibles a partir de un algoritmo de simulación, y estimaciones de una medida para la ordenación y clasificación final de cada par de usuarios de mayor a menor probabilidad de relación. La experimentación realizada mediante el análisis de sensibilidad frente a diversos factores como el número de usuarios, la tasa de envíos, el número de relaciones o el tipo de información obtenida por el atacante permite concluir la validez de la presente propuesta, al obtener unos excelentes resultados en cuanto a la clasificación de las relaciones.

En segundo lugar, el ataque se refina utilizando el algoritmo EM bajo dos tipos de modelización probabilística de la tasa media de mensajes enviados: la distribución de Poisson y una distribución tabulada discreta. En este último caso se obtienen mejoras significativas en la clasificación final y en la estimación del número medio de mensajes por cada par de usuarios. Estos modelos han sido corroborados con datos reales de correo electrónico, facilitados por el Centro de Cálculo de la Universidad Complutense de Madrid, siendo además la primera vez que un tipo de ataque estadístico de revelación se contrasta con datos reales.

Uno de los últimos trabajos sobre ataque de revelación de identidades y referencia imprescindible en el estado del arte, y el que además plantea hipótesis generales que son asimilables a la presente investigación, presenta un esquema basado en el método de mínimos cuadrados en el que se relacionan mediante un enfoque de regresión los totales marginales de los mensajes enviados y recibidos por los usuarios con las probabilidades condicionales de que un mensaje recibido por un determinado usuario haya sido enviado respectivamente por cada uno de los usuarios del sistema. La comparativa de este método de la literatura con el esquema propuesto es tremendamente positiva a favor de este último, obteniéndose resultados superiores en su aplicación sobre datos de correo electrónico en todas las métricas analizadas: estimación del número medio de mensajes entre cada par de usuarios y decisión de clasificación sobre relación o no entre usuarios.

El planteamiento del ataque presentado en este trabajo se asocia a redes de comunicación con emisores y receptores. Estas redes están presentes en contextos como

redes de correo electrónico y redes sociales. Los datos de correo electrónico utilizados para la aplicación y comprobación del comportamiento del ataque mencionado son un tipo particular de redes sociales, para las cuales existen medidas características basadas en los individuos (grado de centralidad, intermediación, etc.) o en la red en sí (coeficiente de agrupamiento, distribución de grado, etc.). Si bien el objetivo inicial de este trabajo era clasificar las relaciones entre usuarios como existentes o no existentes, o estimar el número de mensajes promedio por ronda entre cada par de usuarios, también puede aplicarse para estimar medidas de características propias de las redes sociales relativas a los individuos o a la red en general. Esta idea se aplica obteniendo resultados de interés en cuanto a la estimación de medidas de centralidad y otras características de estas redes.

Palabras clave: Algoritmo EM, Análisis de Tráfico, Anonimato, Ataque de Intersección, Ataque Estadístico de Revelación de Identidades, Comunicación Anónima, Datos de Correo Electrónico, Identidad, Mezcladores, Patrones de las Relaciones, Privacidad, Redes Sociales, Tablas de Contingencia con Marginales Fijos, Tecnologías para Mejorar la Privacidad.

Part I

Description of the Research

Chapter 1

Introduction

1.1 The Importance of Information Security in Protecting Privacy

Today, organizations are placing a tremendous amount of collected data into massive repositories from various sources, such as: transactional data from enterprise applications and databases, social media data, mobile device data, documents, and machine-generated data. Much of the data contained in these data stores is of a highly sensitive nature and would trigger regulatory consequences as well as significant reputation and financial damage. This may include social security numbers, banking information, passport numbers, credit reports, health details, political opinions and anything that can be used to facilitate identity theft.

Our daily activities are developed in a digital society where the interactions between individuals and other entities are through technology. Now, we can organize an event and send the invitation using a social network like Facebook, sharing photos with friends using Instagram, listening to music through Spotify, asking for an address using Google Maps; all of these activities are just some of the ways in which many people are already working on the Internet every day. Personal information in real world is protected from strangers but it is different in the online world, where people disclose it [Kri13].

All available information about a person gets cross-referenced, and the resulting dossier ends up being used for many purposes, lawful and otherwise. This practice has expanded over the years; the companies that compile and sell these dossiers are known as data brokers. The communication systems behaviour has changed and it has been forced to

improve its management in order to protect users privacy and satisfy the new requirements. Economists, sociologists, historians, lawyers, computer scientists, and others have adopted their own privacy definitions, just as the value, scope, priority and proper course of study of privacy. Details about the background, law and history of privacy are showed in [GD11]. According to experts, privacy and intimacy are difficult concepts to define. However, we may consider personal health conditions, identity, sexual orientation, personal communications, financial or religious choices, along with many other characteristics. References from literature on how privacy solutions are applied from economic, social and technical areas are in [BGS05] [NS09] [GA05].

Respect for privacy as a right includes undesirable interference, the abusive indiscretions and invasion of privacy, by any means, documents, images or recording. The legal foundations date back to 1948. In that year, the Universal Declaration of Human Rights was released, in which it was established that no person "shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation". However, despite the legal and political developments that have taken place since then, it has not been possible to solve a fundamental problem to curb abuses every day. The lack of clarity and precision in the right to freedom of expression and information limits is an open issue; cases that threaten these rights are increasing.

The development of digital media, the increasing use of social networks, the easier access to modern technological devices, is perturbing thousands of people in their public and private lives. Examples abound, the most recent was the deputy mayor of a Flemish town, who was caught and recorded on a video while having sex with a man in the Town Hall offices. The recording was made and released for an unknown group of young boys. Another scandal was the president of the Guatemalan Institute of Social Security, whowas shot in his office committing "lewd acts". Unlike the previous one, in this case there was a crime and the action given was justified publicly. All of this stuff is available on the Internet and traditional media, the videos anonymous communications

There are two perspectives on user side: One way is to accept a complete loss of privacy in exchange to the benefits of using technology. The other side is, do not get involved into technological tools, and being outside of the digital world. Both extremes are radical; the optimal way is to maintain control over the personal data and to take advantage of the

benefits of technology without allow a privacy intrusion.

Governments and industry take advantage of sophisticated data storage tools and are using it to profile their users for financial, marketing, or just statistical purposes; organizations are able to acquire and maintain massive infrastructure at bargain prices and this derives to multiple benefits.

Individuals have the right to control their private information and only provide it to certain third parties. In the last decade users privacy concerns have grown [DJR12] [CMD09] [GA05] and since then several technologies have been developed to enhance privacy. Privacy enhancing technologies (PETs) are designed to offer mechanisms to protect personal information, and can be used with high level policy definition, human processes and training in the use of computer and communication systems [GWB97] [Gol03] [Gol07b]. PETs have been proposed to defend users privacy in user, network and server areas. Private and public organizations, as well as individuals should include the protection of privacy besides the typical aspects like integrity, confidentiality and availability of data.

Privacy protection must avoid the disclosure of identities in a communication system. Motivations of these issues include censorship resistance, spies or law enforcement, whistleblowers, dissidents and journalists living under repressive regimes.

There are some technologies used to accelerate the transition to encryption as a service including hardware-based encryption key storage, centralized data protection schemes for applications, databases, storage and virtualized environments, as well as role-based access controls. Despite significant investment in security technology, organizations have a great hole in security effectiveness. This is due to the fact that conventional defenses rely on IP addresses and digital signatures. Signatures used in antivirus and intrusion prevention systems are effective at detecting known attacks at the time attacks are launched. They are not effective, however at detecting new attacks and are incapable of detecting hackers who are still in the reconnaissance phase, probing for weakness to attack. IP reputation databases, meanwhile, rely on the notion that attackers can be identified by their IP addresses, and so share this information across systems. Unfortunately, this is as ineffective method as it uses a postal address to identify someone. Network attacks are a serious threat to an organization. Next generation technologies are encouraged to improve the encryption solutions available. However, it has been proved that traffic and network topology analysis

do not provide enough users privacy protection, even when anonymization mechanisms are applied. Using auxiliary information, adversaries can diminish anonymity properties.

Every year several research centers publish reports after analyzing the tendencies over Security Information subjects. In 2011, the main tendency was on botnets and malware techniques oriented to achieve economical benefits. But, in 2012, tendencies were focused on mobile devices. One year later, the major topic was the huge number of available threats for mobiles, and nowadays, these threats are continuing growing. Now, the main users' concern is centered on privacy data.

The Snowden report including National Security Agency collecting activities was the main motivation for users' concerns on privacy. It is a positive step for bring awareness to society. Concern about privacy is a positive starting point, even when this tendency has not diminishing the feelings about people affected by a malicious code or any other informatics threat. There is a better understanding about privacy that helps society to become aware of information security areas but, more than being aware it is important to take actions to mitigate it. This situation is like a person being worried for her home security that installs a system security alarm but she leaves the windows at home opened.

From the extensive use of Internet and some other services like web browsers, social networks, webmail, and others, privacy has become more important not just for researchers on the subject, enterprises and society are also involved.

Anonymity systems provide mechanisms to enhance user privacy and to protect computer systems. Research in this area focus on develop, analyze and execute anonymous communication networks attacks. Even when communication content has been ciphered, information routing needs to be sent clearly for routers to know the next package's destination in the network. Every data packet traveling in the Internet contains the node addresses of sending and recipient nodes. So, it is well understood that actually any packet cannot be anonymous at this level. Figure 1.1 shows an example of normal communication and anonymous communication model. In the second approach, several clients use a network of mixes in order to hide his identity. The network of mixes provides all clients the same IP address, letting them indistinguishable.

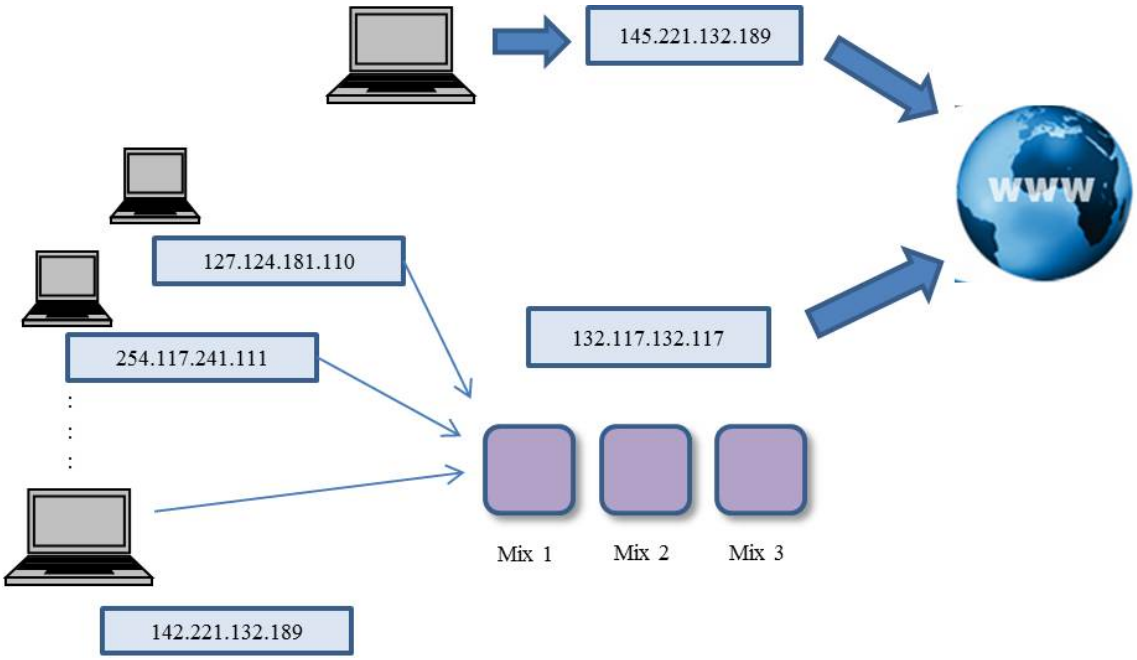


Figure 1.1: Anonymous vs. Regular communication model

Anonymity is the state of absent identity; therefore anonymous communication can only be achieved by removing all the identifying characteristics from the anonymized network. Let's consider a system as a collection of actors, such as clients, servers, or peers, in a communication network. These actors exchange messages via public communication channels. Pitfzmann and Hansen [PH08] defined anonymity as "the state of being not dentifiable within a set of subjects, the anonymity set".

One of the main characteristics of the anonymity set is its variation over time. The probability that an attacker can effectively disclose the message's sender is exactly 1/n, with n as the number of members in the anonymity set. The research on this area has been focused on developing, analyzing and attacking anonymous communication networks. The Internet infrastructure was initially supposed to be an anonymous channel, but now we know that anyone can be spying in the network to reveal our data. Attackers have different profiles such as their action area, users volume capacity, heterogeneity, distribution and location. An outside attacker may identify traffic patterns to deduce who has communication with whom, when, and its frequency.

There are three different perspectives on anonymous communication: (i) Sender anonymity: Sender can contact receiver without revealing its identity; (ii) Receiver anonymity: Sender can contact receiver without knowing who the receiver is; (iii)

Unlinkability: Hide your relationships from third parties. According to [PH08] unlinkability between two items of interest occurs when an attacker of the system cannot distinguish if the two items of interest (in a system) are related or not.

Over the past years, anonymous communications has been classified by two categories: high latency systems and low latency systems. The first ones aim to provide a strong level of anonymity but are just applicable for limited activity systems that do not demand quick responses, such as email systems. On the other hand, low latency systems offer a better performance and are used in real-time systems. Examples include web applications, secure shell and instant messenger. Both systems are built on a reflection of Chaum's proposal [Cha81]. Unlinkability is provided in a similar way in both cases using a sequence of nodes between a sender and its receiver, and using encryption to hide the message content. An intermediate node knows only its predecessor and its successor.

The mix networks systems are the basic building blocks of all modern high latency anonymous communication systems [Cha81]; On the other hand, several designs have been developed to provide anonymity in recent years with for low latency systems, such as Crowds [RR98], Hordes [LS02], Babel [GT96], AN.ON [BFK01], Onion routing [GRS96], Freedom [BGA01], I2P [Del09] and Tor [DMS04]. Nowadays, the most widely used anonymous communication network is Tor; allowing anonymous navigation on the web. Tor forwards traffic through multiple relays. Tor purpose is to keep web traffic anonymous by delaying or altering the packets of data that are sent through servers, making it look like the traffic is coming from a place that it's not actually (the IP address that the server "sees" is called an "exit node"). There have been several attacks to Tor, one of them occurs if the end server of the site visited can detect the origin point also called, the "entry guard" or "entry relay, then anonymity is lost. In Figure 1.2 we show the Tor model.

A comparison of the performance of high latency and low latency anonymous communication systems is showed in [Loe09].

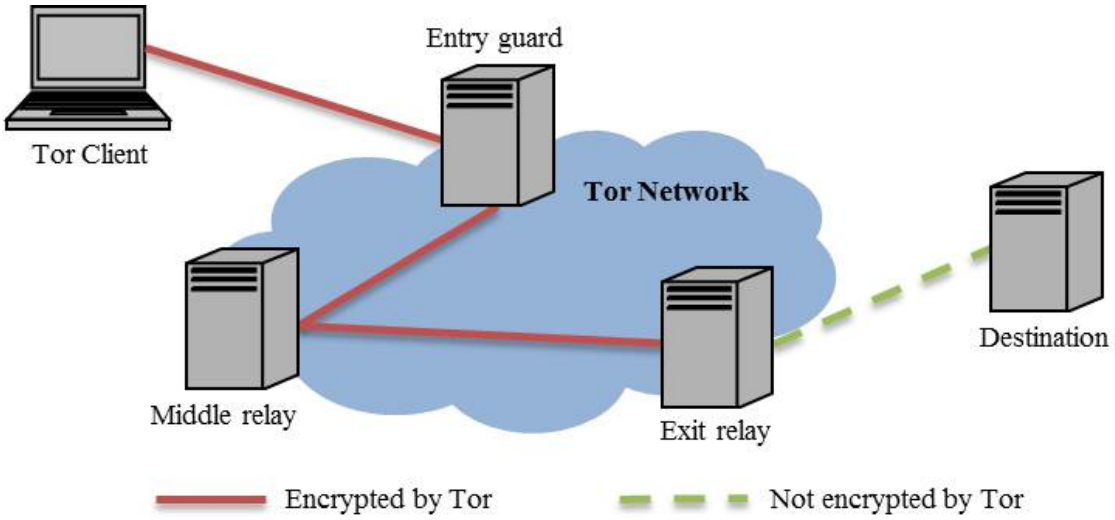


Figure 1.2: The Tor model

1.2 Mix Systems Attacks

The attacks against mix systems are intersection attacks [Ray01b]. They take into account a message sequence through the same path in a network, it means performing traffic analysis. The set of most likely receivers is calculated for each message in the sequence and the intersection of the sets will make it possible to know who the receiver of the stream is. Intersection attacks are designed based on correlating the times when senders and receivers are active. By observing the recipients that received packets during the rounds when Alice is sending, the attacker can create a set of Alice's most frequent recipients, this way diminishing her anonymity.

1.3 Summary of Contributions

This thesis has four main contributions to the field of Statistical Disclosure Attacks. The first and main contribution consists in a novel and general modeling approach to deploy a statistical attack over networks of mixes [PGVST$^+$15] [GMRCnSO$^+$12]. This method is presented in Chapter 4. The second contribution consists in an important improvement of the attack through the use of the EM algorithm. The third contribution consists in the application of the attack over real mail data [STPGMGV14] [PGMGVST$^+$15] [STPGMGV15a]) ; this last is a relevant advancement since this is the first time this genre of attack appears in the literature under real application settings.

In the fourth place, a comparison is made with one of the best state of the art methods, where the attack presented here is showed to perform better on real data. Second, third and fourth results are presented in Chapter 5. The last contribution is the implementation of the attack with the aim of estimating characteristic measures in a social network data framework [STPGMGV15c]. This application is presented in Chapter 6.

1.4 Outline of the Thesis

This thesis is organized as follows:

Chapter 2 introduces the basic concepts of anonymity and privacy. There will be exposed the principal concerns about the development of protection measures and a taxonomy of privacy violations is set. The principal anonymity metrics are also presented. It is shown how the birth of new Privacy Enhancing Technologies addresses the issue of Privacy protection and the history and taxonomy of Privacy Enhancing Technologies is examined. The concept of Mixes for network protection is also developed since this is the protection the methods presented here aims to attack.

Chapter 3 presents the concept of traffic analysis on anonymous communications and establish the different type of attacks present in the literature. A special effort is made in detailing the procedures of the attacks named intersection attacks and statistical disclosure attacks. The evolution of the ideas used in this genre of attacks is studied. Focus is set on the hypothesis assumed on the data retrieved by the attacker for modeling the attack, since it is one of the weak points of the attacks present in the literature [GVSTP15] [STGVD08] [STGV08].

Since previous attack methods lack of realistic assumptions to apply on real data, a new modeling approach for a general statistical disclosure attack is presented in Chapter 4. The use of contingency tables with fixed marginals to represent the data retrieved by the attacker allows to obtain estimates for the messages sent based on an algorithm that generate feasible tables. The results are used to obtain an ordering for the cells of the adjacency matrix that leads to a classification framework where the attacker classifies each pair of users in friends that communicate or not. Sensitivity analysis is developed in a simulation framework. Showing this is a promising method to develop an attack [PGVST$^+$15] [GMRCnSO$^+$12].

The attack presented in Chapter 4 is improved in Chapter 5 through the use of the EM algorithm to refine the estimates. A Poisson distribution version and a discrete tabulated distribution version are compared, obtaining better results in the last modeling approach. The modification of the method is shown to derive on better classification measures than the version method presented in the previous chapter and is then considered the state of the art of our attack from now and on.

In the next section, our method is proved on real email data. This is the first time a statistical disclosure attack is proven against real data. Results are encouraging, obtaining high rates of good classifications with moderate settings of attacker retrieved data [STPGMGV15b] [PGMGVST$^+$15] [STPGMGV15a].

One of the last works about Statistical Disclosure Attacks and first order reference in the state-of-the-art, where the general hypotheses settings are similar to our research framework, presents a modeling setting based on a least squares approach. he comparison between this method en the method presented in our work is very positive favoring the last one. Results obtained are superior in our method on the application over real email data with respect to every metric analyzed: estimation of the mean number of messages between pairs of users and classification decision about the existence or not of relationships between each pair of users.

The attack method employed in this work is related to communications network with senders and receivers. These networks are present in contexts such as email or social networks. The email network data used for the application and performance study of the disclosure attack presented here is a particular case of social networks, where there can be studied different measures of individuals behavior (centrality, betweenness, etc.) or about the network itself (degree distribution, cluster coefficient, etc). Although this work principal aim was to classify relationship between users in existent or non existent, the information rerieved can be used to estimate the social network characteristic measures, relatives to individuals or to the whole network. This idea is performed in Chapter 6, obtaining accurate results when estimating the metrics involved [STPGMGV15c].

Chapter 2

Anonymity and Privacy

The aim of this chapter is to show an introduction of anonymity and privacy considering different knowledge areas. It is organized into ten sections. First, it describes the anonymity types and degrees as well as Privacy definition and general information related. In section 2.4 is identified the Taxonomy of privacy continuing with the development of privacy. Considering the taxonomy of privacy it is shown the main motivation of this area, taking into account the risks and vulnerabilities on real world. Section 2.5 shows the anonymity metrics. Section 2.6 explains the legislation around this topic. Section 2.7 presents the State of Art of Privacy Enhancing Technologies which includes: Its history, Triggers, and Categorizations. Section 2.8 defines the Mix network and the classification of anonymous communication systems. Mix network is a fundamental basis of high and low latency anonymous communication systems, which aims to hide the relationship of a message with its corresponding sender and receiver.

2.1 Types of Anonymity

In order to list the types of anonymity, first we must define anonymity. Anonymity is the state of being not identifiable within a set of subjects the anonymity set. The anonymity set is the set of all possible subjects who might execute an action [PH08]. Figure 2.1 shows the basic model of anonymity.

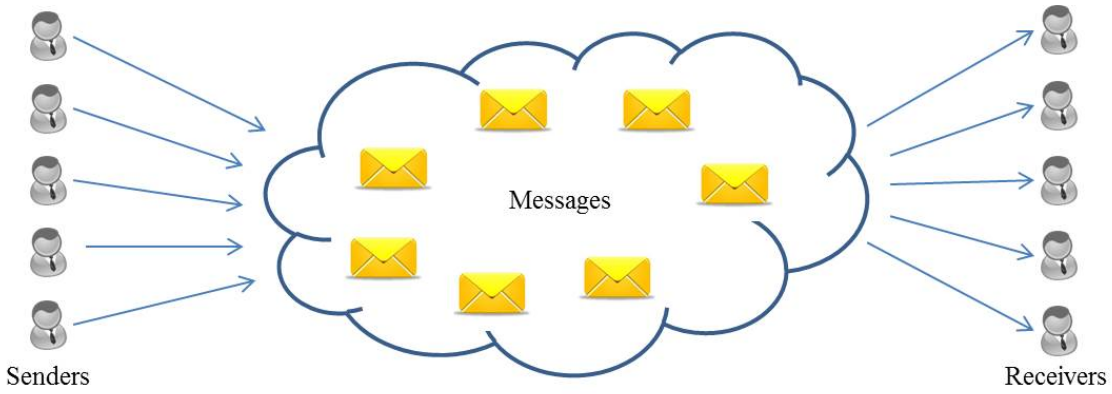


Figure 2.1: Basic anonymity model

Anonymity is a legitimate way in many applications such as web browsing, e-vote, e-bank, e-commerce and others. Popular anonymity systems are used by hundreds of thousands people, such as journalists, whistle blowers, dissidents, and others. It is well known that encryption does not guarantee the anonymity required for all participants.

Attackers can identify traffic patterns to deduce who, when and how often users are in communication. Communication layer is exposed to traffic analysis, it is necessary to anonymize it, as well as application layer that support anonymous cash, anonymous credentials and elections. Anonymity systems provide mechanisms to enhance user privacy and to protect computer systems.

- **Sender**: Receiver / observer can't identify sender.
- **Receiver**: Observer can't identify receiver.
- **Sender-receiver**: Observer can't identify that communication has been sent.
- **Unlinkability**: To hide the association of sender and receiver.

Several researchers have focused their experiments on represent the intuitive properties of anonymity channels. Most proposals fall into two classifications (a) weak definitions, focused on particular applications to require efficiency; (b) stronger definitions, oriented to complex applications most of the time impractical. There is also a list of several mechanisms to achieve anonymity and unobservability. Formal definitions of unlinkability, sender-anonymity, receiver-anonymity, sender-receiver anonymity and unobservability are shown in [HM08].

Sender unlinkability and receiver unlinkability are considered the weakest notions of anonymity. A protocol is sender-unlikable when is able to hide the relationship between senders and receivers. In the other side, the strongest notions are Sender-Receiver Anonymity and Unobservability. Sender-receiver unlinkability strength the requirements for receivers, hiding the sent and received messages values, but not necessarily the total size of exchanged messages. Sender anonymity is defined as the number and values of messages for the sender must keep hidden, but not the values of the received messages for each party. Receiver anonymity works in the same way, the difference is that it reverses the roles of sender and receiver.

2.2 Types of Anonymity

The degrees of anonymity can be modeled as a continuous line divided by six sections that represent each degree [RR98]. In the right side of the line is absolute privacy degree of anonymity, and in the opposite side is provably exposed as is shown in Figure 2.2.

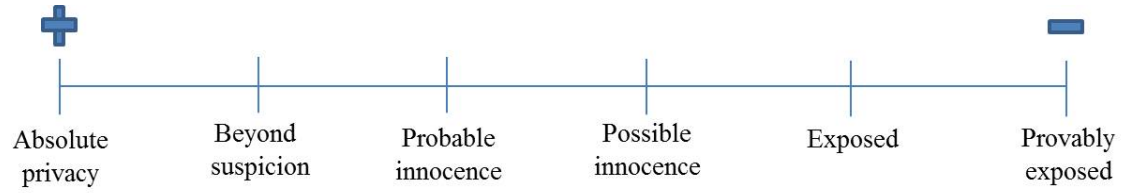


Figure 2.2: Degrees of anonymity

- **Absolute privacy**: Attacker cannot perceive the presence of communication.
- Beyond suspicion Attacker can observe the evidence that a message has been sent, the sender has the same probability of being the originator of that message that any other possible sender in the system.
- **Probable innocence**: A sender is probably innocent if, from the attacker's perspective, the sender appears no more likely to be the originator than to not be th originator.
- **Possible innocence**: A sender is possibly innocent if, from the attacker's perspective, there is a nontrivial probability that the real sender is someone else.

- **Exposed** From the attacker's point of view there is a high probability about who is the sender.

- **Provably exposed**: The attacker can identify the sender/receiver identity and he is able to prove it.

2.3 Privacy

Nowadays privacy is one of the most important topics for security communications. The enormous use of mobile devices, Internet and online applications is a growing phenomenon that has led to develop new techniques and technologies to provide users safe environments. Several research groups have been given the task of developing applications that support this purpose. It is well known that government and organizations want to take advantage of every piece of data we leave on Internet, habits of purchasing, health records, and many other actions for legal (or non legal) purposes. Such practices are sheltered under the argument that monitoring online activities are necessary to detect potential threats that could undermine national security.

Privacy is one of the fundamental human rights. There are several concepts of privacy, from different areas of knowledge such as technological, ethical, philosophical, political, and others. In [Wes68] define privacy as an individual right to control, edit, manage, and delete information about themselves and decide, when, how, and to what extent information is communicated to others.

There are several risks in privacy area such as: unsolicited marketing, price discrimination, disclosure of private information and government surveillance, among others.

Acquisti and Grossklag [AG04] have addressed the topic of why people express high preference of privacy when some interviewed by phone, but on their online behavior show very low preference. In [WSSV15] a research result conducted to 179 people is shown. The research aim was to ask them about their privacy concerns and how to protect themselves of Online Behavioral Advertising. A notable result was the fact of knowing that users with higher levels of study or knowledge actually did not carry out actions to protect their privacy, as they believe that their actions will have little effect against companies that use these techniques to increase the effectiveness of their advertising. In this sense, when a

user visits a website, it is recorded for how long, through which medium it was reached the page, what keywords were used, among other details; with all the information above, the company's website creates a user profile. By using user profiles, it is possible to define audiences based on user preferences. When a user returns to a website using the same browser, such profiles are used to offer products or services that might be of interest.

Various legislative responses have emerged, mainly in European countries, where companies are required to allow consumers to choose whether to participate or not regarding the collection of their data. It means that consumers should explicitly indicate their consent to data collection; meanwhile "opt-out" requires that consumers explicitly prohibit collecting them. An email campaign user on opt-in mode means that, he wants to receive regular news or information, which could include commercial advertising and others. The opt-out mode (choose not to do something) refers to various methods which users can avoid receiving unsolicited information products or services. This ability is usually associated with direct marketing campaigns. For example, in the USA the national credit bureaus offer a toll free number that allows consumers to opt out on all pre-approved credits and insurance offers with just a phone call.

In [VWW05] a study about blacklisting of telemarketing showed that people with more education are the most likely to sign up; but the question arises of the reasons, it will be because they value their time, or because better understand the risks, or because they receive more calls.

The lack of anonymous spaces is a neuralgic issue related to the formation of a particular model of society where it is possible to ensure the protection of minorities, dissident groups, citizens of repressive totalitarian regimes and anyone who simply does not want to disclose his identity.

The right to privacy for everyone is guaranteed under Article 12 of the Organization of United Nations in the Universal Declaration of Human Rights in 1948: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the law protection against such interference or attacks".

In [Wes68] four basic states of individual privacy are identified:

1. Solitude.

2. Privacy.

3. Anonymiy.

4. Reserve.

In 1998, Kang [Kan98] defined privacy as a union of three overlapping sets of ideas:

1. **Physical space**: the extent to which the solitude of a territorial individual is invaded by objects or unwanted signals.

2. **Election**: the ability of an individual to make important decisions without interference.

3. **Personal information flow**: the control of a person over the process, it means, the acquisition, dissemination and use of personal information.

Privacy is recognized around the world in various regions and cultures. It is protected by the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and many other international and regional human rights organizations. Most countries have included in their constitution at least the minimum provisions that guarantee the inviolability of the rights at home and the secrecy of communications.

Of all human rights in the international catalog, privacy is perhaps the most difficult to define [Mic94]. In many countries, the concepts of privacy and data protection have merged, which is interpreted as personal information management.

The EPIC [Cen07] considers the following concepts related to privacy:

1. Privacy of information, which involves the establishment of rules governing the collection and processing of data such as credit information, medical records and government services. It is also known as data protection.

2. Privacy body, which corresponds to the protection of the physical space of individuals against invasive procedures such as genetic and drugs screening, and cavity search.

3. Privacy in communications, covering the security and privacy in mail, telephone, and email services, and any other form of communication.

4. Privacy territorial, which corresponds to the composition of limits on intrusion into domestic environments and others, such as the workplace or public spaces. These intrusions include video surveillance through and identifiers verification.

The PISA Consortium (Privacy Incorporated Software Agent) [Con03] defines privacy as "the right of individuals to be alone, free of surveillance or interference from other individuals, organizations or the state". This definition includes a set of rules of conduct between people and the environment related to personal information management. Personal data can be defined as the collection of all data that are or may be related to an individual, his identification, physical data, social and financial behavior, and other personal data related.

2.4 Taxonomy to Identify Privacy Violations

Each attempt to classify violations of privacy has its restrictions because several cases do not fit within the classifications that have been considered and it were proposed before the digital age of today. In a more recent context Solove [Sol06] exposes the following taxonomy with the intention of identifying and understanding the different types of privacy breaches:

1. **Information Collection**: Surveillance, Interrogation.
2. **Information Processing**: Aggregation, Identification, Insecurity, Secondary Use, Exclusion.
3. **Dissemination of information**: Breach of confidentiality, Disclosure, Exposure, Increase Accessibility, Blackmail, Appropriation, Distortion.
4. **Invasion**: Intrusion, Decisional Interference (government interference on subject personal decisions about his life).

Each of these groups consists of various subgroups of harmful activities. In an attempt to model the relationship between these groups see Figure 2.3.

Violations of the privacy of an individual can be categorized according to the damage or problem incurred. In order to give some examples: If a newspaper article describing a crime publishes the name of the victim; the installation of surveillance cameras in different parts of the city; there are new applications that allow you to view photos of real scenes on the Internet, such as Google maps application, called Street view [Goo15]; the massive use of social networks application based on user information used for different and sometimes non-legal purposes.

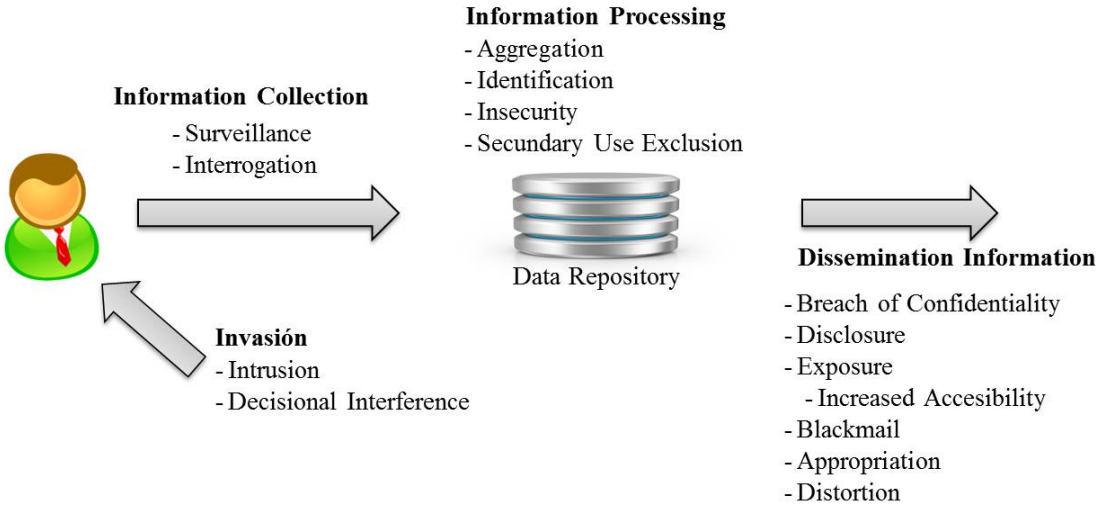


Figure 2.3: A taxonomy of privacy

The argument of "I have nothing to hide" [Sol07] is used when the subject of surveillance activities and data mining carried out by the government is discussed; many people respond: "I have nothing to hide". Many people believe that there is no privacy danger unless the government does not cover illicit activities; in this case a person does not have a legitimate justification to claim that they kept private. The lack of privacy is not an issue for individuals who have nothing to hide; so that if the individual is located only in legal activities, he has nothing to worry about.

In countries like the UK, the government has installed millions of video cameras to monitor public streets of major cities and towns, which are monitored by officers through a closed-circuit television. The slogan of a campaign for the government program said "If you have nothing to hide, you have nothing to fear".

Show personal details in public spaces, may cause many adverse effects, due to the fact that electronic infrastructure today facilitates the collection of information. The main dangers fall into the following three classes [Gro05]:

1. **Loss of confidentiality - abuse of personal information**: In the same way that individuals might feel a physical offense because their homes or other private areas intrusions into personal electronic records or the exposition of personal information it is more often considered an offense or a danger to individuals.

2. **Identity theft**: In many situations a simple identity data such as social security number is accepted as proof of identity to apply for medical services or financial

transactions. On the one hand, this facilitates procedures for citizens, but leaves open the possibility of fraud through identity theft. Within the obvious risks they are of course substantial financial losses to give access to certain services based on a stolen identity, but there may be worse consequences: get false certificates, passports, driving licenses, bank loans, among others. In recent years, it has increased the phenomenon of phishing where through false websites and using social engineering is seeking confidential information fraudulently. The term phishing comes from the English word "fishing", and is the contraction of "password harvesting fishing" (harvest and phishing) referring to the act of fishing users through increasingly sophisticated lures.

3. **Unsolicited messages**: One of the most sensitive issues in privacy protection is the fight against unsolicited emails or spam. Unlike the physical world, where use messaging services has a cost, in current electronically context this cost is negligible, so it becomes an ideal solution for bulk mailing of commercial information. Email is the most used tool for these purposes, but also online services such as logs, SMS's, forums, and others.

The privacy risks are not well defined in the literature. People, who use the technology and give much value to the benefits offered by digital technologies, are willing to disclose their privacy in exchange for those benefits. This behavior is different in different countries. For example in India, consumers seem more willing to negotiate their privacy in exchange for goods; at the opposite extreme is Germany, where consumers have a tremendous notion of risks to compromise their privacy. A study of 15,000 people from 15 countries [Cor14] revealed that although consumers have experienced some type of damage to their privacy, the reality is that, no actions are taken to protect their information, even the basic ones considered such as change passwords regularly or use passwords on mobile devices. Most of the people consider that is responsibility of the government and not from themselves, protecting the privacy of consumers through laws and regulations. Moreover, consumers do not believe the government or companies can truly protect their privacy by the lack of ethics and transparency that has emerged in recent years. To mention some examples are the revelations disclosed by a former NSA consultant, Edward Snowden, or the exposure of various governments as Mexico, Sudan, Colombia, Russia, of having hired the services of a company that sold software to carry out espionage activities of its citizens, or the

multiple penalties on Google and Facebook for violating the rules on data privacy.

There has not been a convincing classification considering the privacy risks associated to cost. In Table 2.1 the risks and costs of privacy from the point of view of the trader and the consumer are examined [Gel02]. If the quality of protection technologies is very low, it could destroy applications completely. The protection of privacy, like most things, it has its benefits, costs and consequences.

Table 2.1: Privacy risk associated with its cost

Dealer	**Consumer**
Lost sales due to lack of privacy	High prices
The loss of sales to a merchant will be the opportunity for another	Spam, Telemarketing
International opportunities lost	Identity Theft
Increases cost legal advice	The effects of Internet: service inefficiency and delays due to spam

We can categorize the properties of privacy in two groups.

- **Anonymity**: Anonymity is defined as the condition of being unknown in a set of entities, the anonymity set [PH08]. The entities which might be related to an anonymous transaction are part of the anonymous set for that particular transaction. An entity performs an anonymous transaction if he cannot be differentiated (by an opponent) of other entities. This definition has been adopted in many literature of anonymity, reflecting the probabilistic information obtained by adversaries trying to identify anonymous characters.

- **Not observable**: A user activity is hidden. Formally, it is defined as the state of being indistinguishable of others user. An anonymous system has no observable property if an attacker cannot determine which user did a particular activity from a set of users that might be senders and a set of users who might be the recipients. This property ensures that a user can use a resource or service without others observes this resource or service is being used. The parties not involved in the communication cannot observed either sending or receiving messages. In the second group the concepts of different actions have been executed by the same identity are included.

- **No link**: The relationship between two or more actions is hidden. Many anonymous systems can be modeled in terms of no linkage. This property is defined as:

Non-linking two or more elements, called items of interest (for example, subjects, messages, events, actions, etc.) means that these elements are within this system, but they are not more or less related according to knowledge a priori, for example sender-reciver (anonymous delivery) merchant-buyer (anonymous authentication with electronic money), electronic voting. Considering the messages sent and received as items of interest (IOIs), anonymity can be defined as a non linking IOI to an identity [PH08].

- **Pseudononimity**: This property allows multiple actions being linked to a single actor whose identity is protected. A pseudonym is an alias name or other identifier that removes the actual name of an entity, but it is used as a means to relate it to an entity. It is the state to use a pseudonym as an identifier. We can distinguish two types of aliases: the one-time use and the persistent.

A pseudonym can model roles, transactions, people, and relationships with different degrees of anonymity.

Figure 2.4 represents an abstract model of anonymity [Dó5]. As mention before anonymity systems hide the relationship between the entities and items of interest. The basic mechanism behind a system of anonymous transactions is based on hiding the relationships between entities and items of interest. The set of entities that may be related to an item of interest is called the anonymity set. If the anonymity set is bigger, all entities involved will have higher level of anonymity.

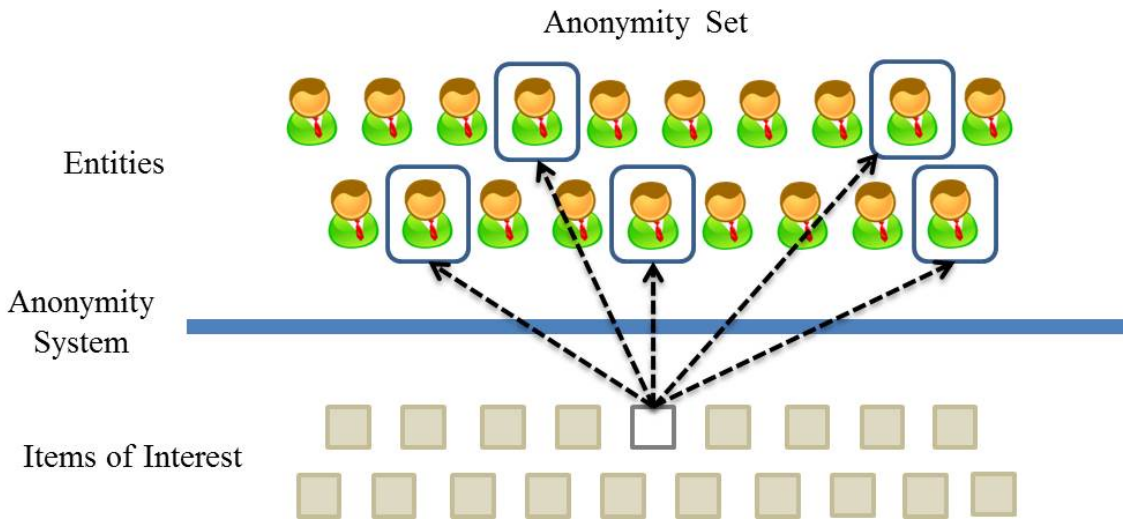


Figure 2.4: Abstract model of anonymity

2.5 Metrics on Anonymity Systems

The need of a metric to measure efficiency in an anonymous system, arise with the development of applications to enhance privacy and allow anonymous electronic transactions, such as electronic voting, electronic money or web browsers.

Questions around this topic are: How to measure the level of anonymity?, how can compare two different anonymous systems?, Is there a general measure of anonymity that can be applied to any anonymous system?, how to evaluate effectiveness of different attacks on anonymous systems?, how can you quantify the loss or gain on anonymity?, how can the metrics of anonymity reflect partial or statistical information collected by an attacker?, what is the optimum level of anonymity?

Several researches have tried to answer these questions, but there are some to be done, for example to define what is a sufficient level of privacy is an issue that requires further investigation, as it depends on the context and the needs of specific application, ranging beyond a technical dimension.

Before the theoretical information anonymity metrics were proposed, there were some attempts to quantify anonymity in communication networks.

Reiter and Rubin [RR98] define the degree of anonymity as a probability $1 - p$, where p is the probability assigned for an attacker to potential receivers. In this model, users are more anonymous than they appear (against some adversary) to be less likely to have sent a message. This metric considered separately users and therefore does not capture well the properties of anonymity. Consider a system with 2 users who appear to be the sender of a message with probability 1/2. Now, consider another system with 1000 users. The user u1 appears as the sender with 1/2 probability while other users are given the chance of having sent 0.001 the previous message. According to the definition of Reiter and Rubin [RR98], the degree of anonymity of u1 and two users of the first system would be the same (50%). Anyway, in the second set, u1 appears more likely to be the sender that any user, while the two users of the first system are indistinguishable to the attacker.

Berthold et al. [BPS01] define the degree of anonymity as A = log2 (N), where N is the number of users in the system. This metric only depends on the number of users of the system and therefore does not express anonymity properties of different systems. The first anonymity metrics approaches are [SD03] and [DSCP03], the first use entropy to measure the size of the effective anonymity set, while the second goes a step further,

normalizing entropy to get a degree of anonymity on the scale of 0-1.

2.6 Privacy Legislation

A significant number of countries governed by democratic systems, have recognized the protection of personal data in agreements, EU directives, national laws and other legal statutes. For example, the guidelines of the OECD provide a backup to privacy individual information and the related rights; they also indicate restrictions and obligations of the parties to the exchange and storage of personal data. The concern of regulate such mechanisms arises with increased automatic personal data processing through computer systems, databases, electronic communications and other devices. Unfortunately, despite the laws that have been made, there is a well-established and globally accepted way to protect privacy.

For more than a century in the Supreme Court of the USA, privacy was defined as the "right to be left alone", which is one of the most esteemed rights. The concept itself of "being alone" is not enough to define the concept of privacy in the current digital environment where communication systems through electronic media are widely used in the interaction between individuals, businesses and public institutions. Many types of information are generally considered private or confidential nature, typical examples of sensitive information are: medical, financial, religious and political preferences records. But depending on the context, even trivial information as buying preferences, records of phone calls and the geographical position can be highly sensitive.

A key element of privacy is the ability to relate different pieces of information. The combination of information from different sources can lead to a violation of an individual privacy, even if the information from each source is considered trivial. This is possible when two or more sources of information are using the same unique identifier in their records such as social security number. When the sources of information are using the same identifier, there is obviously a possibility that his information might be combined and individual's privacy might be violated.

Therefore, the fields of personal information identifiers are often considered the most sensitive part of the information, and for this reason are protected. In other words, any information that is related to an individual can be used to track personal information by connecting different sources. The data fields can be used to link the phone number,

address or shoe size. Not all these fields are unique identifiers, but by combining several factors likely enable the identification of an individual and thus make it possible to combine information.

On privacy protection topics, it can be considered to reduce the minimum collected and stored information, and delete the information as soon as it has fulfilled its purpose. But these principles are deficient by the individuals needs to use convenient electronic services such as e-commerce transactions or e-government. The value of the service depends on whether it has a positive individual identification, and has access to relevant information about it.

If personal information is physically controlled by the individual or it is collected on third parties such as business partners or authorities, the protection will need of electronic tools to control access and the use of the information will be in accordance with individual decisions.

A broad spectrum of tools and technologies have been developed to enhance the privacy of electronic solutions, most of which focus on communications and transactions over the Internet. The purpose of Privacy Enhancing Technologies (PETs) is to protect the privacy of individuals while allowing them to interact with other parties.

There is a fine line between security and the protection of fundamental human rights. The government claims more control, and applied monitoring techniques of mass at the same time trying to establish laws that protect privacy. Every nation legislates differently data protection, the power conferred to the police and intelligence services, monitoring mechanisms, etc. In late 2001 the USA president signed a law called "USA-Patriot Act", which significantly curtails civil rights and attacks fundamental freedoms of americans under the judgment of ensuring national security. This law designed as a legal support called "war against terrorism", contains numerous provisions and amendments to laws and regulations that experts on legislative matters are considered unconstitutional. Given this legal framework, the right to privacy and freedom of expression are terms of the past. Various precursor organizations of human rights have severely criticized such laws.

The European Union regulatory approved by various member states regarding the creation and collaboration of police and intelligence services do not show a significant difference. We have countries like England where multiple devices are installed video surveillance, building on the grounds of national security. A similar case is the recently

passed Public Safety Act in Spain that at July 1, 2015, have initiated, sanctions to offenses related to protests are imposed. The law limits freedom of expression and freedom of assembly under the pretext of maintaining security.

In relation to the protection of personal data it is considered that there are two sides: the European model that seeks to protect the information and ownership of it in order to preserve the dignity and reputation of a person, even after the death. The USA model, which aims to protect the information from people based on the concept of the right to privacy, which is exempt once the person dies.

Several countries have developed laws protecting personal data and each nation has sought to adapt some of the two existing models based on their own cultural, economic and political conditions. The processing of personal data in the European Union is governed by European Directive 95/46 / EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and the free movement of such data within the European Union. The principles of this policy have been implemented in the laws of each member of the European Union.

Another issue that has been controversial is the increasing commercial applications of facial recognition. Among its uses is, to replace the connection password, locating people or in a near future, customize windows or advertisements screens when a user scroll through a store or mall. Considering that hundreds of millions of people get photos and videos daily on social networks, it is a fact that there is now a wealth of biometric data, which opens the possibility that in the future almost all people can be identified by name in public spaces. The numbers of applications are huge and still poorly understood. Defenders of human rights and privacy indicate that the biometric information is extremely sensitive, since it is possible to change a password or a credit card number; but people cannot change their fingerprints, or the patterns of their faces. There is a risk that these technologies can lead to errors as assign someone else's identity and also that these data will be used for commercial or political purposes. It is estimated that in the United States of America at the end of 2015, will be a record of more than 50 million images of faces, which is the basis of the world's largest biometric data, aimed at identifying criminals [Lyn14]. However, there are no rules in the use of facial recognition technologies; there are only two initiatives of USA states, particularly in Texas [Sta09] and Illinois [Sta08] regulating part of this phenomenon.

Another issue that is emerging is the Internet of Things [Ash09], it is a phenomenon that is having a very significant impact on the way people interact with businesses and government through the use of technology. Current technology including wireless devices, mobile phones, electronics, smart homes, all linked to the Internet, with applications where each object operates autonomously giving the user greater benefits. The Internet of Things, is also immersed in automatic making decision and in services optimization in sectors such as transport, healthcare, energy, among others. In this sense, the concerns related to privacy is based on the dangers to connect millions of devices, where details of the daily life of a person can be exposed to hack your refrigerator, or Smart-TV.

It is important to note the existence of deficiencies in relation to consumer privacy, which is why the privacy issues raised by new technologies, serve as another example of the need for privacy laws to reflect the threats of today's world.

Other legislation can be found at:

- OECD.
- In Spain:
  - Law of conservation of data on electronic communications and public communications networks.
  - Using Camcorders Act by the Security Forces in public places.
  - Act police database on identifiers obtained from DNA.
  - Draft Platform for Privacy Preferences (P3P) and Identity Management.

2.7 State of Art of Pets

2.7.1 History of PETs

The discussion of privacy was initiated with the arrival of computers in 1970, resulting in the emergence in Europe of various laws on data protection; meanwhile in the USA, the application of such laws to a number of specific sectors was limited as HIPAA (Health Insurance Portability and Accountability Act), which contains a set of privacy standards for information technologies applied to the health sector.

In 1978, Posner defined privacy in terms of discretion [Pos78a] and the following year the spread terms of insulation [Pos78b]. In 1980, Hirshleifer published an article in which

he argued that instead of being withdrawn from society, privacy was a means of social organization derived from a territorial evolutionary behavior [Hir80].

Anonymous technologies research began in the early 80's with David Chaum [Cha81] who suggested anonymous emails in order to hide the correspondence between sender and receiver through messages using public key encryption. These messages should go through a network of mixes before get its destination. The mix changes the appearance and the flow of messages through encryption techniques, which makes it difficult to relate inputs and outputs.

Shortly afterwards in the nineties, with the expansion of the Internet, the rise of the dot-coms and commercial explosion of personal information in virtual shopping, privacy became an issue of major concern. PETs research increased with the adaptation of the concept proposed by Chaum for internet data traffic [Cha81], routing ISDN [PPW91] mobile [GRS96]. With the emergence of several research projects with public resources, several companies adopted the protection of privacy in their business model [Lac00] [FID14] [PRI07].

The technology can be designed to keep personal data under users' control. It would be desirable that user could expose the minimum amount of information to third parties. Anonymity technologies serve as tools for the protection of privacy in electronic applications, and are the main component of PETs. Anonymous communication networks protect Internet users' privacy. It has been an achievement to keep hidden link between sender and receiver. For applications such as electronic voting or electronic payments, both anonymity and privacy are strictly necessary.

The European Commission define Privacy Enhancing Technologies [Com07] as "The use of PETs can help to design information and communication systems and services in a way that minimizes the collection and use of personal data and facilitates compliance with data protection rules. The use of PETs should result in making breaches of certain data protection rules more difficult and / or helping to detect them".

There is no widely accepted definition of the term PETs nor does there a distinguished classification exist. Literature about categorized PETs according to their main functions, privacy management and privacy protection tools [Fri07][M05][Ada06].

In general PETs are observed as technologies that focus on:

- Reducing the risk of breaking privacy principles and legal compliance.
- Minimizing the amount of data held about individuals.
- Allowing individuals to maintain control of their information at all times.

Several researchers are centered on protection of privacy and personal data through sophisticated cryptology techniques. PET's applications such as individual digital safes or virtual identity managers have been proposed for trusted computing platforms.

PETs have traditionally been restricted to provide "pseudonymisation" [PH08]. In contrast to fully anonymized data, pseudonymisation allows future or additional data to be linked to the current data. These kind of tools are software that allow individuals to deny their true identity from those operating electronic systems or providing services through them, and only disclose it when absolutely necessary.

Examples include: anonymous web browsers, email services and digital cash. In order to give a better explanation about PETs applied in a data center, consider the Solove's Taxonomy [Sol06] used to categorize the variety of activities to infringe privacy. We refer to [PH08] for further definitions of privacy properties in anonymous communication scenarios.

- **Information Collection**: Surveillance, Interrogation.
- **Information Processing**: Aggregation, Identification, Insecurity, Secondary Use, Exclusion.
- **Information Dissemination**: Breach of Confidentiality, Disclosure, Exposure, Increased Accessibility, Blackmail, Appropriation, Distortion.
- **Invasion**: Intrusion, Decisional Interference.

Collecting information can be a damaging activity, not all the information is sensitive but certain kinds definitely are. All this information is manipulated, used, combined and stored. These activities are labeled as Information Processing. When the information is released, this group of activities is called Information dissemination. Finally, the last group of activities is Invasion that includes direct violations of individuals. Data brokers are companies that collect information, including personal information about consumers, from an extensive range of sources for the purpose of reselling such information to their

customers, which include private and public sector entities. Data brokers activities can fit in all of the categories above.

In other sub-disciplines of computer science, privacy has also been the focus of research, concerned mainly with how the privacy solutions are to be applied in specific contexts. In simple terms, they are concerned with defining the process of when and how to apply privacy solutions. Before choosing a technology for privacy protection, several questions have to be answered because there is no certainty that one type of technology solves one specific problem. One of the questions to consider is who defines what privacy is? (The technology designer, the organization's guidelines, or the users) [DG12].

2.7.2 Triggers of PETs

There are three main initiators of the vast number of changes affecting the notions, perceptions and expectations of privacy [oPitIAC07].

1. **Changes in technology**: There are huge differences in the technological environment that currently exist, compared to some decades ago. Physical devices behind the information technologies have potentially increased; improvement in processing speed, storage capacity on hard disks and bandwidth allow data to be collected, stored and analyzed in a previously unimaginable way. Other new technologies are radio frequency chips identification implanted in humans. The presence of the virtual world that is supplied with each event daily. The development of new algorithms for data mining. The low cost of technological devices has allowed that tools for personal information collection and analysis from different sources are easily available to people, businesses and governments.

2. **Social changes**: The evolutionary changes in the activities and practices that make use of the above described technological systems and the transformation in the way we do things in our daily lives. For example, it has been essential and has set a unprecedented event in social participation, to give access of community personal information to institutions and organizations. This demand for information has emerged incrementally to manage or confer benefits on various vulnerable population groups, such as providing services and specific support to unemployed, low-income earners, elderly people.

3. **Discontinuity circumstances**: Events and emerging issues that profoundly transform the national debate on privacy in a very short period of time. It breaks the natural scheme, because currently it is not allowed a gradual adjustment to adapt to a new set of circumstances. Among the most recent examples are the terrorist events in France, Copenhagen or Tunisia; and, the exodus of thousands of Syrian refugees in Western countries, which have transformed the international environment have led to the launch of anti-terrorism and national security strategies.

2.7.3 PETs Categorization

A summary of the possible technologies that can be used to enhance privacy is included in Table 2.2 and it is divided into four categories which indicate the effectiveness and availability, protection of personal data [KvGtH+04]. The items included in the General category have less effectiveness in the protection of personal data, while the Privacy Management Systems provide the most effective protection to involve more complex techniques.

As shown in [Dó5], we can distinguish several subtopics in the field of Privacy-enhancing technologies.

- **Anonymous Communication**: These technologies include anonymous communication networks of general purpose such as Tor [DMS04] Routing layered (onion routing) [STRL01] [GRS96]; Mixes [PPW91] ISDN; anonymous email as Chaum's original proposal [Cha81], Babel [GT96] or Mixminion [DDM03]; anonymous P2P systems such as Tarzan [FM02], MorphMix [RP04], P5 [SBS02], Onion [Bro02] or Herbivore [GRPS03]; DC-nets [Cha98]; and proposals that improve the resistance attacks by anonymous communication systems [CYS+07], other applications are related to accountable anonymity [TAKS07] [TAKS08].

- **Publishing Limited**: The anti-censorship systems aim to provide the ability to post anonymously, so that information cannot be removed. The most significant proposals of anonymous posting systems are: Eternity Service [And96] [Ben01]; TAZ servers [GW98]; The Free Haven Project [DFM00]; Freenet [CSWH01]; Publius [WRC00]; and Tangler [WM01], among others.

Table 2.2: Privacy enhancing technologies for effectiveness

Dealer	**Technogies**
General	- Encryption (storage and communication - logical access controls (authentication and authorization) - Biometrics - technologies that improve the quality - Cryptography based on identifiers - Easy access to government services
Data separation	- Managing profiles - Privacy incorporated into databases - Electronic signature blind - Secure personal data
Anonymization	- Mix Routers - Routers layered (onion routing) - Management Tools cookies - File Management Tools - Smart Cards - Biometrics
Management Privacy Systems	- P3P (Draft Platform for Privacy Preferences) - Privacy Rights Management (based on digital rights management) - Automatic Data Destruction Administration (retention) - PISA (Agent software with built-in privacy) - Privacy Ontology - EPAL (Enterprise Privacy Authorization Language) - Management Software Privacy Policy

- **Censorship resistance communications**: The proposals in this category are: Infranet [FBH+02], anonymous web browsing [KH04], among others. A summary of the history and current phishing attacks is shown in [Oll07]. In [JBBM06] several privacy attacks web browsers are discussed.

- **Electronic money/anonymous credentials**: Electronic money is associated with anonymous digital credential. Anonymous digital credential proves something about its owner, without revealing his identity. Both share common characteristics. Examples of these applications are Idemix [CVH02], Credentica [Inc07] Identity Metasystem Architecture [CJ07], and Generic Bootstrapping Architecture (GBA) [SM08]. On the other hand, it have also conducted researches that protect and manage the identity of users in different areas such as mobile domains [ABD+03], negotiation processes online [SWY+03].

- **Private information recovery**: It is a set of protocols that allow the user to retrieve an item from a database hosted on a server, without revealing which item

is trying to recover [SP07] [Gol07a].

- **Traffic Analysis** In recent years a number of publications have focused on traffic analysis attacks that can be used against anonymous communications systems such as: Sybil attacks [Dou02], fingerprints [Hin02], etc. In order to measure the success of such attacks, anonymity metrics are used.

- **Demonstrable Permutation**: For critical applications such as electronic voting, it is important to prove that the permutation of inputs and outputs have been executed correctly and safely. There are some suggestions that can be found as demonstrable permutations applications, such as Flash mixing [Jak99], mixing universally verifiable [Abe06] and hybrid mixes.

- **Economy**: Over time it has been increased the incentives in the implementation of PETs. The research that has been done focuses on the understanding of the economic aspects of anonymity and in the design of systems that encourage honest reputation user behavior [ADS03] [DA04].

- **Formal Methods**: There have been attempts to formalize the properties of anonymity, and the establishment of a formal framework for analyzing properties of hidden information [HS04].

- **Pseudonymity**: Pseudonyms are used in systems where users need to maintain a permanent identity. Pseudonyms have been proposed for a wide variety of systems, such as email or communications infrastructure [SCM05].

The Table 2.3 shows the categorization into two groups of privacy enhancing technologies: the protection of privacy and privacy management. The first group includes a list of "opacity" tools, while the second has tools of "transparency".

The PETs are correlated with various disciplines, because were developed in context of information systems which is governed by social requirements and business models. The development of specific applications in online scenarios can be identified as telecommunications, PETs, and Economics. These applications are influenced by: (i) laws and legal regulations; (ii) the market context related to its needs and products; (iii) and requirements of users from various disciplines respectively.

Table 2.3: PETs categorized by its function

Category	Subcategory	Description
Privacy protection	Pseudonymity tools	Allow e-business transactions without requiring private information.
	Products and services for anonymization	Provide navigation capability and sending emails without disclosing the address and identity of the user.
	Cyphering tools	Secure e-mail, documents and transactions to be read by third parties.
	Filters and blockers	Prevent email and unwanted web content.
	Remover evidence and monitoring	Remove the electronic monitoring user activity.
Privacy management	Information tools	Creation and verification of privacy policies.
	Administrative tools	Managing user permissions and identity.

2.8 Mixes

2.8.1 Classification of Anonymous Communication Systems

In literature it has been classified anonymous communication systems into two categories: systems of high latency and low latency. The first aim to provide a strong level of anonymity but are applicable to systems with limited activity that does not require quick attention as email. On the other hand, low-latency systems offer better performance and are used in real-time systems, such as web applications, instant messaging, among others. Both types of systems are based on the proposal of Chaum [Cha81], who introduced the concept of mix.

The purpose of a network of mixes is to hide the correspondence between inputs with outputs; it means to hide who communicates with whom. A network of mixes collects a number of different packages called the anonymity set, and through cryptographic operations changes the appearance of the incoming packets, making it difficult for the attacker to know who communicate. The mixes are based building block for all communication systems high latency.

Low latency anonymous systems are also classified according to their routing paradigms: those derived from the Onion Routing [GRS96] and those based on Crowds [RR98]. In the first subcategory are systems like Tor [DMS04], JAP [Fou15], and I2P [I2P15] which use deterministic routing, where the set of proxies utilized for traffic transmission is considered known. Meanwhile, systems as GNUNet [BG15], BitBlender [BMGS08] and OneSwarm [IPKA10] schemes use probabilistic routing traffic as does Crowds. That is, while Tor uses a circuit of 3 nodes, randomly Crowds decides how

many will use for routing nodes, may be 2, 3, or 14. Another difference is that with Crowds each pair of nodes has a different symmetric key, which is used to encrypt data between nodes. A key path is established for each route to encrypt both the request and the response.

Currently the most widely used anonymous network communication is Tor (acronym for The Onion Router), which allows to browse anonymously on the web. The Tor network is a network of distributed communications and superimposed on Internet, which the routing of messages exchanged between users does not reveal his identity. Tor messages travel from source to destination through a series of special routers called "onion routers". Development and maintenance is made possible by a group of organizations and individuals who donate their bandwidth and processing capability as well as an important community that supports it [DMS04]. In [Loe09] is shown a comparison of the performance of communication systems of low and high latency.

2.8.2 Mix Networks

In 1981, Chaum [Cha81] introduced the concept of Mix networks whose purpose is to hide the correspondences between the items in its input and those in its output. Since then, there are different proposals in the literature [DP04]. A mix network collects a number of packets from distinct users called anonymity set, and then it changes the incoming packets appearance through cryptographic operations. This makes it impossible to link inputs and outputs taking into account timing information. Anonymity properties are strongest as well as the anonymity set is bigger, and these are based on uniform distribution of the actions execution of the set subjects. In Figure 2.5 is represented a model of a mix network.

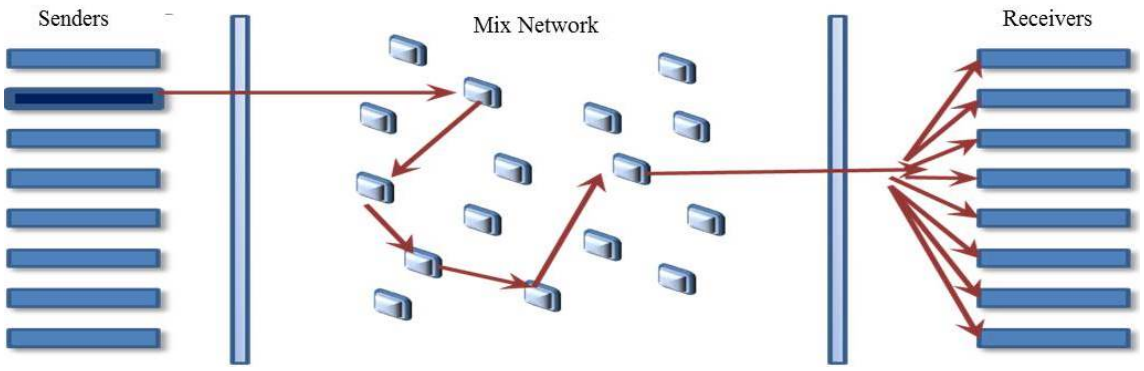


Figure 2.5: Mix network model

A mix is a go-between relay agent that hides a message's appearance, including its bit pattern and length. For example, say Alice generates a message to Bob with a constant length, a sender protocol executes several cryptographic operations through Bob and mix public keys. After that, a mix hides the message's appearance by decoding it with the mix private key.

The initial process for Alice to be able to send a message to Bob using a Mix system is to prepare the message. The first phase is to choose the path of the message transmission; this path must have a specific order for iteratively sending before the message gets its final destination. It is recommended to use more than one mix in every path to improve the security of the system. The next phase is to use the public keys of the chosen mixes for encrypting the message, in the inverse order that they were chosen. In other words, the public key of the last mix initially encrypts the message, then the next one before the last one and finally the public key of the first mix will be used. Every time that the message is encrypted, a layer is built and the next node address is included. This way when the first mix gets a message prepared, this will be decrypted with his correspondent private key and will get the next node address. Figure 2.6 shows the network mix procedures to deliver a message.

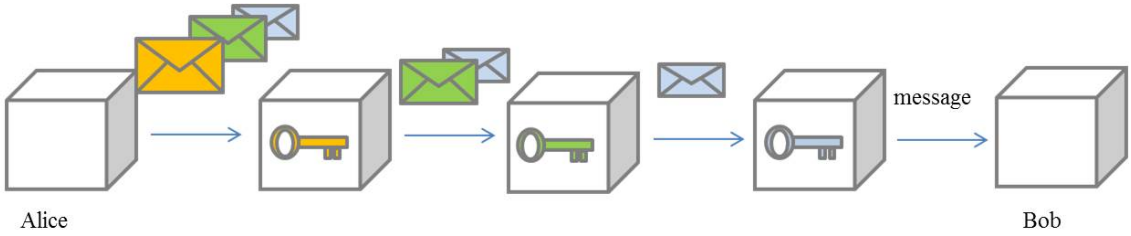


Figure 2.6: Mix networks process

Techniques such as reordering and delay messages, and utilize "dummy traffic" are applied to modify the flow of messages. This first design was a threshold mix, which has two phases: the first phase collects B messages; in phase 2 mixture messages and then delivers them. The graphical representation is shown in Figure 2.7.

Even if an attacker can observe all input and output lines, the goal is that he cannot obtain a correlation between them and being impossible to assume the correspondence between a message sender and its receiver. To achieve this purpose different solutions have been proposed:

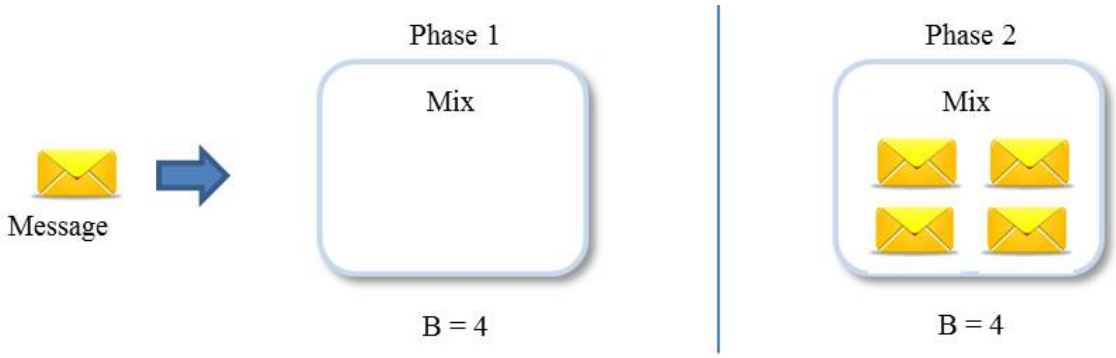


Figure 2.7: Mix model phases

- **Continuous mixes**: Each package suffers a delay before it is sent, in an example proposed by Kesdogan et al. [KEB98], the delay is previously assigned to the sender with an exponential distribution.

- **Pool mixes**: In this case there is a buffer storing the packets received for the mix. If a certain condition is accomplished, which may be temporary and/or after a number of packets are stored, one packet is chosen randomly and sent.

These strategies can be used together, providing higher levels of anonymity, although it is present the dichotomy between the robustness against anonymity attacks and the network latency.

The main drawbacks of the basic mix scheme are:

1. The encryption and decryption with public key on each mix is computationally expensive in terms of time.

2. They have high degrees of latency, so apply well in emails but not in anonymous web browsing.

3. External attacks are executed outside the network, while internal attacks are from compromised nodes, which are actually part of the network. Mix networks are a powerful tool to mitigate outside attacks by making the sender and receiver path untraceable. The participant nodes in a mix network relay and delay messages in order to hide the route of the individual messages through the mix. However, they can be corrupted nodes that perform inside attacks. This kind of problem is addressed [RR98] by hiding the sender or the receiver from the relay nodes.

2.9 Summary

This chapter has mentioned the state of art of PETs (Privacy Enhancing Technologies), as well as privacy and anonymity definitions. It has been identified the anonymity degrees in order to establish a scale of systems. In this sense, it has been shown the taxonomy to identify privacy violations, privacy properties, anonymity systems measures and legislation related. It has been listed the evolution over time of PETs. On the other hand, it has also been shown the categorization of PETs by subtopics; and well as it has been classified the tools according to privacy protection and privacy management. One of the basis of anonymous communications are mixes networks, it has been listed its definition and the classification of anonymous communication system.

Chapter 3

Traffic Analysis on Anonymous Communications

The purpose of this chapter is to give a general approach of Traffic Analysis on Anonymous Communications. The chapter begins with an introduction of the concept of anonymous communication. Then, it has been demonstrated the Dining Cryptographers problem in order to explain the base of hiding the executer identity. It is mention the types of attacks, particularly is defined the Intersection attacks. The chapter continues with a summary of the work related to statistical disclosure attacks found in the literature. Finally in the last section we collect a brief summary of the discussion in the chapter.

3.1 Introduction

Anonymous communications aim to hide the relationship in communication. The anonymity is the state or quality of being unknown to most people; anonymous communications can be achieved by removing all identifiable characteristics of an anonymous network.

Consider a system where a set of actors is concentrated in a communication network, such as clients, servers and nodes. These actors exchange messages through public communication channels. In [PH08] defined anonymity as the state of being unidentifiable within a set of subjects, known as the "anonymous set". One of the main features of anonymous set is its variation over time. The probability that an attacker can effectively reveal who is the recipient of a message is exactly $\frac{1}{n}$, where n correspond to the number

of members in the anonymous set. Research in this area focuses on developing, analyzing and carry out attacks on anonymous communication networks.

Internet infrastructure was initially planned to be an anonymous channel, but now it is well known that anyone can spy a network. Attacks on communication networks are a serious problem in any organization. New technologies have a great challenge to find solutions to improve security. It has been proven that the data encryption and the topology of a network do not provide enough protection for users' privacy even when mechanisms of anonymity are applied. The attacker may be able to diminish its anonymity properties using auxiliary information.

Attackers have different profiles such as area of action, range of users, heterogeneity, distribution and location; they can be internal or external, passive or active, local or global, static or adaptive. Some examples are:

- **Passive global attacker**: observes all links.
- **Rogue operator**: controls one or more significant nodes in the network.
- **External attacker**: capable of add or modify traffic.

An external attacker can identify traffic patterns to deduce who communicate, when and how often. There is no guarantee of complete privacy due the number of potential senders and existing receivers. If there are only two users on the network, an attacker to gain access to such information can determine who communicates with whom.

In traffic analysis the information can be inferred from observable characteristics of the data flowing through the network such as the packet size, its origin and destination, size, frequency, timing, among others.

In anonymous communications context, it is important to point out public key cryptography also known as asymmetric cryptography, which is the cryptographic method that uses a pair of keys for sending messages. The two keys belong to the same person who sent the message. One key is public and can be delivered to anyone, the other key is private and the owner must protect it from third parties. It is assumed that two people have not the same key pair. If the sender uses the recipient's public key to encrypt the message, once encrypted, only the recipient's private key can decrypt this message, because it is the only one who knows it. Therefore confidentiality is achieved by sending the message. In Figure 3.1 diagram shows asymmetric cryptography.

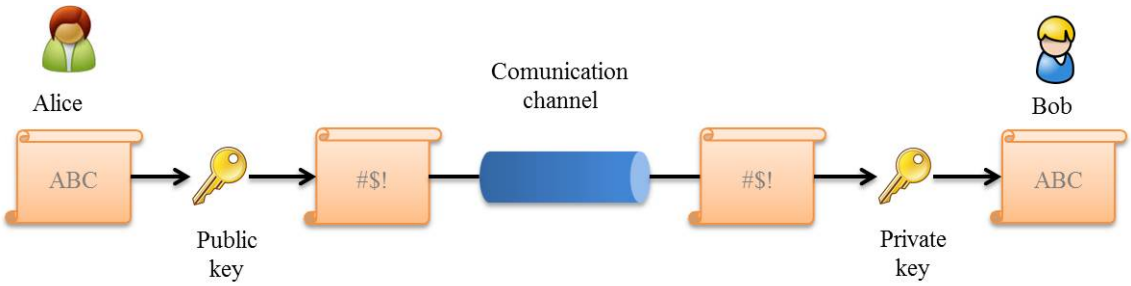


Figure 3.1: Asymmetric cryptography

3.2 Dining Cryptographers

Dining Cryptographers [Cha98] is a protocol to come up from the necessity to ensure the anonymity of a set of users, regardless of the attackers' time and resources.

Dining cryptographers can be described as follows: "Three cryptographers are dining in a restaurant. When it is time to pay, the waiter tells them that the account has already been paid and who did it do not want his identity revealed. Cryptographers want to know if any of the guests was the one who made the payment, or if someone outside was who paid. They only want to know if any of them paid or not. In case of an external paying, anonymity is guaranteed, but if the person that paid was a member of the group, others respect the right to invite and do not want to know the identity of who paid".

The found solution is: "Each one of the diners throws a coin. Observe the result and shares with its neighbor to the left. Then each of them looks exactly two coins, self and neighbor who shares with him. Finally, everyone must indicate whether the two coins that could be observed are the same or different, with the condition that if one of them paid the bill, should lie about their claim".

Under these conditions, if the number of responses that were different coins is odd, it means who paid is in the group of diners. The opposite, indicates that the person who paid is someone outside the group.

In a protocol based on the above, the way it transmits the encrypted message is that the sender sends the message and all other users of the set of senders also transmit some information. A receiver can obtain the encrypted message through the sum of all received messages that were transmitted. This transmission method is used to hide the sender who has sent a message. Its main drawbacks are: when an attacker delete or add messages to the communication channel; when two or more users send messages at once, they cannot

be received correctly; when the number of secret keys is too large to be shared between each pair of users.

3.3 Types of Attacks

Traffic analysis belongs to a family of techniques used to deduce information patterns of a communication system. It has been shown that the encrypted data itself do not guarantee the anonymity [Dan03].

- **Brute Force Attacks**: Dummy type messages are sent to the network in order to confuse the attacker. The pattern of attack is to follow all the possible routes that a message can take. The attack is called brute force because all it does is looking for all possible combinations of routes; it is assumed that at some point, if the mix is not well designed the attacker can obtain the relationship between sender and receiver.

- **Flushing attacks node**: If nodes expect to have t messages before they are delivered, the attacker can send $t - 1$ messages to identify their own messages that have come and gone. He can correlate with corresponding output message by identifying the message he has not sent.

- **Attacks time**: If an attacker knows the time it takes to make each of the routes, he can determine what path a message takes. The attack can be effective even when the flow of messages from participants and the first node is constant. In order to prevent this type of attack, the system mixes must wait a random time prior to sending messages. However, for practical purposes it is not so convenient that the wait is too long.

- **Contextual Attacks**: They are considered the most dangerous but also the most difficult to model because the user behavior in real life is variable. In this family there are attacks called "attacks on communication patterns" in which if one user is active at a time is possible to determine who communicates with whom. Other attack in this area is the "Counting packet attack" where the attacker is able to distinguish some types of communication, for example, when the attacker knows which users have made more deliveries. The patterns communication and counting packet attacks can be combined to create the "frequency messages attack". The

three previous attacks may not work well in large networks. Another contextual attack is the "intersection attack" where the attacker knows which users are active at a given time and therefore can infer who communicate.

3.4 Intersection Attacks

Statistical disclosure attacks are also known as attacks intersection, its base lies in an attack on a simple mix network, which aims to obtain information from a particular sender by linking senders messages sent, recipients with the messages they receive, or link senders and receivers [Ray01a]. An attacker can derive these relationships through the observation network, delaying or modifying the mix messages to compromise systems.

Commonly attacks on anonymous communication systems are two-step processes. The first step is to find the anonymity set, which is the set of users whose identity has been hidden through the use of a scheme of anonymity. It is assumed an attacker can observe the anonymous set but cannot determine their relationships. The second step of the process is to find the source of anonymous traffic through the observation of network connections, for example, monitoring the routers used as anonymous set and where traffic is transported. This is a general outline of how traffic analysis attacks operate.

3.4.1 The Disclosure Attack

The disclosure attack is presented in [AK03], whose aim is to get information of one particular sender, Alice. The attack is global, in the sense that it retrieves information about the number of messages sent by Alice and received by other users, and passive, in the sense that the attacker cannot alter the network (sending false messages or delaying existent ones).

The data observed are structured in rounds, which can be defined for batches or lapses of time. A round is composed of multisets of senders and receivers, where a_k is the number of sent messages of Alice in round k and b_k is the number of messages sent by all senders in that round. In Figure 3.2 it is shown the messages flow through the mix in round k.

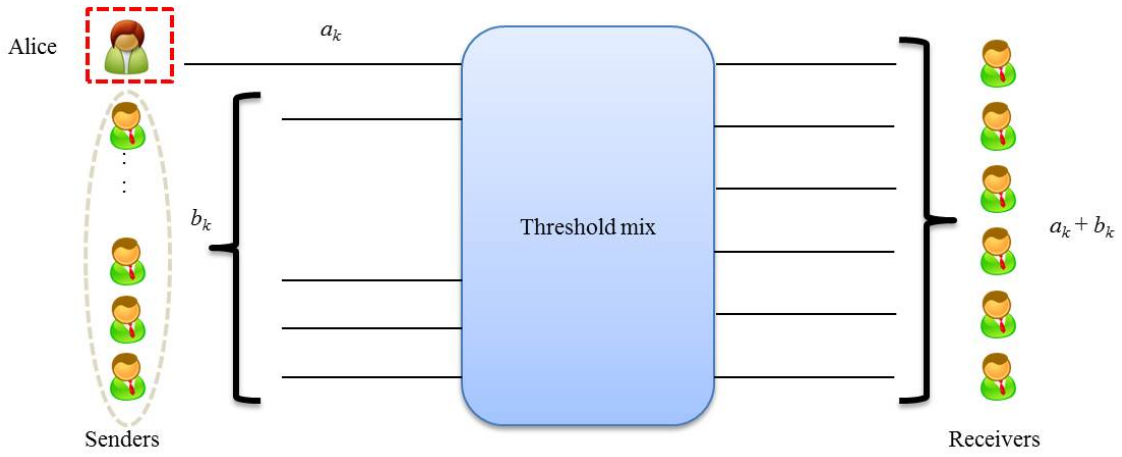


Figure 3.2: Representation of a round with threshold mix

Authors assume Alice sends messages to limited m recipients or friends with the same probability and that she sends one message in a batch of b messages. All users send messages using a simple threshold mix. This kind of mix collects a specific number of messages per round and delivery them in a random order, it is represented in Figure 3.2.

The attack was modeled by considering a bipartite graph $G = (A \cup B, E)$. The set of edges E represents the relationship between senders and recipients A and B. Mixes assume that all networks links are observable. So, the attacker can determine anonymity sets by observing messages to and from an anonymity network; the problem arises for how long the observation is necessary. In this attack, authors make several strategies in order to estimate the average number of observations for achieve the disclosure attack. They assume that: i) Alice participates in all batches; ii) only one of Alice's peer partners is in the recipient sets of all batches.

A disclosure attack has a learning phase and an excluding phase. The attacker should find m disjoint recipients set by observing Alice's incoming and outgoing messages. This attack is very expensive because it takes an exponential time taking into account the number of messages to be analyzed trying to identify mutually disjoint sets of recipients. The main bottleneck for the attacker derives from an NP-complete problem. Test and simulations showed it only works well in very small networks.

3.4.2 The Statistical Disclosure Attack

The SDA proposed by Danezis [Dan03] is based on the previous attack. In this work exists the same assumptions that [AK03]. The attack is developed taking into account a

wide number of observations in a period of time on a mix network where it is possible to calculate the distribution probability of sending / receiving messages in order to diminish privacy in an anonymous communication system.

Figure 3.3 shows the messages distribution in a network of mixes where it is possible to observe that there is higher probability message sender is user A o B in this round. After SDA there were proposed more traffic analysis attacks in order to deduce information considering behavioral patterns in a communication system.

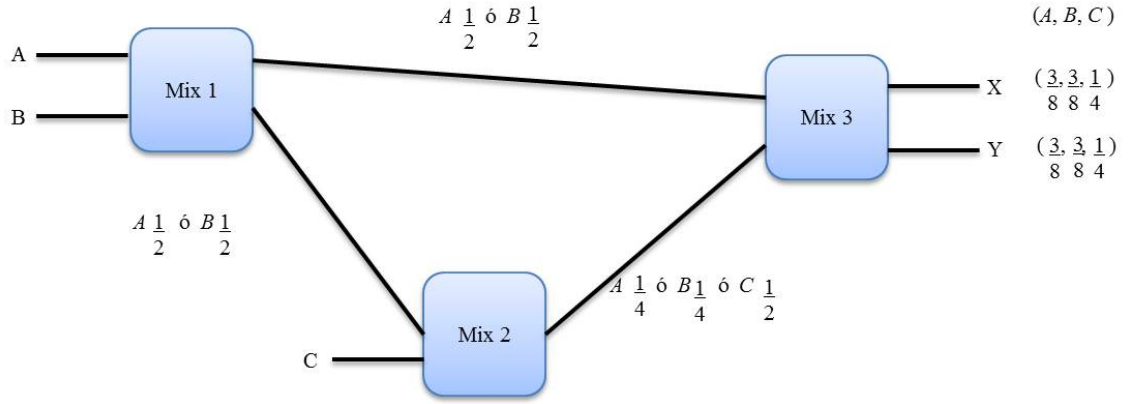


Figure 3.3: Distribution probabilities of sending/receiving messages

The method tries to reveal the most likely set of Alice's friends using statistical operations and approximations. It means that the attacks applies statistical properties on the observations and recognize potential recipients, but it does not solve the NP-complete problem presented in previous attack. It requires less computational effort by the attacker and gets the same results.

Intersection attacks are designed based on time correlation where senders and receivers are active. The attacker can create a set of most likely Alice receivers through observing the elements that receive packets where Alice is sending a message. Consider $\overrightarrow{v}$ as the vector with N elements corresponding to each potential recipient of the messages in the system. Assume Alice has m recipients as the attack above, so $1 - m$ might receive messages by her and it's always $|\overrightarrow{v}| = 1$. The author also defines $\overrightarrow{u}$ as the uniform distribution over all potential recipients N. In each round the probability distribution is calculated, so recipients are ordered according to its probability. The information provided to the attacker is a series of vectors representing the anonymity sets. The highest probability elements will be the most likely recipients of Alice.

Variance on the signal and the noise introduced by other senders is used in order to calculate how many observations are necessary. Alice must demonstrate consistent behaviour patterns in the long term to obtain good results, but this attack can be generalized and applied against other anonymous communication network systems. Distinct to the predecessor attack, SDA only show likely recipients and does not identify Alice's recipients with certainty.

3.4.3 Further SDA Attacks

One of the main characteristics on Intersection Attacks relies on a fairly consistent sending pattern or a specific behaviour for users in an anonymity network; which is not like that on real communication systems. Several works assume Alice has exactly m receivers and sends messages with the equal probability each, or it is also just focus on one user solution as interdependent.

Most models of attacks has focused on systems consider only a threshold mix and just the SDA has been extended to pool mixes where messages can be delayed for delivery by more than one round. For attacks evaluation have been considered two perspectives: 1) Evaluate the revelation from individual posts [DT09] [TGPV08]; 2) Evaluate from the number of rounds necessary to identify a percentage of all Alice receivers [MD05] [MW11a] [PWK10].

Each of the attacks intended to deduce the most likely contacts of Alice, also known as her user profile. All through observing the set of possible recipients of each message that Alice sends; invariably it is used "Alice receivers" and "friends of Alice" to refer to all people with whom Alice communicates. The variants related to the attacks of revelation are the technique used to deduce users' profiles.

Attacks using threshold mix or mixes pool are described in [MD05], it consider the assumption previously assumption. For example, they focus on a single user Alice, and it is assumed that the number of Alice friends and the parameter B corresponding to the threshold are known. One of the main features of this type of attack is that it believes there are constant patterns of sending or a specific behavior of the anonymous network users.

Mathewson and Dingledine in [MD05] make an extension of the original SDA. One of the more significant differences is that they consider that a real social network has a scale-free network behaviour, and also such behaviour changes slowly over time. They do not simulate these kinds of attacks. In order to model the sender behaviour, authors assume Alice sends n messages with a probability $P_m(n)$; and the probability of Alice sending to each recipient is represented in a vector $\overrightarrow{v}$. First the attacker gets a vector $\overrightarrow{u}$ whose elements are $\frac{1}{b}$ the recipients that have received a message in the batch, and 0 for recipients that have not. For each round i in which Alice sent a message, the attacker observes the number of messages m_i sent by Alice and calculates the arithmetic mean. Simulations on pool mixes are presented, taking into account that each mix retains the messages in its pool with the same probability every round. The results show that increasing variability in the message makes the attack slower by increasing the number of output messages. Assuming all senders choose with the same probability all mixes as entry and exit points and attacker is a partial observer of the mixes. The results suggest that the attacker can succeed on a long-term intersection attack even when it partially observes the network. When most of the network is observed the attack can be made, and if more of the network is hidden then the attacker will have fewer possibilities to succeed.

As mentioned before, the attacker is able to view messages in and out of the mix. If we represent the communication that occurs between users in a certain time period (or the threshold where the B parameter mix is determined), through a matrix, the messages will be marginal.

The Figure 3.4 shows three rounds, the observations obtained by the attacker to the network. The system consists of 4 users and $B = 10$. In round 1: user 1 has sent 4 messages and has received 3; user 2 has sent 3 messages but has not received any message; user 3 has submitted 2 and received 4; finally, the user 4 has not sent or received any messages. In this example as $B = 10$, one round is composed of 10 messages. For teaching purposes are marked with a red box those lines / columns where have no sent / received messages. It may be occur that a user in a round is not participating as sender or receiver or it has not been active, in this case it will be marginal zero.

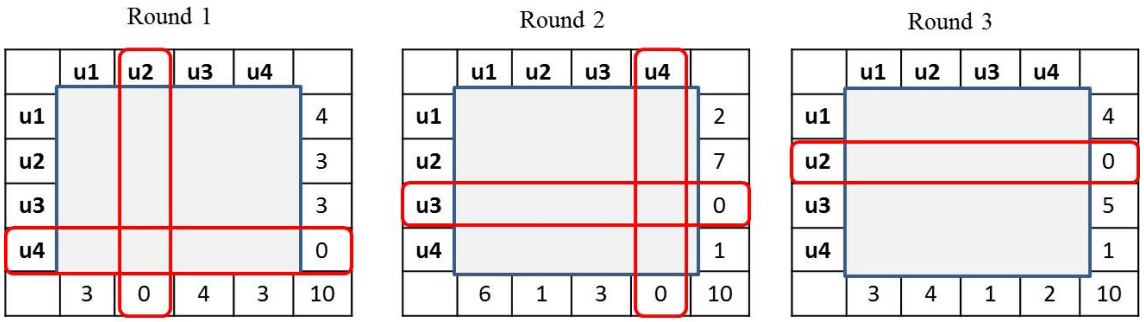


Figure 3.4: Example of three rounds

The Two Sided Statistical Disclosure Attack (TS-SDA) [DDT07], provide an abstract model of an anonymity system considering that users send messages to his contacts, and takes into account some messages sent by a particular user are replies. This attack assumes a more realistic scenario regarding the user behaviour on an email system; its aim is to estimate the distribution of contacts of Alice, and to deduce the receivers of all the messages sent by her. The model considers N as the number of users in the system that send and receive messages through a threshold mix of size B.

Figure 3.5 shows the attack model. Each user n has a probability distribution D_n of sending a message to other users. For example, the target user Alice has a distribution D_A of sending messages to a subset of her k contacts. At first the target of the attack, Alice, is the only user that will be model as replying to messages with a probability r. The reply delay is the time between a message being received and sent again. The probability of a reply r and the reply delay rate are assumed to be known for the attacker, just as N and the probability that Alice initiates messages. Based on this information the attacker estimates: (i) the expected number of replies for a unit of time; (ii) The expected volume of discussion initiations for each unit of time; (iii) The expected volume of replies of a particular message. Authors show a comparative performance of the Statistical Disclosure Attack (SDA) and the Two Sided Disclosure Attack (TS-SDA). It shows that TS-SDA obtains better results than SDA. The main advantage of the TS-SDA is its ability to uncover the recipient of replies or reveal discussion initiations. Inconvenient details for application on real data are the assumption that all users have the same number of friends to which they send messages with uniform probability.

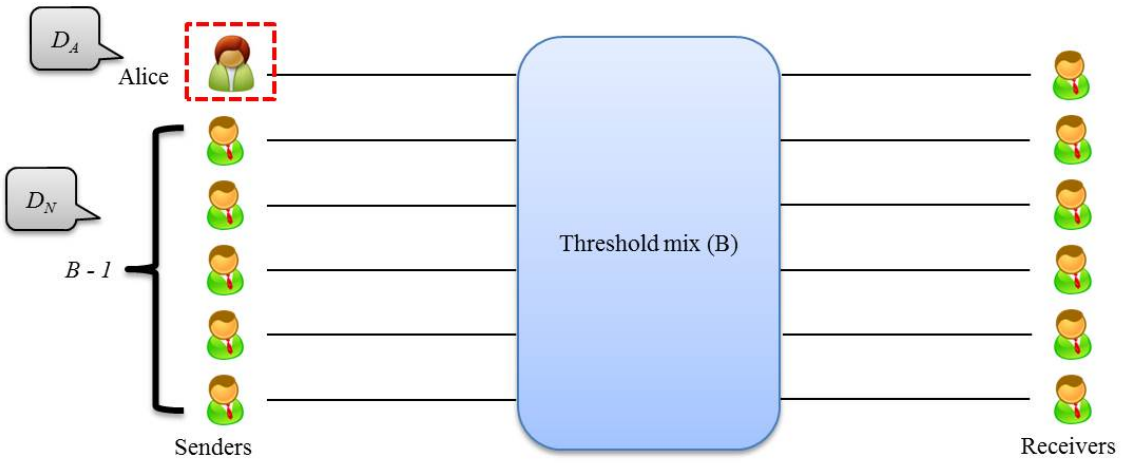


Figure 3.5: TS-SDA model

The PMDA [TGPV08] is based on graph theory, it considers all users in a round at once, instead of one particular user iteratively. No assumption on the users behaviour is required to reveal relationships between them. Comparing with previous attacks where Alice sends exactly one message per round, this model permits users to send or receive more than one message in each round. Bipartite graphs are employed to model a threshold mix, and through this, they show how weighted bipartite graphs can be used to disclosure users communication. A bipartite graph $G = (S \cup R, E)$ considers nodes divided in two distinct sets S (senders) and R (receivers) so that every edge E links one member in S and one member in R. It is required that every node is incident to exactly one edge. In order to build a threshold mix, it is thought that t messages sent during one round of the mix form the set S, and each node $s \in S$ is labeled with the sender's identity $\text{sin}(s)$. Equally, the t messages received during one round form the set R where each node r is labeled with the receiver's identity $\text{rec}(r)$. A perfect matching M on G links all t sent and received messages. Additionally P' is $t \times t$ matrix containing weights w_s, r, representing probabilities for all possible edges in G. The procedure for one round is: (i) sent messages are nodded in S, and marked with their senders identities; (ii) received messages are nodes in R, and marked with their receivers identities; (iii) derive the $t \times t$ matrix: first estimating user profiles when SDA and then de-anonymize mixing round with $P'(s,r) = \tilde{P}_{\text{sin}(S)}, \text{SDA}\left(\text{rec}\left(r\right)\right), s \in S_i \tilde{P}_{\text{sin}}$, r; iv) replace each element of the matrix $P'(s,r)$ with $\log_{10}\left(P'\left(s,r\right)\right)$; v) having each edge associated with a log-probability, a maximum weighted bipartite matching on the graph $G = (S \cup R, E)$ outputs the most likely sender-receiver combination. This work shows that it is not enough to take the

perspective of just one user of the system. Results of experimentation show that this attack does not consider the possibility that users send messages with different frequencies. An extension proposal considers a Normalized SDA. Another related work concerning perfect matchings is perfect matching preclusion [BHVY05] [PS09] where Hamiltonian cycles on the hypercube are used.

A generalization of the disclosure attack model of an anonymity system applying Bayesian techniques is introduced by Danezis et al. [DT09]]. Authors build a model called Vida to represent long term attacks against anonymity systems, which are represented as N_{user} users that send N_{msg} messages to each other. Assume each user has a sending profile, sampled when a message is to be sent to determine the most likely receiver. The main contributions are two models: (1) Vida Black-box model represents long term attacks against any anonymity systems; (2) Vida Red-Blue allows an adversary to performance inference on selected target through traffic analysis. Vida Black Box model describes how messages are generated and sent in the anonymity system. In order to perform inference on the unknown entities they use Bayesian methods. The anonymity system is represented by a bipartite graph linking input messages i_x with its correspondent output messages o_y without taking into account their identities. The edges are labelled with its weight that is the probability of the input message being sent out. Senders are associated with multinomial profiles, which are used to choose their correspondent receivers. Through Dirichlet distribution these profiles are sampled. Applying the proposed algorithm will derive a set of samples that will be used for attackers to estimate the marginal distributions linking senders with their respective receivers. Vida Red-Blue model tries to respond to the needs of a real-world adversary, considering that he is interested in particular target senders and receivers. The adversary chooses Bob as a target receiver, it will be called "Red" and all other receivers will be tagged as "Blue". The bipartite graph is divided into two sub-graphs: one containing all edges ending on the Red target and one containing all edges ending on a Blue receiver. Techniques Bayesian are used to select the candidate sender of each Red message: the sender with the highest a-posterior probability is chosen as the best candidate. The evaluation includes a very specific scenario which considers: (i) messages sent by up to 1000 senders to up to 1000 receivers; (ii) each sender is assigned 5 contacts randomly; (iii) everyone sends messages with the same probability; (iv) messages are anonymized using a threshold mix with a batch of 100 messages.

In [KP04] a new probabilistic method of attack called The Hitting Set Attack is presented. The frequency analysis is used to improve the applicability of the attack, and verification of algorithms is also used to solve the problem of improving the solution space. It is assumed that a subset A' of all senders A send a message to a subset B' of all receivers B. In this model, the attacker can determine the anonymous set, considering for example that mixes assume that all links network are observable. This can be assumed in a real-world scenario, if an attacker is able to observe messages to and from an anonymity system. The following properties of an anonymous system is assumed: i) Each anonymous communication, a subset A' of all senders A sends a message to a subset B' of all receivers B; ii) A sender can send multiple packets per batch; iii) Several senders can communicate with the same receiver. Also it is assumed that Alice chooses one contact of her m contacts in every communication with uniform probability and the attacker knows the number of Alice friends. Finding a minimum set of hits is an NP problem, but there is a property widely used, the frequency analysis. Through the use of frequencies, it can restrict the search space; indeed Alice's friends appear more frequently in the sets of receptors than other users. If an item is more common than others it is more likely to be included in the set of points, so the search is restricted to those groups that are more likely to be a set of hits; of course in order to get better results, attacker should have the frequency of all elements on the observations.

In [MW11a] introduces the Reverse Statistical Disclosure Attack. This attack uses observations of patterns of sending all users to estimate both the sending and receiving patterns of the target users. Estimates patterns are combined to find a set of the most likely contact of target users. It explores how the attacker could extract information from other patterns of sending users to learn more about a specific user and her contacts. The first step is for the attacker applies SDA each user to send messages. In SDA attacker is only interested in senders who send messages to Alice; RSDA instead, the attacker wants all contacts that communicate with Alice, whether those whom send or receive her messages. This is a more realistic scenario; traffic analysis is generally confined to find relationships in one direction. RSDA assumes that the attacker is interested in any contact Alice, no matter if Alice is sending messages to them or not. It is assumed that there are N users, and it established a uniform model to contact each. Specifically, each user, including Alice, has a fixed number of receivers m. The recipients are uniformly

randomly selected from the set of users. Unlike previous studies of statistical disclosure attacks, the set of senders and receivers are not separated. Moreover, each user will be taken as a receptor for some of the users. All users communicating with a particular user are included as contacts of that user. Since the attacker is focused on a specific user Alice, a distinction between the behavior of Alice and the behavior of users is made. Alice sends n_A messages in a given round, n_A is a random variable selected from a Poisson distribution with mean rate λ_A.

One of the most used strategies to attempt against SDA is sending cover traffic which consists of fake or dummy messages mixed with real ones that can hide Alice's true sending behaviour. The Statistical Disclosure Attack with Two Heads (SDA-2H) [AAGW11] is an extension of SDA [Dan03] and takes its predecessor as a baseline to improve it as it considers background traffic volumes in order to estimate the amount of dummy traffic that Alice sends. Dummy traffic serves as a useful tool to increase anonymity and they are classified based on their origin: (i) user cover, generated by the user Alice; (ii) background cover, generated by senders other than Alice in the system; (iii) receiver-bound cover, generated by the mix. This work is centered on background cover which is created when users generated false messages along with their real ones. The objective for the attacker is to estimate how much of Alice's traffic is false based on the observations between the volume of incoming and outgoing traffic. Authors make several simulations and find that for a specific number of total recipients, the increase in the background messages makes it harder for the attacker to succeed having total recipients and Alice's recipients unchanged. They also find that when Alice's recipients stay and the number of total recipients increases, the attacker would need few rounds of observations to find Alice's recipients. A comparative between SDA and SDA-2H shows that SDA-2H may not be better than SDA in all cases, but SDA-2H takes into account the effect of background cover to achieve a successful attack. Other works related with dummy traffic are [OTPG14] [MW11b]; Figure 3.6 shows a model of such systems.

[PGTO14] presents an approach of SDA using minimum squares to retrieve user profiles in a context of pool mixes. This attack models the user profile as a least squares problem by minimizing the error between the actual number of outgoing messages and an estimate based on the n incoming messages. The attack estimates the communication patterns of users in a network mix; the objective is to evaluate the probability that Alice sends

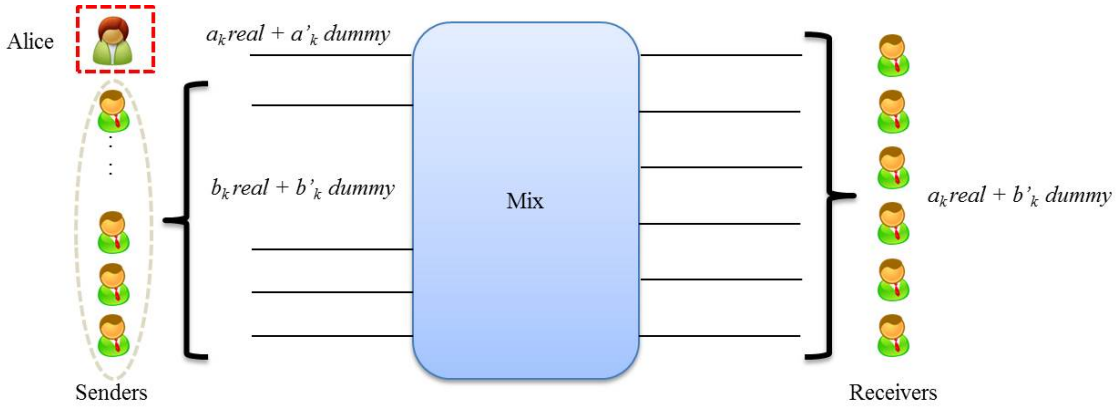


Figure 3.6: Dummy traffic model systems

a message to Bob. The assumptions are: (i) the probability of sending a message from a user to a specific receptor is independent of previous messages; (ii) the behavior of all other users are independent; (iii) any message coming into the mix is considered a priori sent by any user with uniform probability, and; (iv) the parameters used to model the probabilistic behavior does not change over time. It is assumed that each user has n receivers or friends who send messages uniformly. This attack is not only seeks to identify the set of Alice receivers, but also to estimate the probability that Alice sends or receives a message from them.

3.5 Summary

This chapter has shown an introduction to traffic analysis in anonymous communications and it has also defined the characteristics or profile of the attackers, according to their area of action, range of users, heterogeneity, location and distribution. Such details are listed in order to have a broader perspective in the field of traffic analysis. It has defined mix networks, including its description, model and operations; important elements in the construction of anonymous communication systems. It has presented one of the sub-families of the intersection attacks known as statistical disclosure attacks (SDA). In this chapter it has been listed the characteristics of each attack, hypotheses, application scenarios, among other details. Similarly, it has been briefly described techniques and methods used.

Chapter 4

Disclosure Identities Attacks

This chapter aims to present a global disclosure attack to detect relationships between users. It discuss the assumptions and general framework which are more flexible that it is used in the literature, allowing to apply automatically the method to multiple situations such as email data or social networks data. The only information used by the attacker is the number of messages sent and received by each user for each round. It shows how through contingency tables is modeled partial information. It develops a classification scheme based on combinatoric solutions of the space of round retrieved. Finally, the chapter explains the results obtained.

4.1 Framework and Assumptions

This work addresses the problem of retrieving information about relationships or communications between users in a network system, where partial information is obtained. When information is transmitted through the Internet, it is typically encrypted in order to prevent others from being able to view it. The encryption can be successful, meaning that the keys cannot be easily guessed within a very long period of time. Even if the data itself is hidden, other types of information may be vulnerable. In the e-mail framework, anonymity concerns the senders 'identity, receivers' identity, the links between senders and receivers, the protocols used, the size of data sent, timings, etc. Since [Cha81] presented the basic ideas of the anonymous communications systems, researchers have developed many mix-based and other anonymity systems for different applications, and attacks to these systems have also been developed. This work aims to develop a global statistical

attack to disclose relationships between users in a network based on a single mix anonymity system.

The information used is the number of messages sent and received by each user. This information is obtained in rounds that can be determined by equal sized batches of messages, in the context of a threshold mix, or alternatively by equal length intervals of time, in the case the mix method consists in keeping all the messages retrieved at each time interval and then relaying them to their receivers randomly reordered.

The basic framework and assumptions needed to develop our method are the following:

- Attacker knows the number of messages sent and received by each user in each round.
- The round can be determined by the system (batches) in a threshold mix context, or can be based on regular intervals of time where the attacker gets the aggregated information about messages sent and received, in the case of a timed mix where all messages are reordered and sent each period of time.
- Method is restricted, at this moment, to threshold mix with fixed batch size, or, alternatively, to a timed mix where all messages received in a fixed time period are relayed randomly reordered to their receivers.
- No restriction is made from before about the number of friends any user has, or about the distribution of messages sent. Both are considered unknown.
- Attacker controls all users in the system. In our real data application we aim at all email users of a domain send and receive within this domain.

The method introduced in this work allows to address these general settings in order to derive conclusions about relationships between users. Contrary to other methods in the literature, there are no restrictions about users relationships (number of friends, distribution of messages) and therefore can be used in wider contexts. Furthermore, our proposition is new in the methodological sense: this is a novel approach to the problem, by means of a contingency tables setting and extraction of solutions by sampling.

In an email context, this attack can be used if the attacker has access, at regular time intervals, at the information represented by the number of messages received and the number of messages sent for each user, in a closed domain or intranet where all users are

controlled. This situation extended to mobile communications or social networks, could be used, for example, in the framework of police communication investigations.

4.2 Rounds Composing

The attacker obtains, in each round, information about how many messages each user sends and receives. Usually the sender and receiver set is not the same, even if some users are senders and also receivers in some rounds. Also, the total number of users of the system N is not present in each round, since only a fraction of them are sending or receiving messages. Figure 4.1 represents a round with only 6 users.

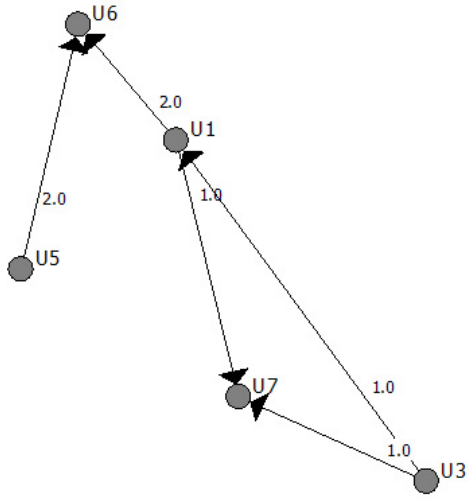


Figure 4.1: Graphical representation of one round

The information of this round can be represented in a contingency table (See Table 4.1) where the element (i, j) represents the number of messages sent from user i to user j.

Table 4.1: Example of contingency table

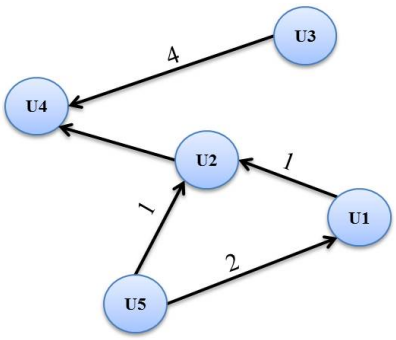
Senders\ Receivers	U1	U6	U7	Total sent
U1	0	2	1	3
U3	1	0	1	2
U5	0	2	0	2
Total received	1	4	2	7

The attacker only sees the information present in the aggregated marginals that means, in rows, the number of messages sent by each user, and in columns, the number of messages

received by each user.

In our example, only the sending pairs of vectors (U1 U3 U5) (3 2 2) and receiver pairs of vectors (U1 U6 U7) (1 4 2) are known.

Let's show another example of a round composing. Attacker need to wait and observe the network. The contingency table presented in Figure 4.2 is elaborated trough observing a communication system composed of five users. In the contingency table are the senders and receivers. It can be observed that u1 send one message to u2; u2 send 4 messages to u4 and so on. The total of messages sent and received are the marginals. Such marginals are information attackers.



Senders\ Receivers	U1	U2	U4	Total sent
U1	0	1	0	1
U2	0	0	2	2
U3	0	0	4	4
U5	2	1	0	3
Total received	2	2	6	10

Figure 4.2: Contingency table example

Consider contingency tables of equal size with batch of 10. For practical purposes, marginals rows and columns with values of 0 are eliminated, but in order to clarify the model of rounds (attacker information), the entire table is shown in Figure 4.3.

Round 1

	u1	u2	u3	u4	
u1					4
u2					3
u3					3
u4					0
	3	0	4	3	10

Round 2

	u1	u2	u3	u4	
u1					2
u2					7
u3					0
u4					1
	6	1	3	0	10

Round 3

	u1	u2	u3	u4	
u1					4
u2					0
u3					5
u4					1
	3	4	1	2	10

Figure 4.3: Rounds example

For example, the senders active elements of round 1 are u1, u2 and u3; in the other hand, the active elements acting as receivers are u1, u3 and u4. At round 2, senders active elements are u1, u2 and u4; receivers active elements are u1, u2 and u3. Finally, in round 3, senders active elements are u1, u3 and u4, and receivers active elements are u1,

u2, u3 and u4. A border red line indicates the marginals with values of 0, meaning that sender/receiver has not been active in the round.

To carry out the creation of rounds of X size, the attacker must wait X number of messages. In addition, it has been considered different periods of time of observation. Obviously, the number of rounds that will result from a time horizon of one month will be lower to the number of rounds when consider a period of 12 months.

It can build a table A summarizing all rounds with all messages sent and received for each user in the entire period of the attack. The resulting table A for the rounds shown in Figure4.3 is presented in Figure 4.4.

	u1	u2	u3	u4	
u1					10
u2					10
u3					8
u4					2
	12	5	8	5	30

Figure 4.4: Matrix A

There are many possible tables that can lead to the table with given marginals which the attacker is seeing, making it impossible in most cases to derive direct conclusions about relationships. The feasible space of tables solution of the integer programming problem can be very large. In the example, there are only 16 possible different solutions, and only one true solution.

Solutions (feasible tables) can be obtained via algorithms such as the branch and bound algorithm or other integer programming algorithms. In general they do not guarantee covering evenly all possible tables-solutions, since they are primarily designed to converge to one solution. The simulation framework presented in this chapter allows us to obtain a large quantity of feasible tables (in the most problematic rounds it takes approximately 3 minutes to obtain one million feasible tables). In many of the rounds with moderate batch size all feasible tables are obtained.

An algorithm that takes into account the information contained over all the rounds retrieved is developed in the next sections.

4.3 Hypothesis

The main objective of the proposal algorithm is to derive relevant information about the relationship (or not) between each pair of users. The information obtained by the attacker are the marginal sums, by rows and columns, of each the rounds $1, \ldots, T$ where T is the total number of rounds. Note that in each round the dimension of the table is different, since we do not take into account users that are not senders (row marginal=0), nor users that are not receivers (column marginal=0). We say element (i, j) is "present" or "active" at one round if the i and j corresponding marginals are not zero. That means that user i is present in this round as sender and user j is present as receiver.

As mention before, a final aggregated table A can be build, summing up all the rounds and obtaining a table with all messages sent and received from each user at the whole time interval considered for the attack. Each element (i, j) of this final table would represent the number of messages sent by i to j in total. Although the information obtained in each round is more precise and relevant (because of the lower dimension and combinatoric possibilities), an accurate estimate of the final table is the principal objective because a zero in elements (i, j) and (j, i) would mean no relationship between these users (no messages sent from i to j nor from j to i). A positive number in an element of the estimated final table would mean that some message is sent in some round, while a zero would mean no messages sent in any round, that is, no relationship.

4.4 Feasible Tables

4.4.1 Algorithm

We consider all rounds as independent events. The first step is to obtain the higher number of feasible tables it is possible for each round, taking into account time restrictions. This will be the basis of our attack. Our method is based in [CDHL05], and it consists in filling the table column by column, and computing the new bounds for each element before it is generated.

1. Begin with column one, row one: Generate n_{11} from an integer Uniform distribution in the bounds according to equation 4.1 where $i = 1, j = 1$. Let r be the number of rows.

2. For each row element n_{k1} in this column, if row elements until $k-1$ have been obtained, new bounds for n_{k1} are according to next equation

$$\max(0, (n_{+1} - \sum_{i=1}^{k-1} n_{i1}) - \sum_{i=k+1}^{r} n_{i+}) \leq n_{k1} \leq \min(n_{k+}, n_{+1} - \sum_{i=1}^{k-1} n_{i1}) \tag{4.1}$$

   The element n_{k1} is then generated by an integer uniform in the fixed bounds.

3. Last row element is automatically filled since lower and upper bounds coincide, letting $n_{(k+1)+} = 0$ by convenience.

4. Once this first column is filled, row margins n_{i+} and total count n are actualized by subtraction of the already fixed elements, and the rest of the table is treated as a new table with a column less.

The algorithm fixes column by column until the whole table is filled.

Consider the following example in Figure 4.5, composed of one round of batch 9 and just four feasible tables obtained applying the algorithm above. For didactical purposes, each round is displayed with a table of equal size. When a user is not "present" or "active" as sender/ receiver in that round, the corresponding row or column has the value of 0. In the other side, for practical purposes, each table corresponding to a round has different dimensions because it does not take into account users who do not send or receive messages. We say that an element (i, j) is "present" in a round if its marginal is different from zero.

4.4.2 Algorithm Performance

Time employed depends on the complexity of the problem (number of elements, mean number of messages) In our email data even for large number of elements it has not been a problem. For large table sizes in our applications, It takes approximately 3 minutes to obtain one million feasible tables in rounds with 100 cells and 10in a PC with Intel processor 2.3 Ghz and 2 Gb Ram.

Repeating the algorithm as it is written for each generated table does not lead to uniform solutions, that is, some tables are more probable than others due to the order used when filling columns and rows. Since we must consider a priori all solutions for a determined round equally possible, two further modifications are made:

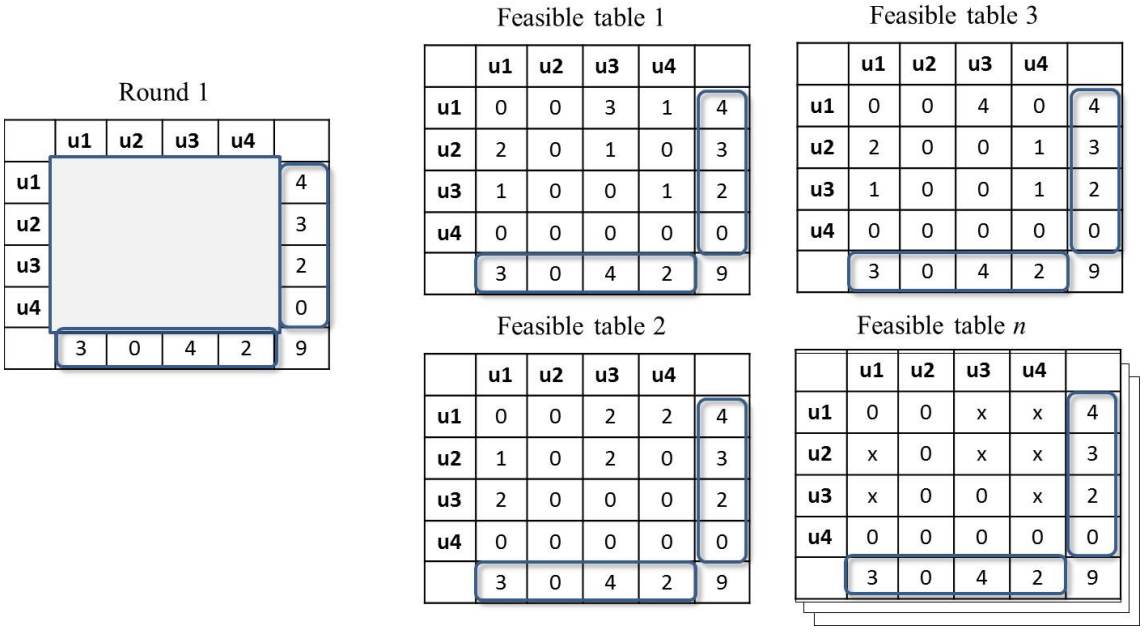


Figure 4.5: Example of one round and its feasible tables

i. Random reordering of rows and columns before a table is generated

ii. Once all tables are generated, only distinct tables are kept to make inferences.

These two modifications have resulted in an important improvement of the performance of our attack, lowering the mean misclassification rate about a 20% in our simulation framework.

4.4.3 Calculating the Number of Feasible Tables

Deciding the number of tables to be generated poses an interesting problem. Computing the number of distinct feasible tables for a contingency table with fixed marginals is still an open problem, which has been addressed via algebraic methods [Rap03], and by asymptotic approximations [GM08], but in our case the margin totals are small and depend on the batch size; therefore it is not guaranteed that asymptotic approximations hold. The best approximation so far to count the feasible tables is to use the generated tables.

[CDHL05] shows that an estimate of the number of tables can be obtained by averaging over all the generated tables the value $\frac{1}{q(T)}$ a ccording to next algoritm:

Algorithm 2

1. $q(T)$ is the probability of obtaining the table T, and is computed iteratively imitating the simulation process according to Equation 4.2.

2. $q(t_1)$ is the probability of the actual values obtained for column 1, obtained by multiplying the uniform probability for each row element in its bounds. $q(t_1 \mid t_2)$ and subsequent terms are obtained in the same way, within the new bounds restricted to the precedent columns fixed values.

$$q(T) = q(t_1)q(t_1 \mid t_2)q(t_3 \mid t_1,t_2) \dots q(t_c \mid t_1,t_2,\dots,t_{c-1}) \tag{4.2}$$

The number of feasible tables goes from moderate values as 100,000, that can be easily addressed, getting all possible tables via simulation, to very high numbers as 1013. Generating all possible tables for this last example would take with the computer we are using, a Windows 7 PC with 2.3 Ghz and 4 Gb Ram, at least 51 days. The quantity of feasible tables is the main reason why it is difficult that any deterministic intersection-type attack works, even with low or moderate users dimensions. Statistical attacks need to consider relationships between all users to be efficient, because the space of solutions for any individual user is dependent of all other users marginals. Exact trivial solutions can be however found at some time in the long run, if a large number of rounds are obtained.

In our settings we try to obtain the largest number of tables we can, given our time restrictions, obtaining a previous estimate of the number of feasible tables and fixing the highest number of tables that can be obtained for the most problematic rounds. However, an important issue is that once a somewhat large number of tables is obtained, good solutions depend more on the number of rounds treated (time horizon or total number of batches considered) that on generating more tables. In our simulations, there is generally a performance plateau in the curve that represents misclassification rate versus the number of tables generated, since a sufficiently high number of tables are reached. This minimum number of tables to be generated depends on the complexity of the application framework.

4.5 Results on Matrix

The final information obtained consists of a fixed number of generated feasible tables for each round. In order to obtain relevant information about relationships, there is a need to fix the most probable zero elements. For each element, the sample likelihood function at zero $\hat{f}(X\|p_{ij}=0)$ is estimated. This is done by computing the percent of tables with that element being zero in each round the element is present, and multiplying the estimated likelihood obtained in all these rounds (the element will be zero at the final table if it is zero at all rounds).

If we are estimating the likelihood for the element (i,j), and are generating M tables per round, we use the following expressions:

$n_t^{(i,j)}$=number of tables with element (i,j)=0 in round t.

$N_{present}$=number of rounds with element (i,j) present.

X=sample data, given by marginal counts for each round.

$$log(\widehat{f}(X \mid p_{ij}=0)) = -N_{present}\log(M) + \sum_{t=1,(i,j)present}^{T} \log(n_t^{(i,j)}) \tag{4.3}$$

Final table elements are then ordered by the estimated likelihood at zero, with the exception of elements that were already trivial zeros (elements that represent pair of users that have never being present at any round).

Elements with lowest likelihood are then considered candidates to insert as "relationship". The main objective of the method is to detect accurately:

a. Cells that are zero with a high likelihood (not relationship $i \to j$)

b. Cells that are positive with high likelihood (relationship $i \to j$).

In our settings the likelihood values at $p_{ij}=0$ are bounded in the interval $[0,1]$. Once these elements are ordered by most likely to be zero to less, a classification method can be derived based on this measure.

A theoretical justification of the consistency of the ordering method is given below.

Proposition 1

Let consider, a priori, that for any given round k all feasible tables, given the marginals, are equiprobable.

Let p_{ij} the probability of element (i, j) being zero at the final matrix A, which is the aggregated matrix of sent and received messages over all the rounds. Then the product of the proportion of feasible tables with $x_{ij} = 0$ at each round, Q^{ij}, leads to an ordering between elements such that if $Q^{ij} > Q^{i'j'}$ then the likelihood of data for $p_{ij} = 0$ is bigger than the likelihood of data for $p_{i'j'} = 0$.

Proof

If all feasible tables for round k are equiprobable, the probability of any feasible table is $p_k = \frac{1}{\#[X]_k}$, where $\#[X]_k$ is the total number of feasible tables in round k.

For elements with $p_{ij} = 0$, it is necessary that $x_{ij} = 0$ for any feasible table. The likelihood for $p_{ij} = 0$ is then

$P([X]_k \mid p_{ij} = 0) = \frac{\#[X|x_{ij}=0]_k}{\#[X]_k}$

where $\#[X \mid x_{ij} = 0]_k$ denotes the number of feasible tables with the element $x_{ij} = 0$.

Let $k = 1, ..., t$ independent rounds. The likelihood at $p_{ij} = 0$, considering all rounds, is

$Q^{ij} = \prod_{k=1}^{t} P([X]_k \mid p_{ij} = 0) = \prod_{k=1}^{t} \frac{\#[X|x_{ij}=0]_k}{\#[X]_k}$

and the log likelihood:

$\log(Q^{ij}) = \sum_{k=1}^{t} \log(\#[X \mid x_{ij} = 0]_k) - \sum_{k=1}^{t} \log(\#[X]_k)$.

Then the proportion of elements with $x_{ij} = 0$ at each round leads to an ordering between elements such that if $Q^{ij} > Q^{i'j'}$ then the likelihood of data for $p_{ij} = 0$ is bigger than the likelihood of data for $p_{i'j'} = 0$.□

Our method is not based on all the table solutions, but on a consistent estimator of Q^{ij}. For simplicity, let consider a fixed number of M sampled tables at every round.

Proposition 2

Let $[X]_k^1$,...,$[X]_k^M$ a random sample of size M of the total $\#[X]_k$ of feasible tables for round k. Let $w_k^{(i,j)} = \frac{\#[X|x_{ij}=0]_k^M}{M}$ the sample proportion of feasible tables with $x_{ij} = 0$ at round k. Then the statistic $q^{ij} = \prod_{k=1}^{t} \frac{\#[X|x_{ij}=0]_k^M}{M}$ is such that, for any pair of elements (i, j) and (i', j') , $q^{ij} > q^{i'j'}$ implies, in convergence, higher likelihood for $p_{ij} = 0$ than for $p_{i'j'} = 0$.

Proof

a. Let $\#[X]_k$ the number of feasible tables at round k. Let $[X]_k^1, \ldots, [X]_k^M$ a random sample of size M of the total $\#[X]_k$. Random reordering of columns and rows in Algorithm 1, together with the elimination of equal tables, assures it is a random sample. Let $\#[X \mid x_{ij} = 0]_k^M$ the number of sample tables with element $x_{ij} = 0$. Then the proportion $w_k^{(i,j)} = \frac{\#[X|x_{ij}=0]_k^M}{M}$ is a consistent and unbiased estimator of the true proportion $W_k^{(i,j)} = \frac{\#[X|x_{ij}=0]_k}{\#[X]_k}$. This is a known result from finite population sampling. As $M \to \#[X]_k$, $w_k^{(i,j)} \to W_k^{(i,j)}$.

b. Let $k = 1, ..., t$ independent rounds. Then given a sample of proportion estimators $w_1^{(i,j)}, ..., w_t^{(i,j)}$ of $W_1^{(i,j)}, ..., W_t^{(i,j)}$, consider the function $f(w_1^{(i,j)}, ..., w_t^{(i,j)}) = \sum_{k=1}^{t} \log(w_k^{(i,j)})$ and $f(W_1^{(i,j)}, ..., W_t^{(i,j)}) = \sum_{k=1}^{t} \log(W_k^{(i,j)})$. Given the almost sure convergence of each $w_k^{(i,j)}$ to each $W_k^{(i,j)}$ and the continuity of the logarithm and sum functions, the continuous mapping theorem assures convergence in probability, $f(w_1^{(i,j)}, ..., w_t^{(i,j)}) \xrightarrow{P} f(W_1^{(i,j)}, ..., W_t^{(i,j)})$. Then $log(q^{ij}) = f(w_1^{(i,j)}, ..., w_t^{(i,j)})$ converges to $log(Q^{ij}) = f(W_1^{(i,j)}, ..., W_t^{(i,j)})$. Since the exponential function is continuous and monotonically increasing, applying the exponential function in both sides leads to the convergence of q^{ij} to Q^{ij}, so that $q^{ij} > q^{i'j'}$ implies, in convergence, $Q^{ij} > Q^{i'j'}$, and then higher likelihood for $p_{ij} = 0$ than for $p_{i'j'} = 0$.□

Given all pairs of senders and receivers (i, j) ordered by the statistic q^{ij}, it is necessary to select a cut point in order to complete the classification scheme and decide whether a pair do communicate ($p_{ij} > 0$) or not ($p_{ij} = 0$). That is, it is needed to establish a value c such that $q^{ij} > c$ implies $p_{ij} = 0$, and $q^{ij} \leq c$ implies $p_{ij} > 0$.The defined statistic q^{ij} is bounded in $[0, 1]$, but this is not strictely a probability, so fixing a priori a cutpoint such as 0.5 is not an issue. Instead, there are some approaches that can be used:

1. In some contexts (email, social networks) the proportion of pairs of users that communicate is approximately known . This information can be used to select the cut point from the ordering. That is, if about 20% of pairs of users are known to communicate, the classifier would give value '0' (not communication) to the upper 80% elements (i, j), ordered by the statistic q^{ij}, and value '1' (communication) to the lower 20% of elements.

2. If the proportion of zeros is unknown, it can be estimated, using the algorithm for obtaining feasible tables over the known marginals of the matrix A and estimating the proportion of zeros by the mean proportion of zeros over all the simulated feasible tables.

4.6 Performance on Email Data

In this section simulations are used to study the performance of the attack.

Each element (i, j) of the matrix A can be 0 (not communication) or strictly positive. The percentage of zeroes in this matrix is a parameter, set a priori to observe its influence. In a closed-center email communications, this number can be between 70% and 99%. However, intervals from 0.1 (high communications density) to 0.9 (low communications density) are used here to aim to different practical situations. Once this percentage is set, a randomly chosen percent of elements are set to zero and then are zero for all the rounds.

The mean number of messages per round for each positive element (i, j) is also set a priori. This number is related, in practice, to the batch size the attacker can get. As the batch size (or time length interval of the attack) decreases, the mean number of messages per round decreases, making the attack more efficient.

Once the mean number of messages per round is determined for each positive element (λij), a Poisson distribution with mean λij, $P(\lambda ij)$, is used to generate the number of messages for each element, for each of the rounds.

External factors, given by the context (email, social networks, etc.) that have an effect upon the performance of the method are monitorized to observe their influence:

1. **The number of users**: In a network communication context with N users, there exist N potential senders and N receivers in total, so that the maximum dimension of the aggregated table A is N^2.

2. As the number of users increases, the complexity of round tables and the number of feasible tables increases so that it could affect negatively the performance of the attack.

3. **The percent of zero elements in the matrix** A: These zero elements represent not communication between users. As it will be seen, it influences the performance of the method.

4. **The mean frequency of messages per round for positive elements**: It is directly related to the batch size, and when it increases, performance is supposed to be affected negatively.

5. **The number of rounds**: As the number of rounds increases, it is supposed to improve the performance of the attack, since more information is available. One factor related to the settings of the attack method is also studied.

6. **The number of feasible tables generated by round**: This affects computing time, and it is necessary to study at what extent it is useful to obtain too many tables. This number can be variable, depending on the estimated number of feasible tables for each round.

The algorithm results in a binary classification, where 0 in an element (i, j) means no relationship sender-receiver from i to j, and 1 means positive relationship sender-receiver. Characteristic measures for binary classification tests include the sensitivity, specificity, positive predictive value and negative predictive value. Letting TP as the true positives, FP as false positives, TN as true negatives, and FN as false negatives:

Sensitivity=$\frac{TN}{TN+FP}$ measures the capacity of the test to recognize true negatives.

Specificity=$\frac{TP}{TP+FN}$ measures the capacity of the test to recognize true positives.

Positive predictive value=$\frac{TP}{TP+FP}$ measures the precision of the test to predict positive values.

Negative predictive value =$\frac{TN}{TN+FN}$ measures the precision of the test to predict positive values.

Classification rate=$\frac{TN+TP}{TN+TP+FN+FP}$ measures the percent of elements well classified.

The figures 4.6 and 4.7 show simulation results. When it is not declared, values of $p_0 = 0 : 7$; $\lambda = 2$, $N = 50$ users and number of rounds= 100 are used as base values.

Figure 4.6 shows that as the number of cells (N^2, where N is the number of users) increases, and percent of cells that are zero decreases, the number of feasible tables per round increases. For moderate number of users such as 50, the number of feasible tables is already very high, greater than 1020. This does not have a strong effect over the main results, except for lower values. As it can be seen in Figure 4.6, once a sufficiently high number of tables per round is generated, increasing this number does not lead to significant improvement on the correct classification rate.

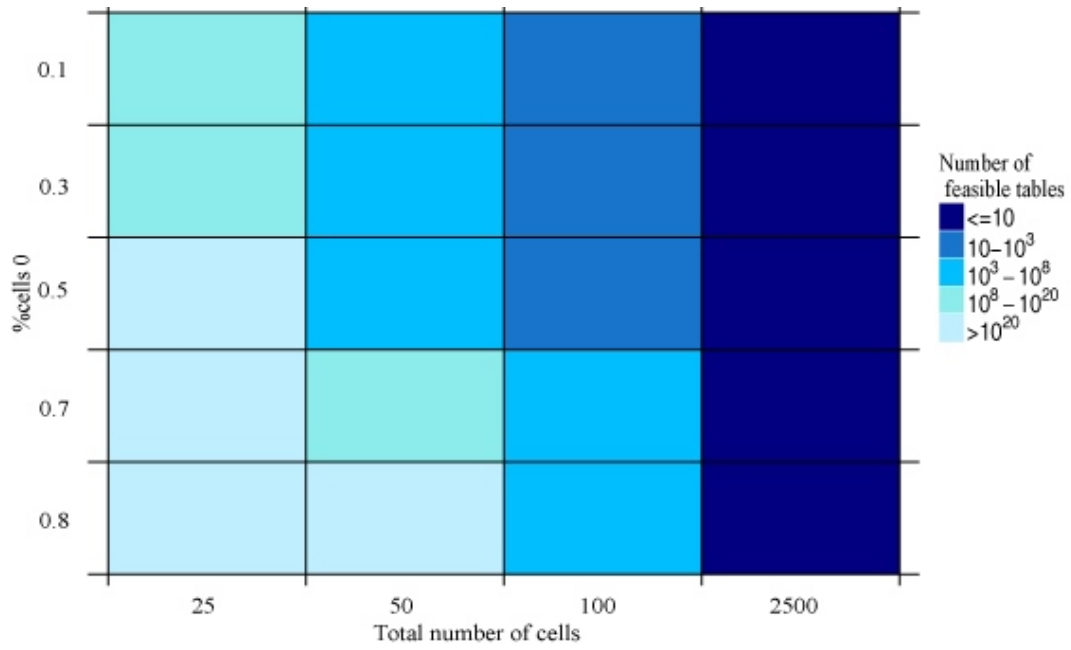


Figure 4.6: Number of feasible tables per round, depending on % of cells zero and total number of cells

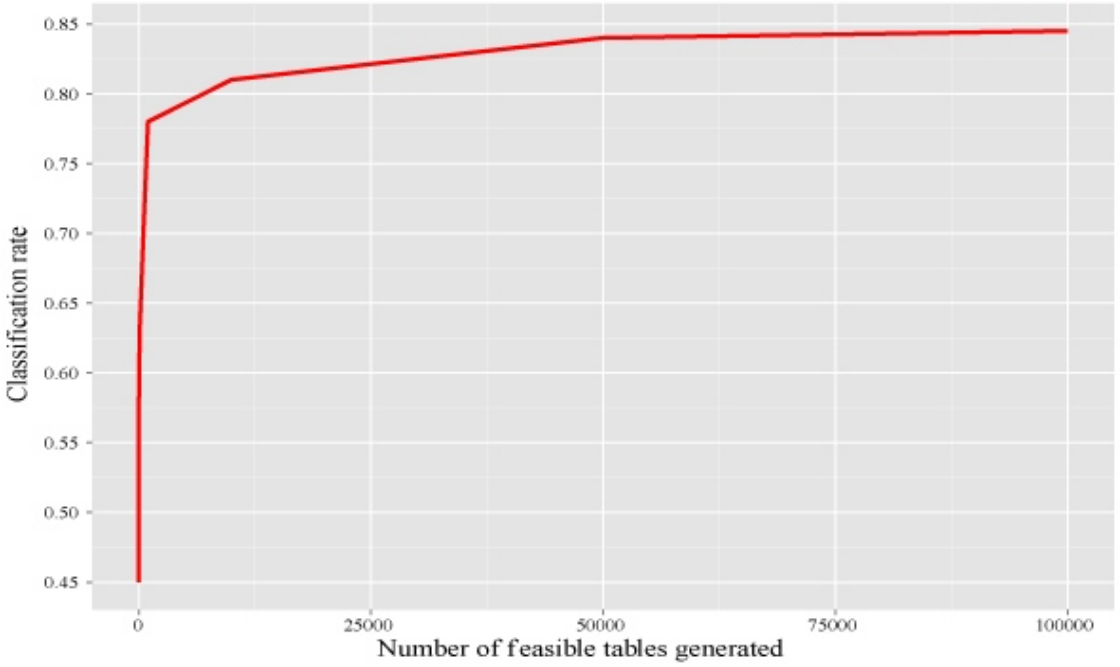


Figure 4.7: Classification rate as function of the number of feasible tables per round

Figure 4.8 shows that the minimum classification rate is attained at a percent of cells zero (users that do not communicate) near 0.5. As this percent increases, the true positive rate decreases and true negative rate increases.

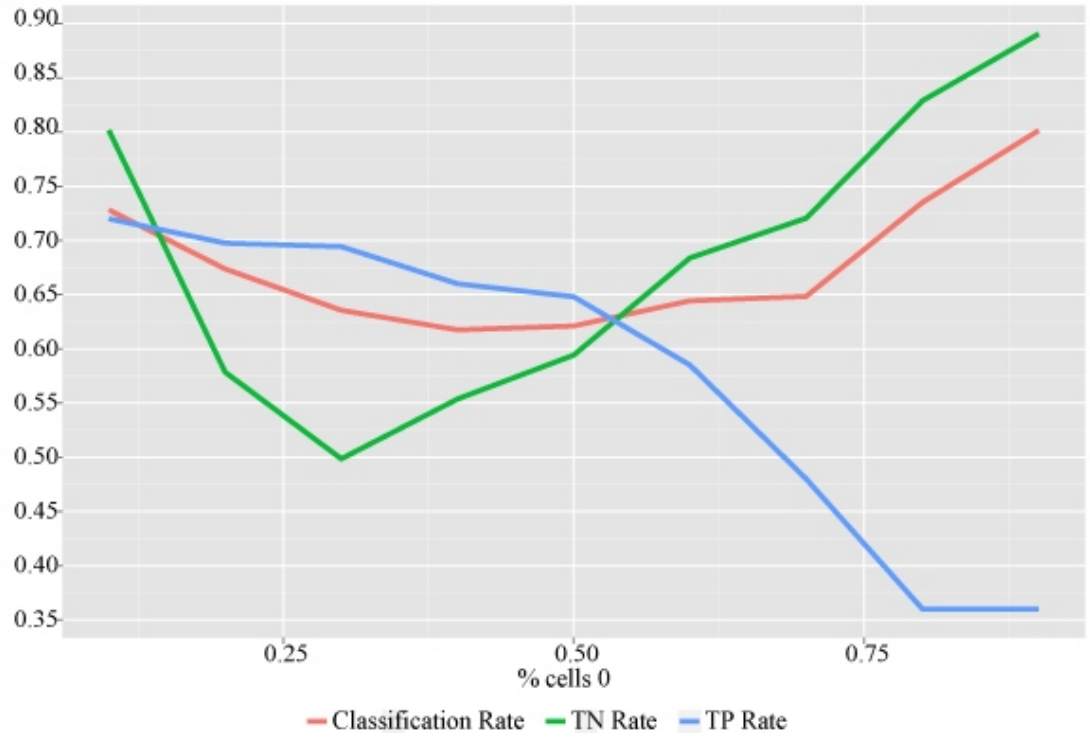


Figure 4.8: Classification rate, true positives rate and true negatives rate

As the attacker gets more information, that is, more rounds retrieved, the classification rate gets better. Once a high number of rounds is obtained, there is no further significative improvement, as it is shown in Figure 4.9.

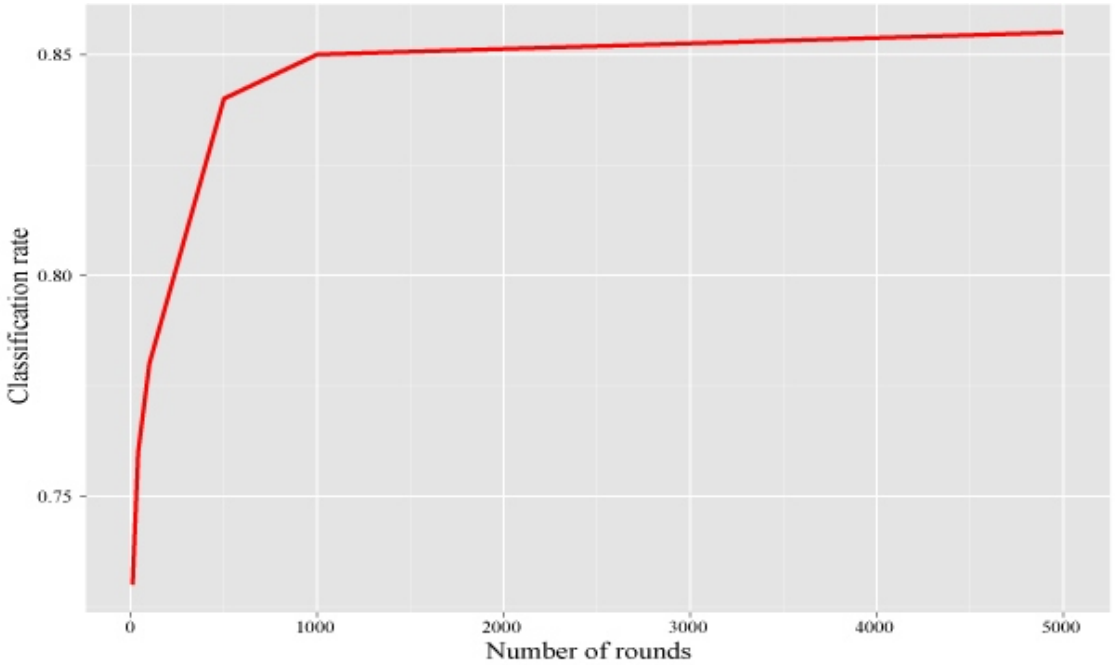


Figure 4.9: Classification rate vs. Number of rounds obtained

In Figure 4.10, it is shown that as the number of messages per round (λ) for users that communicate increases, the classification rates decrease. This is a consequence of the complexity of the tables involved (more feasible tables). This number is directly related to the batch size, so it is convenient for the attacker to obtain data in small batch sizes, and for the defender to group data in large batch sizes, leading to lower latency.

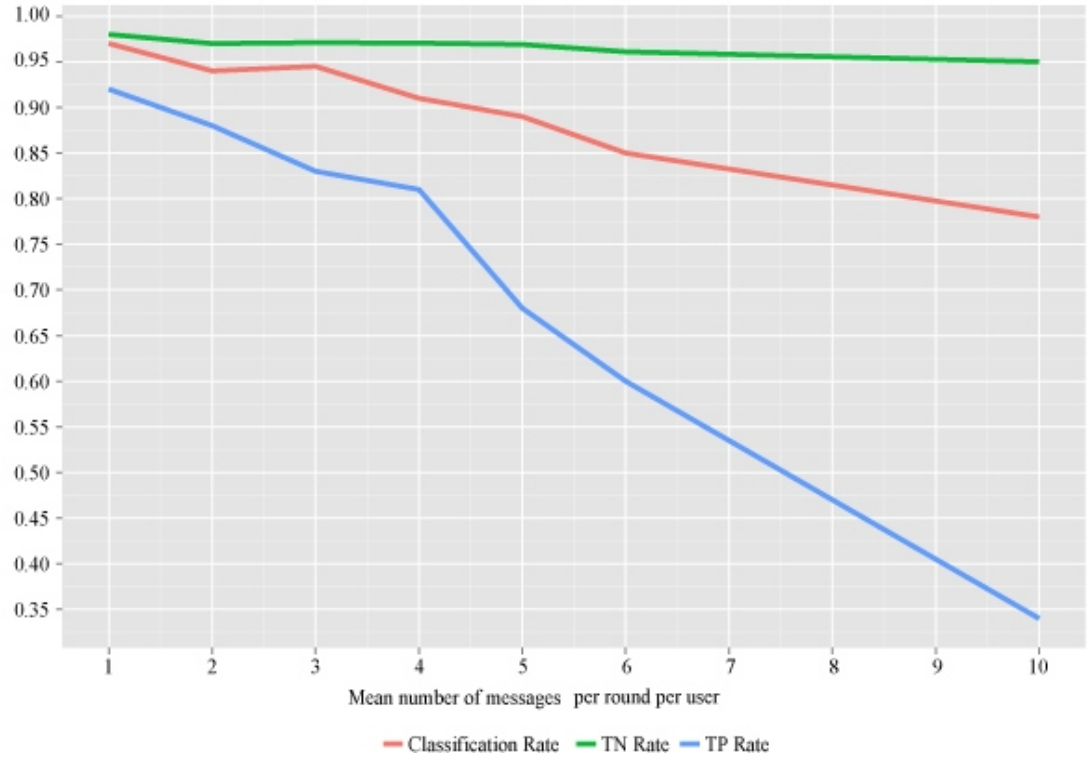


Figure 4.10: Classification rate vs. Mean number of messages per round

The complexity of the problem is also related to the number of users, as can be seen in Figure 4.11, where the classification rate decreases as the number of users increases.

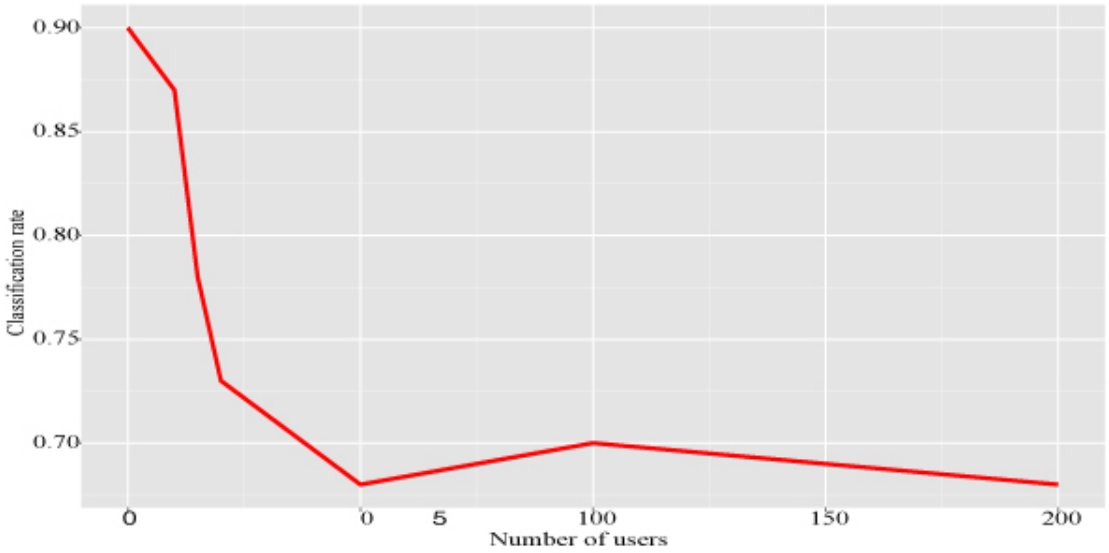


Figure 4.11: Classification rate vs. Number of users

4.7 Summary

This chapter presented a method to detect relationships (or non-existent relationships between users in a communication framework, when the retrieved information is incomplete. It has detailed the framework and assumptions in order to execute the proposal attack. The attacker gets partial information that can be modeled through contingency tables. It has been described the algorithm to calculate as much as possible solutions or feasible tables for each contingency table. This method can also be used to other communications framework, such as social networks or peer to peer protocols, and to real de-anonymization problems not belonging to the communications domain. It has been demonstrated the selection of optimal cut points, optimal number of generated tables and further refinements of the final solution. Finally it showed the results obtained after applying the algorithm to email data.

Chapter 5

Method Improvements and Behavior

This chapter aims to show an enhancement of a previously presented statistical disclosure attack. The improvement of the attack is based on the use of the EM algorithm to get better estimation of the messages sent by users and to derive what pair of users really communicates. It presents the framework and assumptions of the attack considering that attacker gets partial information based on his observations. It develops two methods using the EM algorithm to improve the estimation of messages sent, and the best method is used over real email data over 32 different network domain. It also presents a comparative between our method and an algorithm based on the Maximum Likelihood approach. This chapter concludes with a brief review of this chapter.

5.1 Refinement Method Based in the EM Algorithm

This method intends to solve the problem of retrieving information about relationships or communications between users in a network system, where partial information is obtained. The information used is the number of messages sent and received by each user. This information is obtained in rounds that can be determined by equally-sized batches of messages, in the context of a threshold mix, or alternatively by equal length intervals of time, in the case that the mix method consists of keeping all of the messages retrieved at each time interval and then relaying them to their receivers, randomly reordered.

The attacker can only retrieve the number of messages sent and received for each user

in each round as represented in Figures 5.1 and 5.2. In each round, not all the users must be present. A final adjacency matrix A that represents aggregated information from all the rounds is also built by the attacker.

	Receivers			
Senders	u2	u3	u5	
u1	2	1	0	3
u4	1	0	2	3
u7	1	0	1	2
	4	1	3	**8**

Figure 5.1: Round example: The attacker only sees the unshaded information

Round 1

	u2	**u3**	**u5**	
u1				3
u4				3
u7				2
	4	1	3	**8**

Round 2

	u3	**u4**	**u6**	**u8**	
u1					2
u3					1
u6					2
	1	1	2	1	**5**

Round n

	u5	**u6**	**u10**	
u1				1
u3				1
u4				1
u8				3
	2	2	2	**6**

Aggregated Matrix A

	u1	**u2**	**u6**	**u20**	
u1					16
u2					10
…					…
u20					15
	10	13	…	10	**130**

Figure 5.2: Information retrieved by the attacker in rounds

The main objective of the attacker is to derive, for each pair of users ij, if there has been positive communication or not during the study period. Considering a final true adjacency matrix A' where all messages from the original A matrix for all rounds are summed up, and marking as 1 matrix elements that are strictly positive (there has been communication in at least one round) and allowing that 0 elements that are already zero, the objective of the attacker is to develop a classifier that predicts each cell into 1 (communication) or 0 (not communication). This classifier would lead to an estimate matrix $\hat{A}'$ and diagnostic measures could be computed based on the true matrix A' and its estimate $\hat{A}'$. This is an unsupervised problem, since generally the attacker does not know a priori any communication pattern between users.

5.1.1 Framework and Definitions

Let's consider the following general settings for the attack:

- The attacker knows the number of messages sent and received by each user in each round.

- The round can be determined by the system (batches) in a threshold mix context or can be based on regular intervals of time, where the attacker gets the aggregated information about messages sent and received, in the case of a timed mix, where all messages are reordered and sent each period of time.
- No restriction is made from before about the number of friends any user has nor about the distribution of messages sent. Both are considered unknown.
- The attacker controls all users in the system. In our real data application, we aim at all email users of a domain sent and received within this domain.

In [PGVST+15] the obtaining of feasible tables was addressed through a generalized extraction, attempting to obtain feasible tables over all combinatorial regions, giving equal weight to every table. Three features of the algorithm were used in order to achieve this global representation: uniform generation of table cell values, successive random rearrangement of rows and columns before table generation, and deletion of equal feasible tables once a number of tables were obtained.

The method was applied to simulated data with good results, and the refinement presented here is applied on real email data. The performance of the method is affected by these features:

1. **The number of users**: As the number of users increases, the complexity of round tables and the number of feasible tables increases, so that it negatively affects the performance of the attack.
2. **The percentage of zero elements in the matrix A**: These zero elements represent no communication between users.
3. **The mean frequency of messages per round for positive elements**: This is directly related to the batch size, and when it increases, the performance is negatively affected.
4. **The number of rounds**: As the number of rounds increases, this improves the performance of the attack, since more information is available.
5. **The number of feasible tables generated by round**: This affects computing time, and it is necessary to study to what extent it is useful to obtain too many tables. This number can be variable.

A further modification of the method is presented in the next section.

In spite of the interesting results obtained in the previous research using this algorithm, feasible tables that match table marginal values usually have different probabilities of being true. A further refinement of the algorithm taking in account this fact is developed in this section.

In a first setting of the following refinement of the algorithm, the number of messages sent per round by user i to the user j is modeled by a Poisson distribution with parameter λ_{ij}. This is a simplification of the underlying non-homogeneous Poisson process (this rate may change over time). This simplification is motivated by the fact that the rounds, defined by the attacker, may be constructed by batches of messages or alternatively, by time periods.

Also, approximating a non-homogeneous Poisson process by a homogeneous Poisson process is a frequent decision when information is limited, as is the case in the problem treated here.

Within this modeling approach, the number of messages sent by round by user i will follow a Poisson distribution with parameter $\lambda_i = \sum_{j=1}^{receivers} \lambda_{ij}$ and the number of messages received by round by user j will follow a Poisson distribution with parameter $\lambda_j = \sum_{i=1}^{senders} \lambda_{ij}$. Pairs of users that do not communicate will have a degenerated distribution with fixed rate $\lambda_{ij} = 0$.

Each round is an independent realization of a batch of messages sent and received. In each round the attacker observes the number of messages sent by each user i, x_i^r, and the number of messages received by each user j, y_j^r. It should be noted that an unbiased estimator $\widehat{\lambda}_i$ of the rate λ_i is the average number of messages sent per round by the user $i, \overline{x_i} = \frac{1}{n}\sum_{r=1}^{n} x_i^r$. In the same way $\overline{y_i} = \frac{1}{n}\sum_{r=1}^{n} y_i^r$ is an unbiased estimator of λ_j. An initial estimator of λ_{ij} can be obtained through the independence assumption in the final aggregated table A obtained aggregating all the round marginals . In this case, using the well known statistical results in contingency tables under the independence hypothesis, $(\sum_{r=1}^{n} x_i^r \sum_{r=1}^{n} y_i^r)/N$ is an estimator of the total number of messages sent from user i to j for all the rounds and λ_{ij} can be estimated by $\widehat{\lambda}_{ij} = (\sum_{r=1}^{n} x_i^r \sum_{r=1}^{n} y_i^r)/Nn$ where N is the total number of messages sent in all rounds. Obviously the independence hypothesis does not apply, since senders have different preferences over the space of receivers, but it is a good departure point given the limited information available.

5.1.2 Algorithm

In order to refine the estimation of λ_{ij}, it has used the EM algorithm [DLR77]. This algorithm allows to estimate parameters by means of maximum likelihood approach, in situations where it is too difficult to obtain direct solutions from the maximum likelihood optimization equations. Generally this algorithm is used when a probabilistic model exists where X, is the observed data, θ is a vector of parameters, and Z is the latent, non observed data.

The likelihood function is $L(\theta; X, Z) = p(X, Z \mid \theta)$. Since Z is unknown, .likely function of θ is set as $L(\theta, X) = \sum_z P(X, Z \mid \theta)$. This function is not easy to maximize due to the complexity of sum up in Z (frequently multidimensional). The EM algorithm (Expectation-Maximization) allows us to approach the problem in two steps iteratively, after the assignment of an initial value $\theta^{(1}$. In each step t the next two operations are made:

1. **Expectation Step (E-Step)**: The expectation of $L(\theta, X, Z)$ under the distribution of Z conditional to the values of X and $\theta^{(t)}$ is derived: $Q(\theta \mid \theta^{(t)}) = E_{Z|X,\theta^{(t)}}[L(\theta, X, Z)]$

2. **Maximization Step (M-Step)**: $Q(\theta \mid \theta^{(t)})$ is maximized in θ, obtaining a new value $\theta^{(t+1)}$ for θ.

This process is realized iteratively until convergence.

In the present problem, X^r is the information observed by the attacker and represents the marginal sums in each round r. Z^r are the unknown values of the cells of the table in the round r.. The parameter vector is denoted by λ.

For each round, Z^r cell values are a priori pairwise independent, and rounds are generated independently .Also, Z^r values that do not match the round marginals X^r have 0 probability. Then

$$P(Z^r \mid X^r, \lambda) \propto \prod_{i,j} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

for all Z^r compatible with X^r marginal values. Proportionality is fixed with respect

to the sum over all feasible tables in round r:

$$\sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!}$$

where T^r represents the set of all feasible tables with marginals X^r and z_{ijt}^r is referred to the cell values for each table t from the set T^r .

$$P(Z^r \mid X^r, \lambda) = 0 \ \ \forall \ \ Z^r$$

incompatible with X^r marginal values.

Calling X and Z the information for all rounds:

$$P(Z \mid X, \lambda) = \prod_{r=1}^{n} \prod_{i,j} \left(\sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!} \right)^{-1} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

for all Z^r compatible with X^r marginal values.

Since $P(X = x \mid \lambda)$ is the probability of all feasible tables leading to x,

$$P(X \mid \lambda) = \prod_{r=1}^{n} \sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!}$$

.

Then the likelihood is

$$L(\lambda; X, Z) = P(X, Z \mid \lambda) = P(Z \mid X, \lambda) \cdot P(X \mid \lambda) = \prod_{r=1}^{n} \prod_{i,j} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

In this expression z_{ij}^r (the cell values in each round) are latent values, not observed . The EM algorithm is applied in this context to estimate the λ_{ij} values by maximum likelihood. The initial value for λ_{ij} will be set as the independence hypothesis estimate

$$\widehat{\lambda}_{ij} = \frac{\overline{x_i y_j}}{n}$$

1. **E-Step**: In this step it is necessary to approach the expectation of $L(\lambda, X, Z)$ under the distribution $P(Z \mid X, \lambda)$. The Monte Carlo method is used to approach $E_{Z|X,\lambda}[L(\lambda, X, Z)]$ by $\frac{1}{m} \sum_{k=1}^{m} L(\lambda, X, Z_k)$, where Z values are obtained by k generations from the conditional distribution $P(Z \mid X, \lambda)$ for each round.Since

working with the logarithm, the likelihood leads to the same optimization process the following approximation is applied:

$$\widehat{E}_{Z|X,\lambda}[log(L(\lambda, X, Z))] = \frac{1}{m}\sum_{k=1}^{m}\sum_{r=1}^{n}\sum_{i,j}\log\left(\frac{\lambda_{ij}^{z_{ijk}^r}e^{-\lambda_{ij}}}{(z_{ijk}^r)!}\right)$$

In order to obtain samples from $P(Z \mid X, \lambda)$ for each round the algorithm presented in [PGVST+15] is applied, but in this case the feasible tables are generated in each cell generation, instead of the uniform distribution, a Poisson distribution with rate $\widehat{\lambda}_{ij}$ truncated by X^r marginal limitations.

2. **M-Step**: In order to maximize the expression $\widehat{E}_{Z|X,\lambda}[L(\lambda, X, Z)]$ with respect to λ_{ij}, the maximization process is developed as is usual in the Poisson distribution parameter estimation. This results in $\widehat{\lambda}_{ij} = \overline{\overline{z}}_{ij}$ where the mean is taken over the sample feasible tables and all the rounds.This estimated value $\widehat{\lambda}_{ij}$ will be used subsequently in the Monte Carlo table generation referred in step 1.

Steps 1 and 2 are repeated iteratively until convergence.

This application of the EM algorithm leads to the final estimates $\widehat{\lambda}_{ij}$. In order to obtain an estimate of the adjacency matrix $\widehat{A}'$ the ordering of cells is then fixed based on probability of zero for each cell, that is, under the Poisson modeling, $P(z_{ij} = 0) = e^{-\lambda_{ij}}$. A cut point is then selected to apply to ordering list. It can be based on external information, or based on estimation through extracting feasible tables from the A matrix, restricted to sure zero and positive cells already detected by the EM algorithm. The chosen cut point is used to classify cells ij into 0 or 1 obtaining the estimate $\widehat{A}'$ of the true adjacency matrix A'.

The later approach uses the Poisson distribution to model the number of messages sent per round, as is usual in applications. Next, another approach is applied.

Let's model the distribution of the number of messages sent per round by user i to the user j as a discrete tabulated distribution with parameters $(p_{ij0}, p_{ij1}, p_{ij2}, ...)$ where p_{ijt} represents the probability the sender i sends t messages to the user j in a round.

In order to develop a new version of the EM algorithm above, denoting p by the matrix of parameters, it results in

$$P(Z \mid X, p) = \prod_{r=1}^{n} \prod_{i,j} \left(\frac{1}{\sum_{T^r} \prod_{i,j} p_{ijz_{ij}^r}} \right)^{-1} p_{ijz_{ij}^r}$$

for all Z compatible with the marginals X, and the E-Step gives

$$\widehat{E}_{Z|X,p}[log(L(p, X, Z))] = \frac{1}{m} \sum_{k=1}^{m} \sum_{r=1}^{n} \sum_{i,j} \log \left(p_{ijz_{ij}^r} \right)$$

Simple maximization in each p_{ijt} leads to estimate p_{ijt} through the sample proportion the cell ij takes the value t :

$$\widehat{p}_{ijt} = \frac{1}{nm} \sum_{k=1}^{m} \sum_{r=1}^{n} I(z_{ij}^r = t)$$

Since the range of values for each cell is a priori unknown the estimators $\widehat{p}_{ijk}$ are finally adjusted to sum up to one for each cell ij. For the Monte Carlo approach in the E-Step, each value z_{ij} is generated in each round through the algorithm applied in [PGVST$^+$15], using in each cell generation the discrete distribution with parameters $(\widehat{p}_{ij0}, \widehat{p}_{ij1}, \widehat{p}_{ij2}, ...)$ truncated by marginal limitations.

The initialization of $(\widehat{p}_{ij0}, \widehat{p}_{ij1}, \widehat{p}_{ij2}, ...)$ in the EM algorithm in this version is set as in the base algorithm (uniform distribution).

5.2 Disclosure Relationships on Email Data

Data obtained from the Computation Center of the Universidad Complutense of Madrid is used as a basis to study the performance of the method. Time of sending, sender to receiver (both anonymized) for each message are obtained for 12 months, in 32 Faculty subdomains. Messages that evidently are sent to lists, institutional messages and messages that come from out of the subdomain or that are sent out of the subdomain are deleted. E-mail data patterns are very specic. This is a very sparse data, and true A adjacency matrix for each faculty ranges between 90% - 96% zero cells (not communication between pairs). User's activity has high variance, ranking from about 10 messages to 2500. The number of different receivers for each user is also disperse, from 1 to 40. These numbers affect the detection of communications since active users are more likely to be detected. Figure 5.3 shows the variability between faculty subdomains in terms of senders

and receivers.

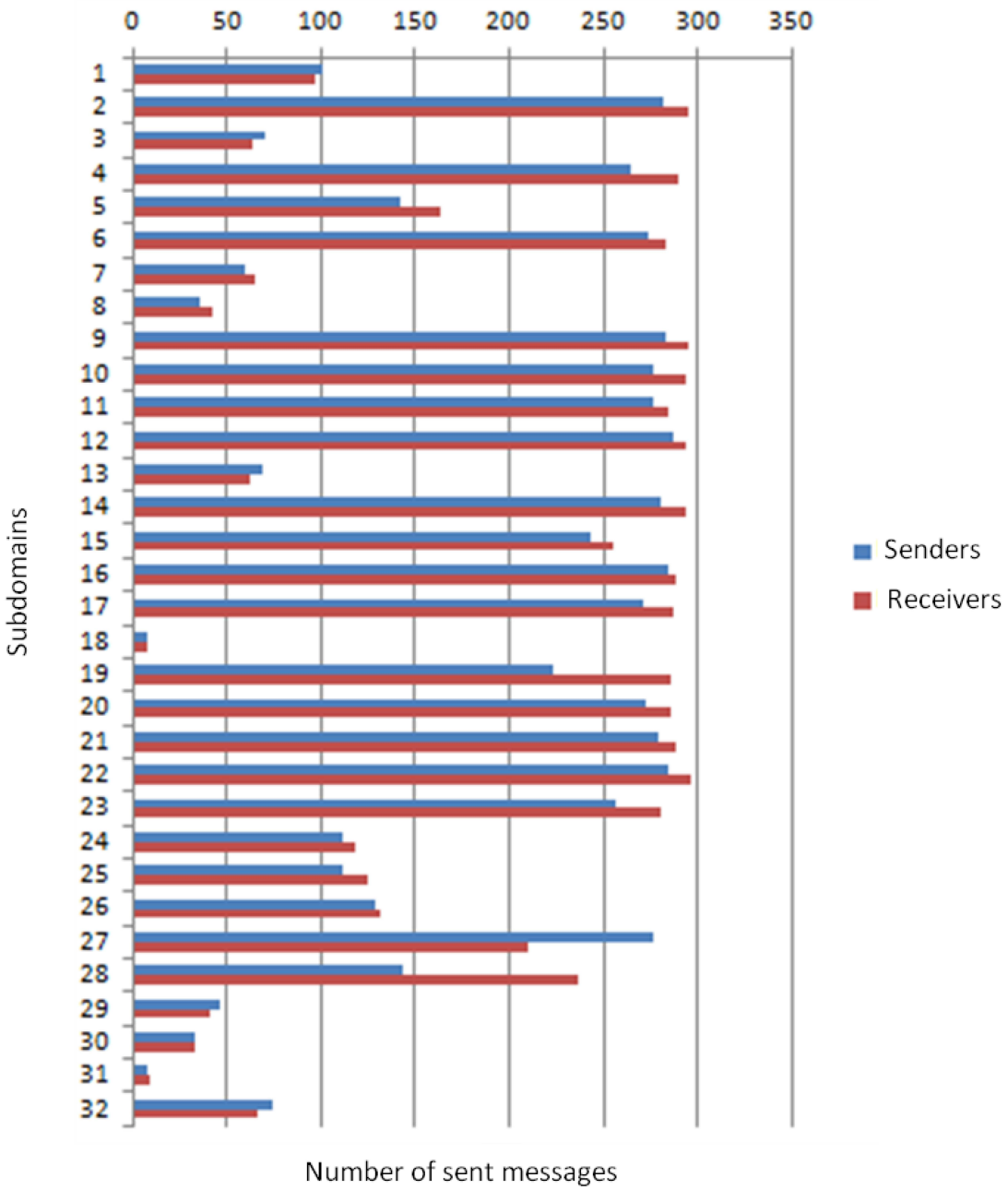


Figure 5.3: Number of senders and receivers in different faculty subdomains

The classification algorithm is initially applied to 10 faculties in order to study its performance under the three forms presented:

a. The original form in [PGVST$^+$15], that is, obtaining feasible tables in an uniform setting, trying to explore the space of feasible tables giving equal weight to all the tables.

b. The application of the EM algorithm under the Poisson model approach.

c. The application of the EM algorithm under the discrete tabulated distribution model approach.

Given that the method is computationally demanding, the EM algorithm is realized only for 5 iterations since it has been observed there is no further improvement. It has also been developed with different batch sizes. As can be seen in Table 5.1, results show that the simple discrete tabulated distribution outperforms the base algorithm and the Poisson modeling approach. Classification rate is the percent of cells i j of the aggregated matrix A that are correctly classified as 0 (not communication) or 1 (communication).

Table 5.1: Classification rate after 5 iterations for the three forms of the algorithm and different batch size, for 4 faculties

Faculty	Batch size	Basic method (Uniform)	EM Poisson	EM discrete
1	7	0.997	0.989	0.997
1	15	0.984	0.983	0.986
1	20	0.976	0.977	0.980
2	7	0.985	0.984	0.99
2	15	0.976	0.977	0.981
2	20	0.965	0.966	0.974
3	7	0.975	0.976	0.98
3	15	0.902	0.91	0.92
3	20	0.89	0.88	0.91
4	7	0.991	0.991	0.991
4	15	0.985	0.986	0.988
4	20	0.972	0.974	0.977

Batch size and complexity of data in terms of percent of zero cells determine the performance of the attack. For the low batch sizes presented in Table 5.1, classification rate is high, since many trivial solutions are detected besides the use of the algorithm to detect communications.

In Figure 5.4 the algorithm is applied in the EM-discrete tabulated form to all faculties for different batch sizes over the 12 months horizon. As batch size increases, performance rapidly decreases. For batch sizes over 100 classification rate is often lower than 80% (not shown in the figure).

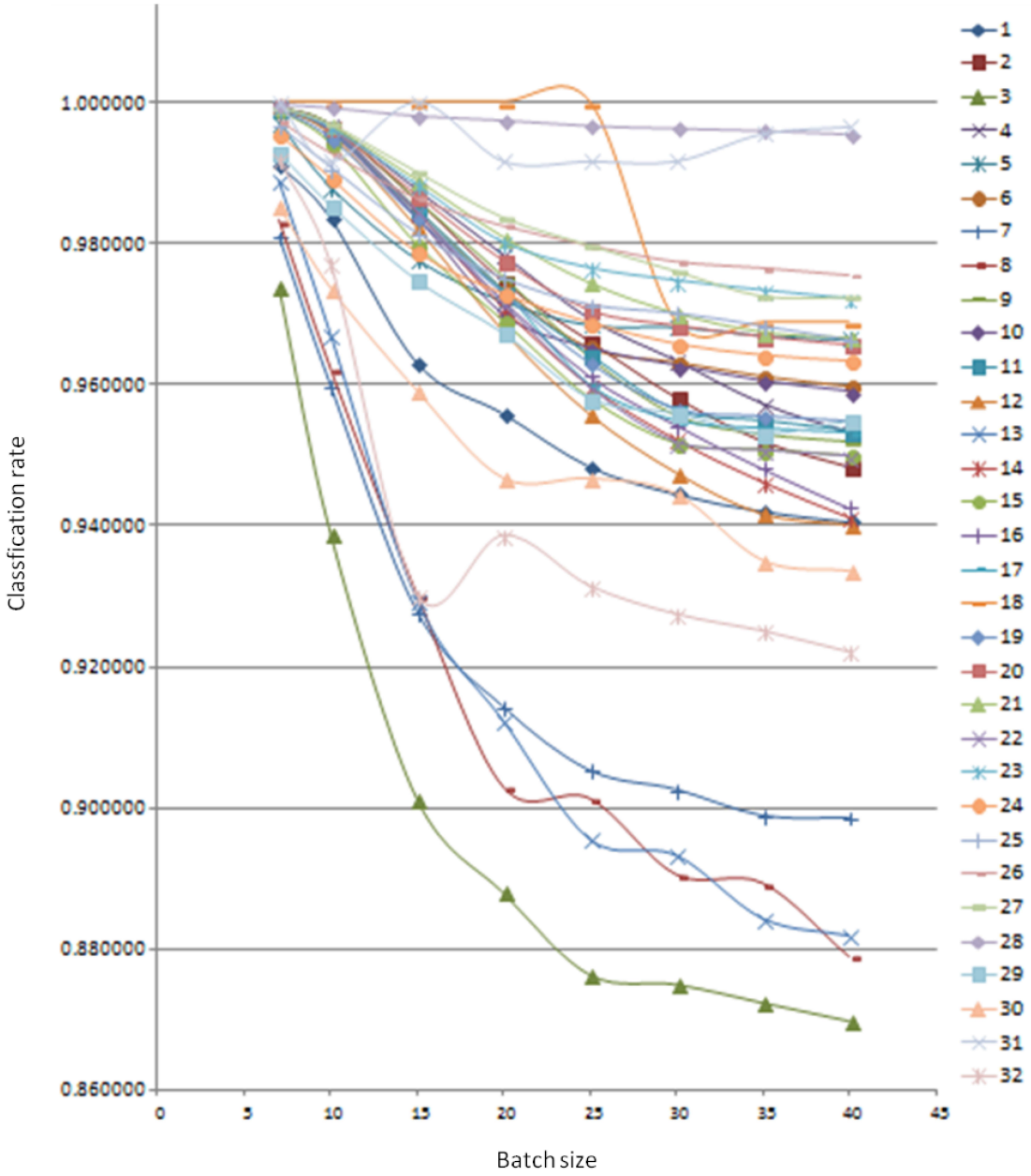


Figure 5.4: Classification rate for all the faculty domains, and different batches sizes

For the method presented here, conservative cut points for the classification based on the cells ordering are used. This leads to results with high positive predictive value, and a somewhat lower negative predictive value (many false negatives). That is, when the algorithm classifies a pair ij as "communicating pair", it is very accurate. When the cell is classified as "not communicating pair", it is less accurate. Table 5.2 presents the True and False positives and negatives, positive predictive value ($TP/(TP+FP)$) and negative predictive value $TN/(TN+FN)$) for some faculties and batch 20. The drawback of the

algorithm is that it does not detect a high percentage of the true communications (low sensitivity). If the aim of the attacker is simply to capture as many communicating pairs as possible with high reliability, the algorithm presented here is very appropriate.

Table 5.2: Rates for different faculties after 5 iterations of the EM algorithm with discrete distribution, batch 20

Faculty	TP	FP	TN	FN	PPV	NPV	Sensitivity
1	264	0	6818	143	1	0.979	0.648
2	1259	1	88831	510	0.999	0.994	0.711
3	231	1	4088	304	0.995	0.924	0.43
4	973	0	89177	451	1	0.994	0.68
5	415	0	28322	504	1	0.98	0.45

5.3 Comparative of Least Squared Method

Most papers on statistical disclosure attacks base their performance on restriction and assumptions that make the comparisons with our method difficult, since this is a general, unrestricted method. Some of the assumptions that can be find in the research literature are:

- The number of receivers for a particular user is known.
- The distribution of messages from an user to another is uniform
- Mix threshold parameter or other system parameters are known.

An expectation to this rule is [PGT12] where they address the problem in a global way, similar to our settings. In this paper the authors develop a method to estimate, given a message sent to a receiver, the conditional probabilities that this message comes from each one of different potential senders. Table 5.3 shows an example of the notation.

Where p_{ij} represent the conditional probabilities of, given a message sent to receiver j, this message has been sent by sender i.

Consequently, there is the restriction

$$\sum_i p_{ij} = 1 \quad \forall \quad j$$

Table 5.3: Example of the notation

sender	**1**	**2**	**3**	**4**
1	p_{11}	p_{12}	p_{13}	p_{14}
2	p_{21}	p_{22}	p_{23}	p_{24}
3	p_{31}	p_{32}	p_{33}	p_{34}
4	p_{41}	p_{42}	p_{43}	p_{44}

In the mentioned article the authors propose a solution based on the least squares method to estimate p_{ij}. Noting by x^k the marginals vector (total messages sent) related to all the senders in round k, that is,

$$x^k = (x_1^k, \ldots, x_N^k)$$

the matrix U is formed by the combined information of all the rounds retrieved:
$U = [x^1, x^2, \ldots x^n]$
And , noting y_j as the receiver j marginals vector in each round:

$$y_j = (y_j^1, \ldots, y_j^k)$$

The authors propose a solution based on the least squares method to estimate p_{ij}:

$$\hat{p}_j = (U^t U)^{-1} U^t y_j$$

where $\hat{p}_j$ is the conditional probabilities $\hat{p}_j = (\hat{p}_{1j}, \ldots, \hat{p}_{Nj})$

This is equivalente to set no intercept regression model for each receiver j:

$y_j = p_j X = p_{1j} x_1 + ... + p_{1j} x_N$

where X represents the marginals vector (total messages sent by each sender) and where each round represents one observation.

These settings have several drawbacks:

1. The least squares solution proposed in the article can lead to negative values for the estimated p_{ij}. These estimations are incorrect since there should be $0 \leq p_{ij} \geq 1$. Also, the sum of column estimators must be 1 and this condition does frequently

not hold .These possibilities make impractical the first proposal matricial and it is necessary to add to the regression modeling problem the mentioned restrictions over the parameters. For the model with positivity restrictions the procedure nlin of SAS package is usedOnce estimates are obtained these are rescaled to sum up to 1.

2. The proposal is limited in the sense that estimators are obtained independently for each receiver j. The interdependence of marginals in each round is therefore not taken into account. This makes comparisons favorable to our method that uses this information.

Our method allows to estimate the mean number of messages per round for each cell (sender, receiver) and also a hard adjacency classification for each cell in 1 if there exists communication or 0 if there is not communication. In order to establish a comparison with the least squares method this is necessary to estimate first the mean of messages per round. This is accomplished by using

$$\hat{\lambda_{ij}} = \sum_{r=1}^{\text{Nrounds}} Y_{ij}$$

Where ${y_j}^r$ is the marginal value (total number of messages received by j in round r).

Besides that, in order to obtain a classification of cells, the estimated percent of cells p_0 used in our algorithm is applied to classify as zero that cells that are below that percent in the table derived from ordering the cells by λ_{ij}.

In order to study the differences in performance of both methods, faculty 8 data, from our email database are used.

The data retrieved by attacker have the following characteristics:

- 28 senders.
- 40 receivers.
- 41 total users.
- 103 rounds in batches of 8 messages.

In order to compare λ_{ij} estimators, quadratic error distance $d = \sum_i \sum_j (\lambda_{ij} - \lambda_{ji})^2$ is used, where λ_{ij} is the true value of the mean number of messages sent from user i to user j per round and $(\hat{\lambda_{ij}})$ its estimated value, obtained by each of the two models. In this data

$d_{algorithm} = 0.75$ and $d_{least-squares} = 7.46$ giving a significative difference in the estimates, with our method having lower estimation error.

In Figure 5.5 a plot that represents the estimated value $(\hat{\lambda_{ij}})$ with respect the true value λ_{ij}, in both models is presented.

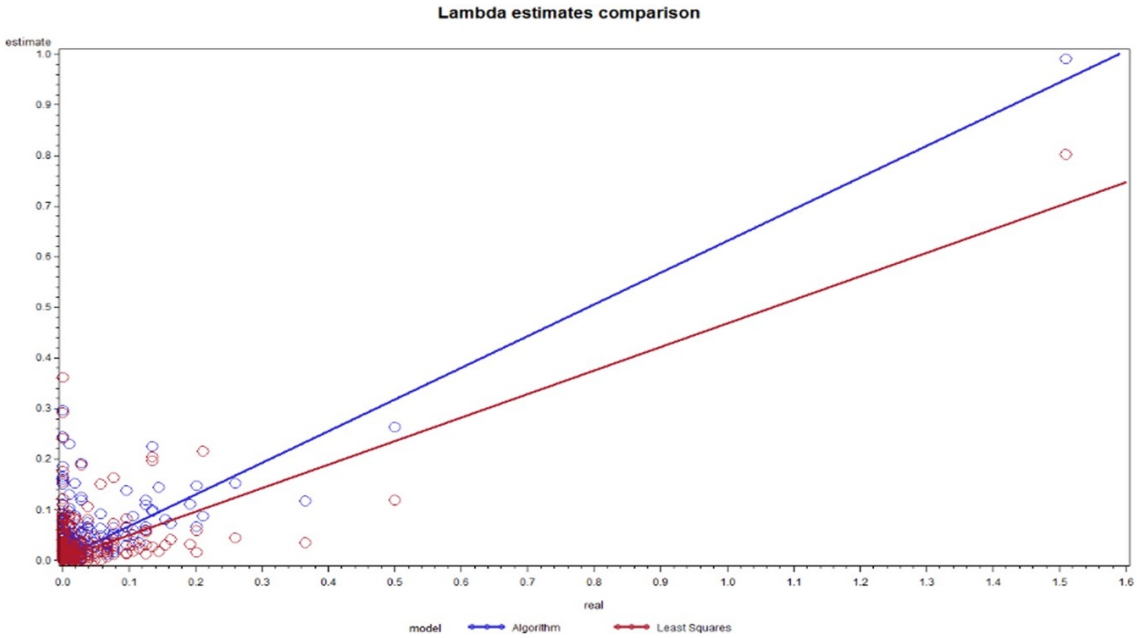


Figure 5.5: Comparative of lambda estimates

There is a linear relationship between true values and estimated values in both methods. The regression R^2 from our model is 0.6, whereas the regression R^2 from the least squares model is 0.45, showing a better fit for our model.

In second place, a cells classification, where cells are set to 1 if a pair of users communicate and 0 if the y do not communicate, is applied with both methods. The estimated percent of zeros to apply as cut point is 86%. Results are showed in Table 5.4.

Table 5.4: Comparative results

	Classification rate	**Sensitivity**	**Specificity**
Algorithm	0.122	0.55	0.93
Least Squares	0.145	0.47	0.91

5.4 Summary

At first, we have shown an improvement of our previously presented statistical disclosure attack. It has been presented the framework and assumptions, as well as the description of the method using the EM algorithm. The method was applied to simulated data with good results, and the refinement presented here is applied on real email data. Additionally, we have presented a comparative between two algorithms whose aim is to recover the communication patterns of users anonymously communicating through a threshold mix. It has been shown our statistical disclosure attacks get better results.

Chapter 6

Application of the Estimation of Features Users Network Email or Social Networks

This chapter deals with the results obtained to estimate the features of network users and network email or social networks users. First, we mainly discuss the properties and measures in social networks in Section 6.1. Then, the results obtained of applying our method are provided in Section 6.2. The chapter ends in Section 6.3 with a brief summary of the above in it.

6.1 Properties and Measures of Social Networks

A social network is a social structure made of individuals, which are connected by one or more types of relationships. Its representation can be made through a graph where the vertices represent individuals or entities and the edges the relations among them. Formally a simple social network is modeled as a graph $G = (V, E)$ where:

- $V = (v_1, \ldots, v_n)$ is the set of vertices o nodes que represented as entities or individuals.
- E is the set of social relationships, represented as edges in the graph, where $E = \{(v_i, v_j) | v_i, v_j \in V\}$

One way to categorize the networks is based on its nature and the number of sets of actors in it. Thus, one can distinguish different modes of networks. The classification is as follows:

a. Unimodal networks. These networks are the most common where structural variables are for one set, it means where all actors come from a single set. Examples friendly relationships.

b. Bimodal networks. Contains structural variables measured for two sets of entities. For example, we can analyze actors from two different sets, one made up of companies and other formed by civil organizations. These networks are also known as affiliate networks.

c. Mode N networks. In this mode three or more sets of social entities are studied. It become complex given the number of actors, the system of relationships and analytical methods for study involved.

In literature exist three levels of analysis within the Social Network Analysis [KS08] [SG92] [WF94]: i) analysis of egocentric networks; ii) analysis focused on subgroups of actors; iii) analysis focused on the overall structure of the network.

The objective of the analysis of egocentric networks is to study how a behavior actor evolves, taking into account that is focus solely on that actor and his relationships with the rest of the participants The second type of analysis allow to understand the logical of networks clustering and the existence of cooperation and competition patterns, which are adapted or maintained over time. Finally, in the analysis of overall structure of the network are considered the morphological characteristics adopted, the existence, role and subgroups interaction, the distribution of relationships between actors involved, the geodesic distance between actors, among others. According to the type of problem to solve some of the three levels of analysis is chosen.

From the structural characteristics, the Social Network Analysis is based on developing a matrix representing the relations between users and the construction of a corresponding graph.

According to [SG92] the properties of social networks can be classified into two types, relational and structural properties.

- Relational Properties. This type is based on the relationships within the network and is focused between two elements: a) transactions within the system, the information flows within it, its directionality and density; b) the nature of relationships. Relational properties deal with the content of relations.

- Structural Properties. These describe the way members fit together to form social networks, and can be divided into three levels of analysis: individual members, sub-groups, and total networks. Measures of individual members describe differences among their connections to other members of the network.

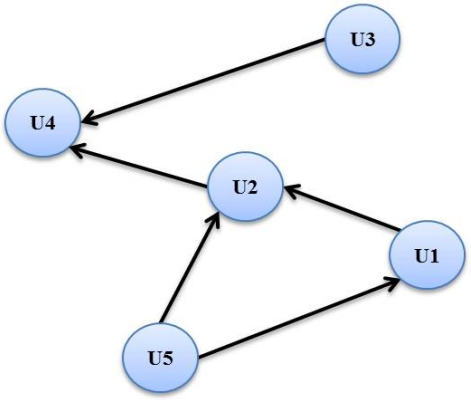
In graph theory, actors can be called nodes or vertices; and edges are called relational links, lines or arches (on directed networks). A graph is an appropriate representation of a social network. However, in some cases the relationship between two pairs of actors are rated or labeled, it is also known as weighted graph, where the strength and intensity of the edges is recorded. Therefore, a valued graph is a graph where each line has a value.

There are two types of graphs:

- Directed graph, where the relationship between two actors is not bidirectional, it means, a node has a relationship with another, but it does not mean that there is an inverse relationship. These graphs also are called digraphs.

- Undirected graph, where the relationship between two nodes is reciprocal. For graph representation all depends on the characteristics of the graph and which technique is wanted to use to manipulate it. One of the simplest structures is the matrix. It is common to use an adjacency matrix M for a graph representation of n^2 size, where n is the number of nodes. If there is an edge between node i and node j, 1 is placed in the cell (i, j) and 0 otherwise. In Figure 6.1 shows the example of an adjacency matrix of a directed graph composed of 5 nodes.

The graph can also be classified according to various topological measures. For example, in order to analyze social networks is important to know if it is possible to reach a node through another node. In that case, it may be interesting how many ways exists to get that node and what is the optimal way.



User	U1	U2	U3	U4	U5	Total
U1	0	1	0	0	0	1
U2	0	0	0	1	0	1
U3	0	0	0	1	0	1
U4	0	0	0	0	0	0
U5	1	1	0	0	0	2
Total	1	2	0	2	0	5

Figure 6.1: Example of adjacency matrix of a directed graph

Before describing the measures it is important to know the paths concept. For studying social networks is important to know if it is possible to reach a node through another node. In this case, it is interesting to know how many ways exists and which one is the best. To calculate the distance between two nodes, the paths are used.

Path is a series of nodes and different lines. The path length is the number of lines in it. The first node is called the origin and the final destination. A shortest path between two nodes is the minimal length path of all the possible paths between nodes.

One of the most common paths is called geodesic path, which is the shortest path between two nodes. The length of a geodesic path is called geodesic distance and is denoted as $d(i, j)$, which is the distance between the nodes n_i and n_j. Both directed and undirected graphs, the geodesic distance is the number of relationships in the shortest possible path from one actor to another.

Distances are important in network analysis; they are mainly used in some of the centrality measures below. One of the main uses of graph theory in social networks is to identify the most important nodes. To calculate the importance of a node centrality network measures are used, in cluster strongly connected, in positions that are structurally equivalent, or the existence of unique positions [VA05]. At network level there are cohesion measures that allow comparison of the whole network structure such as density, diameter and transitivity (also known as clustering coefficient). Below we detail the characteristics of each of the measures.

A level node there are centrality measures such as the node degree, nodal transitivity degree, betweenness and closeness. Measures related to the whole network are for example density, distribution degree, clustering coefficient, diameter among others. Next we

describe the most important measures in social network analysis.

- **Degree**: The centrality degree of a node is the number of users or nodes that are directly related to it. Two nodes of a graph are adjacent or neighbors if there is a branch that connects them. In the case of directed graphs there are two types of degrees:

  1. The degree of an input node $d_I(n_i)$ is the number of arcs that end in it. An entry degree is the sum of the arcs of the way $l_k =< n_j, n_i >$ for all $l_k \in L$ and $nj \in N$.
  2. The output degree of a node $d_O(n_i)$ is the number of arcs that originate from it. A output degree is the sum of the arcs of the way $l_k =< n_j, n_i >$ for all $l_k \in L$ and $n_j \in N$.

  Therefore the overall degree of such graphs is the sum of both. It is said that a degree is regular if all nodes have the same degree.

- **Nodal transitivity or clustering coefficient**: It is a metric that calculates the level of interconnection of a node with its neighbors.

- **Closeness**: The degree of closeness is the ability of a node to reach all others in the network, it is calculated by counting the geodesic distances from one actor to the others. An actor is important if it is close to all others.

  To calculate the closeness of a node n_i, the following formula is used:

$$C_c(n_i) = \left[\sum_{j=1}^{g} d\left(n_i, n_j\right)\right]^{-1}$$

  where: $d(n_i, n_j)$ is the minimum distance between node i and node j.

- **Betweeness**: In order to calculate the importance of a node in a network by Betweeness, it has to be measured the intermediation of this node with the rest, that is, the possibility for an actor to intervene between two nodes. The Betweeness of node n_i is the frequency of n_i appears on the shortest paths (geodesic) between two nodes. An actor with a high degree of Betweeness means it is an important node for the network, because he can control the flow of its communication.

A node must have at least a degree of input and output, to have a value of Betweeness on the network, besides being in the geodesic paths of two nodes. Calculating Betweeness:

$$C_B(n_i) = \sum j < k \frac{g_{jk}(n_i)}{g_{jk}}$$

where:

g_{jk} = the number of geodesic paths between nodes jk. $g_{ik}(n_j)$ = the number of connections in which the node i is at the geodesic path between jk.

- **Density**: It is the percentage of the number of relationships and the number of possible relationships. Networks with high density respond differently to the challenges of those with low density.

- **Diameter**: The diameter of a network is the longest existing network geodesic distance. This is a useful measure that can be used to determine the size of the entire network.

- **Transitivity**: The clustering coefficient is calculated network by Watts and Strogatz [WS98] as the average clustering coefficient of all vertices of the network.

- **Distribution degrees**: A network can be an extremely complex structure, since the connections between nodes may have complicated patterns. One challenge at studying complex networks is to develop simple metrics that capture the structural elements in an understandable form. One such simplification is to ignore any pattern between different nodes, and observe each node separately.

- **Undirected networks**: The degree of a node i is the number of its connections. In terms of a adjacency matrix A, the degree of a node i is only part of the row i of A.

$$k_i = \sum_j a_{ij}$$

where the sum is over all nodes in the network. By counting the number of nodes that each degree, it can be established the grade distribution $P_{\deg}(k)$ defined as:

$P_{\deg}(k)$ = percentage of nodes in the graph with degree k.

An example of the distribution of degrees of an undirected graph shown in Figure 6.2, degrees are $k_1 = 1$, $k_2 = 3$, $k_3 = 1$, $k_4 = 1$, $k_5 = 2$, $k_6 = 5$, $k_7 = 3$, $k_8 = 3$, $k_9 = 2$ y $k_{10} = 1$. The grade distribution is $P_{\text{deg}}(1) = 4/10$, $P_{\text{deg}}(2) = 2/10$, $P_{\text{deg}}(3) = 3/10$, $P_{\text{deg}}(5) = 1/10$.

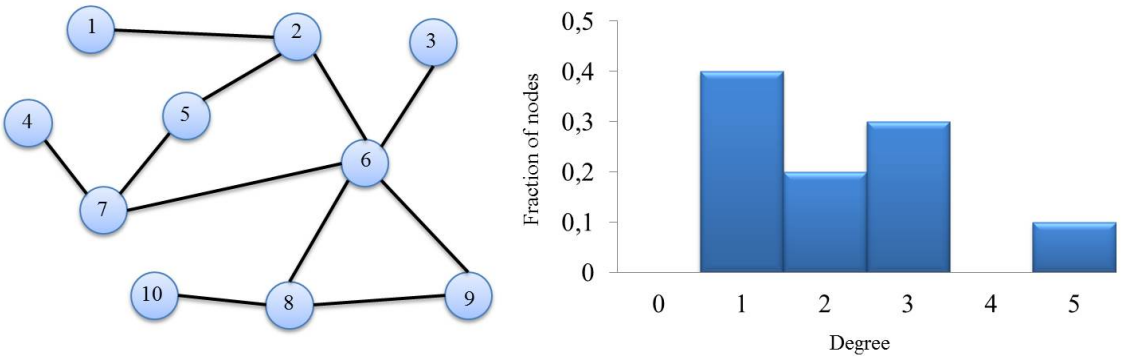


Figure 6.2: Distribution degrees example

The distribution of degrees clearly captures only a small amount of network information. But this information still gives important clues within the structure of a network. For example, in the simplest types of networks, it is common to find that most nodes in the network have similar degrees. In real-world networks usually they have very different degrees distribution. In such networks, most nodes have a relatively small degree, but there are few nodes with a very high degree. The nodes with highest degrees are known as hubs.

- **Directed networks**: The degree distribution of directed networks is a bit more complex than undirected networks; because, the node degree in a directed network cannot be captured for a single number. If we focus on one node of a directed network, we will see some edges that enter the node and other coming out of it. At ignoring the direction of the edges and simply add up the total number of edges, a lot of valuable information will be lost. An input edge and an output edge can mean very different things and in some cases it would be important to maintain this distinction.

There are several works in the literature that suggest real-world social networks have very specific characteristics. Complex networks as www or social networks do not have an organized architecture, but rather have been promoted organized themselves according to the actions of many individuals. From these interactions global phenomenon, can

emerge for example, properties of small world or free scale distribution. These two global properties have considerable implications for the behavior of the network under attack, as well as the dissemination of information or epidemiological issues. In late 1950, Erdos and Renyi [ER59] marked a precedent in classical mathematical theory to model problems of complex networks describing a network using a random graph, defining the foundations of the theory of random networks.

Networks composed of people connected through the exchange of emails exhibit characteristics of small world networks and scale-free networks.

Almost every real-world networks follow a power law. In [BR99] the term "scale-free network" that describes the kind of networks that exhibit a power-law distribution is introduced. The characteristic of such networks is that the distribution of links results in a straight line if plotted on a logarithmic scale twice. The power law is a member of the family of distributions skewed toward the extremes, so describing events in which a random variable reaches high values infrequently, while medium or low values are much more common. Seen from another angle, the power law probability of occurrence of small events is relatively high, while the probability of occurrence of large events is relatively low.

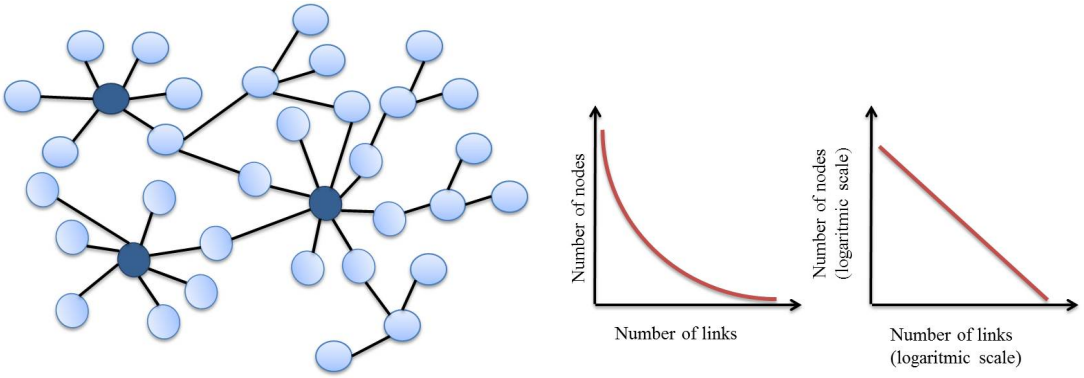


Figure 6.3: Power law distribution

In [MOT12] are analyzed the structural properties of email networks, the results concluded that traffic from a legitimate email system results in small-world networks and scale-free. On the other hand, it is also argued that considering an email system as a single whole, does not display a scale-free behavior completely antisocial behavior as spam.

Inspired on social networks, in 1998 a simple network model was proposed, it was called "small world" [WS98]. In Figure 6.4 we show on the left, a regular network built with value $p = 0$, and the right a random network with value $p = 1$. The p value indicates the probability that any node redirects a connection to any other network node randomly.

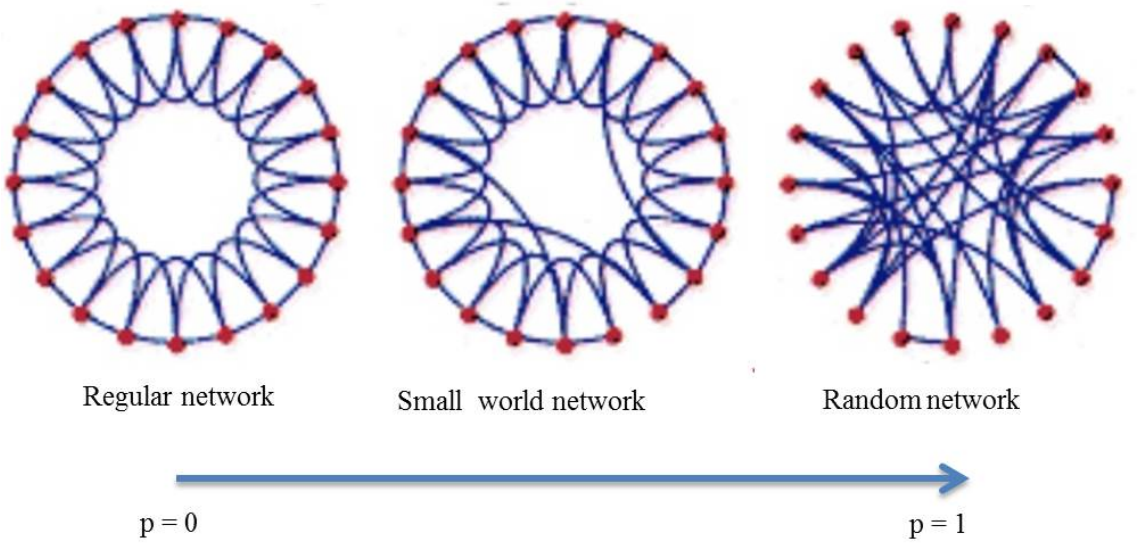


Figure 6.4: Networks topology

The name "Small World" comes from the phenomenon in which two strangers find they have a mutual friend. Human social networks typically exhibit the characteristic that in any set of friends, each friend is also connected to other groups of friends; there is a high probability that two neighbors of a node are connected among themselves, and a small average length the shortest path between two nodes. Networks whose characteristics are of a high clustering coefficient and low average shortest paths are known as small-world networks. In late 1960, an experiment stated that between two people there are on average 6 connections friendship, no matter where in the world be developed. This hypothesis is known as "six degrees of separation" [Mil67].

In [EMB02] first consider the study of the structure of emails networks considering the log files of a university. Considering the network topologies of email address where emails are nodes and edges are the communications among them. The resulting network also shows a distribution of links or relationships with pronounced free scale and small-world behavior. In [LKF07] the evolution of various types of real networks are shown. In other hand, other works utilize communication patterns in the dataset Enron email to: detect social tensions [FMMN09]; discover structures within the organization [CKY05]; identify

the most relevant actors in the network over time [UMH10].

A more detailed work studied more than 100 real-world networks to reveal their groups or communities, the authors note that large networks have a very different structure compared to the small-world networks [LLDM09]. And there is an inverse relationship between the size of the community and the high quality of the community. The largest networks of 100 nodes do not show good conductivity which can be translated as not having the ability to be a good community; the best communities are quite small, in the range of 10 to 100 nodes.

6.2 Application of the Method to Estimate Characteristics of Network Users and a Network Email

It has been shown that an attacker can reveal the identities of mix users by analyzing network traffic, watching the flow of incoming and outgoing messages. In the literature there are researches where an attacker can get partial information to study an anonymous social network, taking into account the vulnerability to attacks capture path [TWH05] [CBM04]. Such attacks using the vulnerability of the network traffic to compromise the identity of users to compromise the network.

We have applied our algorithm to data provided by the Computer Centre of the Complutense University of Madrid who were previously anonymous. Such information is divided into 32 sub domains or faculties that make up the email system. For demonstration purposes we have chosen only the Faculty A.

In Table 6.1 we present the results obtained after applying our algorithm to Faculty A of 3 month data and Table 6.2 for 12 months. We can see that the estimated batches of messages smaller values are closer to the actual values of the network.

Table 6.1: Results of Faculty A for 3 months observations

	Batch	Nodes	Edges	Average degree	Density	Clustering Coefficient
	10	85	406	4.776	0.057	0.335
Estimate	30	85	406	4.776	0.057	0.335
	50	85	403	4.741	0.056	0.334
Real	-	85	406	4.776	0.057	0.335

Table 6.2: Results of Faculty A for 12 months observations

	Batch	Nodes	Edges	Average degree	Density	Clustering Coefficient
	10	116	929	8.009	0.070	0.482
Estimate	30	116	923	7.957	0.069	0.490
	50	116	924	7.966	0.069	0.479
Real	-	116	929	8.009	0.070	0.482

Figure 6.5 shows the estimated and real graph of Faculty A composed of 85 users, with a time horizon of three months. Figure 6.6 shows the results for a 12 month period. Because the differences are lower, we have placed the two overlapping graphs for 3 and 12 months, where green edges correspond to relations that our algorithm has not detected. We also note that both networks exhibit small world and scale-free characteristics.

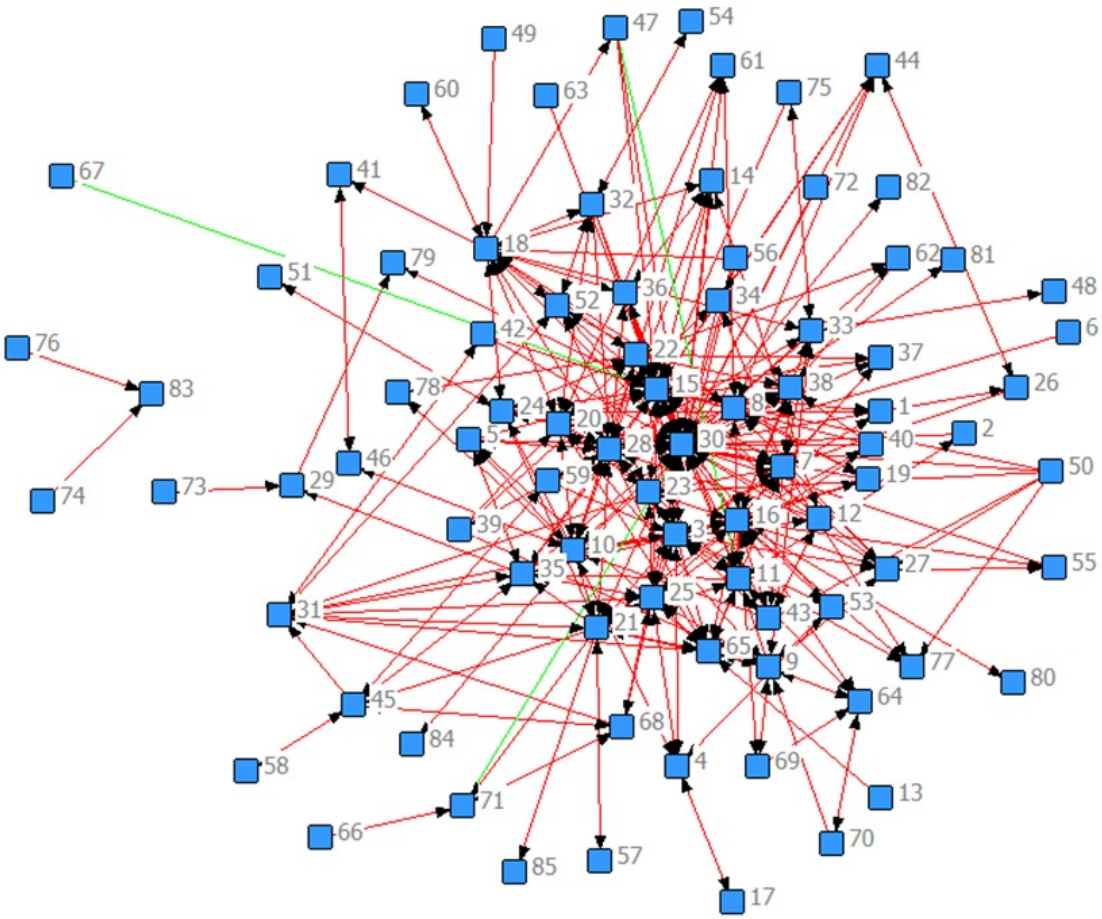


Figure 6.5: Simulated vs. Real graph of Faculty A for 3 months

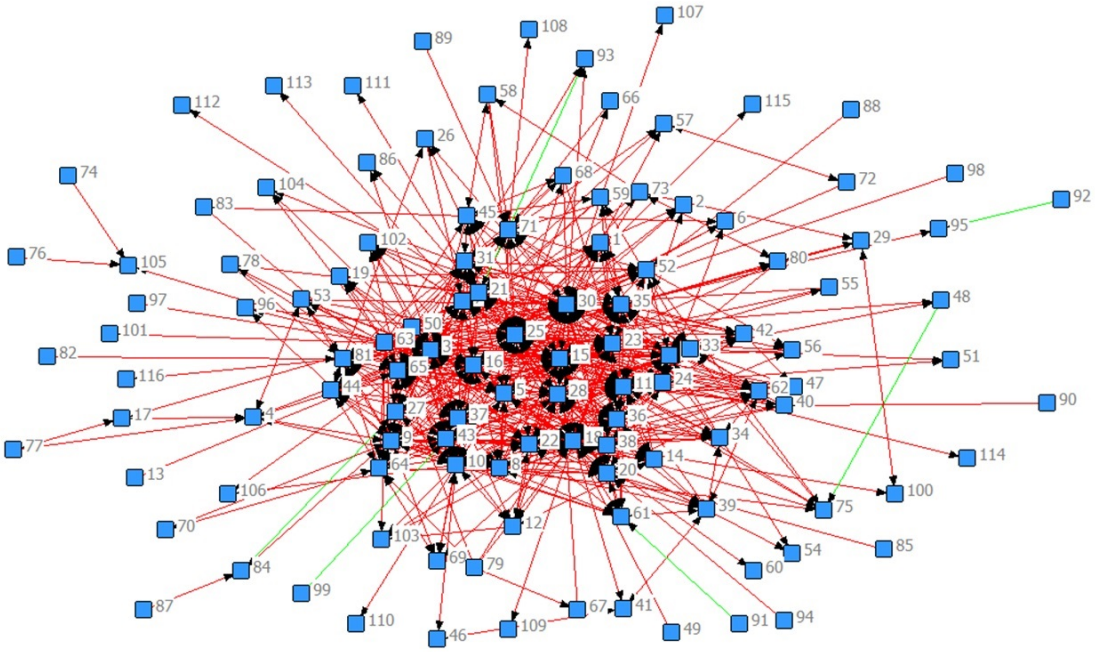


Figure 6.6: Simulated vs. Real graph of Faculty A for 12 months

Table 6.3 presents the five highest degree of centrality calculated for each network estimated with different batch (10, 15, 20 and 30), the last column corresponds to the real network, on the other hand, in Table 6.4 we show the five lowest centrality degrees.

Table 6.3: Five highest degree centrality nodes of Faculty A

Batch 10	Batch 30	Batch 50	Real
0.286	0.286	0.286	0.286
0.214	0.214	0.214	0.214
0.190	0.190	0.190	0.190
0.167	0.167	0.167	0.167
0.167	0.167	0.167	0.167

Table 6.4: Five lowest degree centrality nodes of Faculty A

Batch 10	Batch 30	Batch 50	Real
0.286	0.286	0.286	0.286
0.214	0.214	0.214	0.214
0.190	0.190	0.190	0.190
0.167	0.167	0.167	0.167
0.167	0.167	0.167	0.167

In Figure 6.7 we present the comparison of estimated and actual degrees of the Faculty A for 3 to 12 months; the closer to the diagonal point is better estimate. Otherwise, the points are above or below the diagonal.

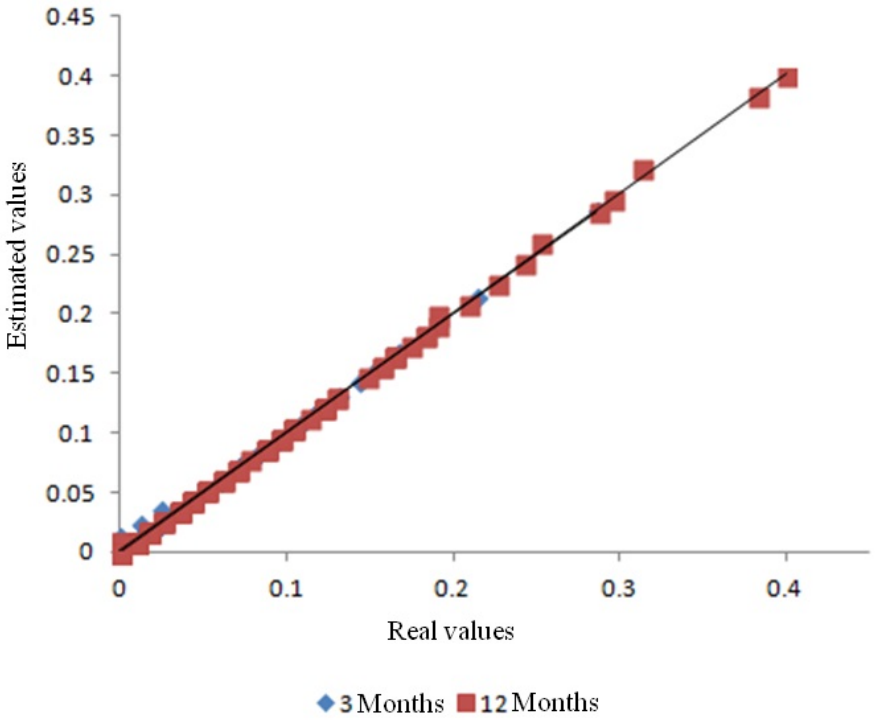


Figure 6.7: Estimated vs. Real centrality degrees for 3 and 12 months

6.3 Summary

In this chapter we described the characteristics and metrics of social networks. Using social network analysis techniques and getting several social network measures we were able to know user's centrality to detect which are the most influential users in a network. We have applied a probabilistic attack disclosure of identities on an university anonymous email system, representing such system as a social network. We showed that analysis of social networks helps to get the metrics that provide insight into the centrality of the users involved to detect the most important elements which control the flow of information. From the results we found the attack is better with small batches.

Chapter 7

Conclusions and Future Works

Anonymity in network private communications is a real concern present in current research. Disclosing if there exists communication or not between a pair of users in a network communication system is the object of the attacker in the present work. The disclosure, or revealing the links between users is a topic already treated in the research on PETS. Statistical Disclosure Attacks are claimed to be efficient on obtaining this important information, but generally the assumptions assumed on the different strategies proposed in previous research limit the scope of the attacks.

The method presented here leads to results in different dimensions: estimation of the number of messages sent by round or unit of time for each pair sender-receiver, ordering of the pairs from highest likelihood of communication to lowest, hard classification of pairs of users in communication-not communication. Another important result derived is the occasional detection of some pairs that have certainly communicate (without any doubt, based on combinatorial deduction) and the detection of some pairs that did never communicate in the time horizon of the attack. Besides, the estimation of different feasible tables for each round can serve as a measure of complexity of the problem when dealing with real data. Then the attack can be seen as a multipurpose or multiobjective attack, reducing or projecting the round basic information to another set of information that can be used in other contexts.

The attack is first studied with simulated data, but is soon applied over real email data in order to imitate the circumstances an attacker would face. The data can be very different from the simulation settings observed in previous work. For example, in our email data limited to faculty domains, a high percentage, near 85% of pairs of users, do

not communicate. This affects seriously precedent algorithms in the literature, leading to specification or computational problems. Our work addresses these problems naturally and also takes advantage of them.

A second version of the procedure, including a second pass on the data through the use of the EM algorithm, improves significantly the estimation and classification results. More passes on the data looking to reach convergence are not useful and can even lead to erratic results. There is certainly space for improvement in this line.

The comparison with the method of least squares presented on (IEEE) leads to very good results favoring our method. This is encouraging since there isn't other research that is clearly comparable to ours in terms of the general previous hypotheses.

The framework for the attack scopes communication networks protected by mixes and, as it has been proved, it can be used to estimate user centrality characteristics and more global network parameters. However, the scope of the method does not limit to internet or local network communications. The schema can be abstracted to other contexts; for example, repeated polls or elections in small populations, where the attack can be used to obtain an ordering of the likelihood users vote to some political groups or anti terrorist research, where the method can use phone calls information in repeated contexts to link senders with recipients.

The strategy proposed here has the advantage of addressing a general framework; however, it is limited by the scale of the information retrieved by the attacker. Since the information obtained consists only on the number of messages sent and received by the users in each round, the size of the rounds (batch size) is an important parameter that affects seriously the results. In general, for small to moderate batch sizes the missclassification rates can be low, but the combinatorial possibilities derived of high batch sizes can lead quickly to bad performances. Other parameters that affect the performance of the method include the number of users and complexity of relationships or number of users that really communicate

7.1 Future Works

Future Work includes the following ideas:

- **Extend to other anonymity protocols**: The attack is supposed to act over single, pool or threshold mixes; there is a need to explore if it could be used on stronger

protections such as Tor (Onion Routing).

- **Applications in other contexts and data**: Social network data can be used to further investigate the performance of the strategy developed here. There are also other standard databases that could be used as benchmark (Enron email data, for example). Other applications in the field of disclosure of public data could be considered.

- **Further comparisons with other methods**: There exist other disclosure attacks to compare with, for example Vida or Perfect matching disclosure attack.

- **Estimation of the percent of zero cells**: The estimation of this percent is still an open problem; it is used to establish a cut point for the hard classification of pair of users. In this work this is made taking into account the feasible tables information, but better estimates could be obtained using previous information in the data context, may be based on Bayesian estimates.

- **EM method improvement**: The fact that two iterations of the EM algorithm improves the performance but further iterations can deteriorate it is still an open question to address.

Bibliography

[AAGW11] Mahdi N. Al-Ameen, Charles Gatz, and Matthew Wright. SDA-2H: Understanding the Value of Background Cover Against Statistical Disclosure. In *Computer and Information Technology (ICCIT), 2011 14th International Conference*, pages 196–201, December 2011.

[ABD+03] Tero Alamäki, Margareta Björksten, Péter Dornbach, Casper Gripenberg, Norbert Györbíró, Gábor Márton, Zoltán Németh, Timo Skyttä, and Mikko Tarkiainen. Privacy Enhancing Service Architectures. In *Proceedings of the 2Nd International Conference on Privacy Enhancing Technologies*, pages 99–109, Berlin, Heidelberg, April 2003.

[Abe06] Masayuki Abe. Universally Verifiable MIX with Verification Work Independent of the Number of MIX Servers. In *Proceedings of the Advances in Cryptology (EUROCRYPT'98)*, pages 437–447, May 2006.

[Ada06] Carlisle Adams. A Classification for Privacy Techniques. *University of Ottawa Law & Technology Journal*, 3(1):35–52, 2006.

[ADS03] Alessandro Acquisti, Roger Dingledine, and Paul Syverson. On the Economics of Anonymity. In *Proceedings of the Financial Cryptography (FC '03)*, pages –, January 2003.

[AG04] Alessandro Acquisti and Jens Grossklags. Privacy and Rationality: Preliminary Evidence from Pilot Data. In *Third Workshop on the Economics of Information Security*, 2004.

[AK03] Dakshi Agrawal and Dogan Kesdogan. Measuring Anonymity: The Disclosure Attack. *IEEE Security and Privacy*, 1(6):27–34, November 2003.

[And96] Ross Anderson. The Eternity Service. In *Proceedings of PRAGOCRYPT 96*, pages 242–252, Prague, Czech Republic, September 1996.

[Ash09] Kevin Ashton. That 'Internet of Things' Thing. *RFiD Journal*, 22(7):97–114, June 2009.

[Ben01] Tonda Benes. The Strong Eternity Service. In *Proceedings of the 4th International Workshop on Information Hiding*, pages 215–229, London, UK, UK, April 2001.

[BFK01] Oliver Berthold, Hannes Federrath, and Stefan Köpsell. Web MIXes: A System for Anonymous and Unobservable Internet Access. In *International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability*, pages 115–129, New York, NY, USA, January 2001.

[BG15] Krista Bennett and Christian Grothoff. Gnunet: Gnu's Decentralized Anonymous and Censorship-Resistant P2P Framework. `http://www.i2p2.de`, 2015.

[BGA01] Adam Back, Ian Goldberg, and Shostack Adam. Freedom Systems 2.1 Security Issues and Analysis. White Paper, Zero Knowledge Systems, Inc., 2001.

[BGS05] Bettina Berendt, Oliver Gunther, and Sarah Spiekermann. Privacy in E-Commerce: Stated Preferences vs. Actual Behavior. *Communications of the ACM*, 48(4):101–106, April 2005.

[BHVY05] Robert C. Brigham, Frank Harary, Elizabeth C. Violin, and Jay Yellen. Perfect-Matching Preclusion. *Congressus Numerantium*, 174:185–192, January 2005.

[BMGS08] Kevin Bauer, Damon McCoy, Dirk Grunwald, and Douglas Sicker. BitBlender: Light-weight anonymity for BitTorrent. In *Proceedings of the Workshop on Applications of Private and Anonymous Communications (AlPACa 2008)*, pages 1–8, New York, NY, USA, September 2008.

[BPS01] Oliver Berthold, Andreas Pfitzmann, and Ronny Standtke. The Disadvantages of Free MIX Routes and How to Overcome Them. In *International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability*, pages 30–45, New York, NY, USA, January 2001.

[BR99] Albert-László Barabási and Albert Réka. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999.

[Bro02] Zach Brown. Cebolla: Pragmatic IP Anonymity. In *Ottawa Linux Symposium*, pages 55–64, Ottawa, Ontario, Canada, June 2002.

[CBM04] Aron Culotta, Ron Bekkerman, and Andrew McCallum. Extracting Social Networks and Contact Information from Email and the Web. In *Proceedings of Conference on Email and Anti-Spam*, pages –, 2004.

[CDHL05] Yuguo Chen, Persi Diaconis, Susan P. Holmes, and Jun S. Liu. Sequential Monte Carlo Methods for Statistical Analysis of Tables. *Journal of the American Statistical Association*, 100(469):109–120, 2005.

[Cen07] Electronic Privacy Information Center. *Privacy in Human Rights Report 2006: An International Survey of Privacy Laws and Developments*. Electronic Privacy Information Center & Privacy International, 2007.

[Cha81] David L. Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *Communications of the ACM*, 24(2):84–90, February 1981.

[Cha98] David Chaum. The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability. *Journal of Cryptology*, 1(1):65–75, January 1998.

[CJ07] Kim Cameron and Michael B. Jones. Design Rationale behind the Identity Metasystem Architecture. In *ISSE/SECURE 2007 Securing Electronic Business Processes*, pages 117–129, 2007.

[CKY05] Anurat Chapanond, MukkaiS. Krishnamoorthy, and BÃ¼lent Yener. Graph Theoretic and Spectral Analysis of Enron Email Data. *Computational & Mathematical Organization Theory*, 11(3):265–281, Octuber 2005.

[CMD09] Emily Christofides, Amy Muise, and Serge Desmarais. Information Disclosure and Control on Facebook: Are They Two Sides of The Same Coin or Two Different Processes? *CyberPsychology and Behavior*, 12(3):341–345, June 2009.

[Com07] European Commision. Privacy Enhancing Technologies (PETs) The existing Legal Framework. `http://europa.eu/rapid/press-release_MEMO-07-159_en.htm?locale=en/`, 2007.

[Con03] Privacy Incorporated Software Agent Consortium. *Handbook of Privacy and Privacy-Enhancing Technologies*. College bescherming persoonsgegevens, 2003.

[Cor14] EMC Corporation. The EMC Privacy Index. `http://www.emc.com/campaign/privacy-index/index.htm`, 2014.

[CSWH01] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W. Hong. Freenet: A Distributed Anonymous Information Storage and Retrieval System. In *International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability*, pages 46–66, New York, NY, USA, January 2001.

[CVH02] Jan Camenisch and Els Van Herreweghen. Design and Implementation of the Idemix Anonymous Credential System. In *Proceedings of the 9th ACM*

Conference on Computer and Communications Security, pages 21–30, New York, NY, USA, November 2002.

[CYS+07] Bogdan Carbunar, Yang Yu, L. Shi, Michael Pearce, and Venu Vasudevan. Query Privacy in Wireless Sensor Networks. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference*, pages 203–214, June 2007.

[DÓ5] Claudia Díaz. Anonymity and Privacy in Electronic Services. PhD. Thesis, Katholieke Universiteit Leuven, 2005.

[DA04] George Danezis and Ross Anderson. The Economics of Censorship Resistance. In *Proceedings of the Workshop on Economics and Information Security (WEIS04)*, May 2004.

[Dan03] George Danezis. Statistical Disclosure Attacks: Traffic Confirmation in Open Environments. In *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*, pages 421–446, May 2003.

[DDM03] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a Type III Anonymous Remailer Protocol. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, pages 2–15, Washington, DC, USA, May 2003.

[DDT07] George Danezis, Claudia Diaz, and Carmela Troncoso. Two-sided Statistical Disclosure Attack. In *Proceedings of the 7th International Conference on Privacy Enhancing Technologies*, pages 30–44, Berlin, Heidelberg, June 2007.

[Del09] Laurie Delmer. L'émergence au Sein d'Internet de Communautés Virtuelles et Anonymes, Freenet et I2P. Master's Thesis, Université catholique de Louvain - Département des sciences politiques et sociales, 2009.

[DFM00] Roger Dingledine, Michael J. Freedman, and David Molnar. The Free Haven Project: Distributed Anonymous Storage Service. In *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 67–95, July 2000.

[DG12] Claudia Diaz and Seda GÃ$\frac{1}{4}$rses. Understanding the Landscape of Privacy Technologies. In *Proceedings of the Information Security Summit*, pages 58–63, November 2012.

[DJR12] Ratan Dey, Zubin Jelveh, and Keith Ross. Facebook Users Have Become Much More Private: A Large-Scale Study. In *Pervasive Computing and*

Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference, pages 346–352, New York, NY, USA, March 2012.

[DLR77] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[DMS04] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The Second-Generation Onion Router. In *Proceedings of the 13th Conference on USENIX Security Symposium*, pages 303–320, Berkeley, CA, USA, August 2004.

[Dou02] John R. Douceur. The Sybil Attack. In *Proceedings of the 1st International Peer To Peer Systems Workshop (IPTPS 2002)*, pages 251–260, London, UK, UK, March 2002.

[DP04] Claudia Díaz and Bart Preneel. Reasoning About the Anonymity Provided by Pool Mixes That Generate Dummy Traffic. In *Proceedings of the 6th International Conference on Information Hiding*, pages 309–325, Berlin, Heidelberg, May 2004.

[DSCP03] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards Measuring Anonymity. In *Proceedings of the 2nd International Conference on Privacy Enhancing Technologies*, pages 54–68, Berlin, Heidelberg, April 2003.

[DT09] George Danezis and Carmela Troncoso. Vida: How to Use Bayesian Inference to De-anonymize Persistent Communications. In *Proceedings of the 9th International Symposium on Privacy Enhancing Technologies*, pages 56–72, Berlin, Heidelberg, July 2009.

[EMB02] Holger Ebel, Lutz-Ingo Mielsch Mielsch, and Stefan Borhholdt. Scale-Free Topology of E-mail Networks. *Physical Review E*, 66(3):1–4, September 2002.

[ER59] Paul Erdos and Alfred Rényi. On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[FBH+02] Nick Feamster, Magdalena Balazinska, Greg Harfst, Hari Balakrishnan, and David Karger. Infranet: Circumventing Web Censorship and Surveillance. In *Proceedings of the 11th USENIX Security Symposium*, pages 247–262, Berkeley, CA, USA, August 2002.

[FID14] FIDIS. Future of Identity in the Information Society: The IST FIDIS Network of Excellence. `http://www.fidis.net/`, 2014.

[FM02] Michael J. Freedman and Robert Morris. Tarzan: A Peer-to-peer Anonymizing Network Layer. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, pages 193–206, New York, NY, USA, November 2002.

[FMMN09] Santo Fortunato, Giuseppe Mangioni, Ronaldo Menezes, and Vincenzo Nicosia. Identification of Social Tension in Organizational Networks. In *Complex Networks*, pages 209–223, 2009.

[Fou15] German Research Foundation. JAP: The JAP Anonymity and Privacy Homepage. `http://anon.inf.tu-dresden.de/`, 2015.

[Fri07] Lothar Fritsch. State of the Art of Privacy-Enhancing Technology (PET). Technical Report, (Norwegian Computing Center, 2007.

[GA05] Ralph Gross and Alessandro Acquisti. Information Revelation and Privacy in Online Social Networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, pages 71–80, New York, NY, USA, November 2005.

[GD11] Robert Gellman and Pam Dixon. *Online Privacy: A Reference Handbook*. ABC-CLIO, 2011.

[Gel02] Robert Gellman. Privacy, Consumers, and Costs How The Lack of Privacy Costs Consumers and Why Business Studies of Privacy Costs are Biased and Incomplete. `https://epic.org/reports/dmfprivacy.html`, 2002.

[GM08] Catherine Greenhill and Brendan D. McKay. Asymptotic Enumeration of Sparse Nonnegative Integer Matrices with Specified Row and Column Sums. *Advances in Applied Mathematics*, 41(4):459–481, Octuber 2008.

[GMRCnSO+12] Javier Portela García-Miguel, Delfín Rupérez Cañas, Ana Lucila Sandoval Orozco, Alejandra Guadalupe Silva Trujillo, and Luis Javier García Villalba. Ataque de Revelación de Identidades en un Sistema Anónimo de Correo Electrónico. In *Actas de la XII Reunión Española de Criptología y Seguridad de la Información (RECSI)*, pages 411–416, Donostia-San Sebastián, Spain, September 2012.

[Gol03] Ian Goldberg. Privacy-Enhancing Technologies for the Internet, II: Five Years Later. In *Proceedings of the 2nd international conference on Privacy enhancing technologies*, pages 1–12, Berlin, Heidelberg, Abril 2003.

[Gol07a] Ian Goldberg. Improving the Robustness of Private Information Retrieval. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, pages 131–148, Washington, DC, USA, May 2007.

[Gol07b] Ian Goldberg. Privacy-Enhancing Technologies for the Internet III: Ten Years Later. In *Digital Privacy: Theory, Technologies and Practices*, pages 3–18, New York, London, December 2007.

[Goo15] Google. Google Street View. `http://maps.google.com`, 2015.

[Gro05] Meta Group. Privacy Enhancing Technologies Ministry of Science, Technology and Innovation. `https://danskprivacynet.files.wordpress.com/2008/07/rapportvedrprivacyenhancingtechlologies.pdf`, 2005.

[GRPS03] Sharad Goel, Mark Robson, Milo Polte, and Emin Sirer. Herbivore: A Scalable and Efficient Protocol for Anonymous Communication. Technical Report, Cornell University, 2003.

[GRS96] David M. Goldschlag, Michael G. Reed, and Paul F. Syverson. Hiding Routing Information. In *Proceedings of the First International Workshop on Information Hiding*, pages 137–150, London, UK, May 1996.

[GT96] Ceki Gulcu and Gene Tsudik. Mixing Email with Babel. In *Proceedings of the 1996 Symposium on Network and Distributed System Security (SNDSS '96)*, pages 2–16, Washington, DC, USA, February 1996.

[GVSTP15] Luis Javier García Villalba, Alejandra Guadalupe Silva Trujillo, and Javier Portela. *Handbook on Data Centers*, chapter Privacy in Data Centers: A Survey of Attacks and Countermeasures, pages 1029–1043. Springer, 2015.

[GW98] Ian Goldberg and David Wagner. TAZ Servers and the Rewebber Network: Enabling Anonymous Publishing on the World Wide Web. *First Monday*, 3(4), August 1998.

[GWB97] Ian Goldberg, David Wagner, and Eric Brewer. Privacy-Enhancing Technologies for the Internet. In *Proceedings of the 42Nd IEEE International Computer Conference*, pages 103–109, Washington, DC, USA, February 1997.

[Hin02] Andrew Hintz. Fingerprinting Websites Using Traffic Analysis. In *Proceedings of the 2nd International Conference on Privacy Enhancing Technologies*, pages 171–178, Berlin, Heidelberg, April 2002.

[Hir80] Jack Hirshleifer. Privacy: Its Origin, Function, and Future. *The Journal of Legal Studies*, 9(4):649–664, December 1980.

[HM08] Alejandro Hevia and Daniele Micciancio. An Indistinguishability-Based Characterization of Anonymous Channels. In *Proceedings of the 8th International Symposium on Privacy Enhancing Technologies*, pages 23–43, Berlin, Heidelberg, July 2008.

[HS04] Dominic Hughes and Vitaly Shmatikov. Information Hiding, Anonymity and Privacy: A Modular. *Journal of Computer security*, 12(1):3–36, January 2004.

[I2P15] I2P. The Invisible Internet Project (I2P). `http://www.i2p2.de`, 2015.

[Inc07] Credentica Inc. Credentica. `http://www.credentica.com`, 2007.

[IPKA10] Tomas Isdal, Michael Piatek, Arvind Krishnamurthy, and Thomas Anderson. Privacy-preserving P2P Data Sharing with OneSwarm. *SIGCOMM Computer Communication Review*, 40(4):111–122, August 2010.

[Jak99] Markus Jakobsson. Flash Mixing. In *Proceedings of the Eighteenth Annual ACM Symposium on Principles of Distributed Computing*, pages 83–89, New York, NY, USA, May 1999.

[JBBM06] Collin Jackson, Andrew Bortz, Dan Boneh, and John C. Mitchell. Protecting Browser State from Web Privacy Attacks. In *Proceedings of the 15th international conference on World Wide Web*, pages 733–744, New York, NY, USA, May 2006.

[Kan98] Jerry Kang. Information Privacy in Cyberspace Transactions. *Standford Law Review*, 50(4):1193–1294, April 1998.

[KEB98] Dogan Kesdogan, Jan Egner, and Roland BÃ$\frac{1}{4}$schkes. Stop-And-Go-MIXes Providing Probabilistic Anonymity in an Open System. In *Proceedings of the Information Hiding Workshop*, pages 83–98, November 1998.

[KH04] Stefan Köpsell and Ulf Hillig. How to Achieve Blocking Resistance for Existing Systems Enabling Anonymous Web Surfing. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society*, pages 47–58, New York, NY, USA, October 2004.

[KP04] Dogan Kesdogan and Lexi Pimenidis. The Hitting Set Attack on Anonymity Protocols. In *Proceedings of the 6th International Conference on Information Hiding*, pages 326–339, Berlin, Heidelberg, May 2004.

[Kri13] Bharadwaj Krishnamurthy. Privacy and Online Social Networks: Can Colorless Green Ideas Sleep Furiously? *Security and Privacy*, 11(3):14–20, May 2013.

[KS08] David Knoke and Yang Song. *Social Network Analysis*. SAGE publications, 2008.

[KvGtH+04] Ronald Koorn, Herman van Gils, Joris ter Hart, Paul Overbeek, Raul Tellegen, and J. Borking. Privacy Enhancing Technologies, White Paper for Decision Makers. Technical Report, Ministry of the Interior and Kingdom Relations, the Netherlands, 2004.

[Lac00] Gerard Lacoste. *Semper–Secure Electronic Marketplace for Europe.* Springer-Verlag New York, Inc., 2000.

[LKF07] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph Evolution: Densification and Shrinking Diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1):1–41, March 2007.

[LLDM09] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Mathematics*, 6(1):29–123, 2009.

[Loe09] Karsten Loesing. Privacy-Enhancing Technologies for Private Services. PhD. Thesis, University of Bamberg, 2009.

[LS02] Brian Neil Levine and Clay Shields. Hordes: A Multicast Based Protocol for Anonymity. *Journal of Computer Security*, 10(3):213–240, September 2002.

[Lyn14] Jennifer Lynch. FBI Plans to Have 52 Million Photos in its NGI Face Recognition Database by Next Year. `https://www.eff.org/deeplinks/2014/04/fbi-plans-have-52-million-photos-its-ngi-face-recognition-next-year`, 2014.

[M05] Group M. Privacy Enhancing Technologies. Technical Report, (Ministry of Science, Technology and Innovation, 2005.

[MD05] Nick Mathewson and Roger Dingledine. Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. In *Proceedings of the 4th International Conference on Privacy Enhancing Technologies*, pages 17–34, Berlin, Heidelberg, May 2005.

[Mic94] James Michael. *Privacy and Human Rights: An International and Comparative Study, with Special Reference to Developments in Information Technology.* Dartmouth Pub Co, 1994.

[Mil67] Stanley Milgram. The Small World Problem. *Psychology Today*, 2(1):60–67, May 1967.

[MOT12] Farnaz Moradi, Tomas Olovsson, and Philippas Tsigas. Towards Modeling Legitimate and Unsolicited Email Traffic Using Social Network Properties. In *Proceedings of the Fifth Workshop on Social Network Systems*, pages 1–6, New York, NY, USA, April 2012.

[MW11a] Nayantara Mallesh and Matthew Wright. A Practical Complexity-Theoretic Analysis of Mix Systems. In *Proceedings of the 12th International Conference on Information Hiding*, pages 221–234, Berlin, Heidelberg, September 2011.

[MW11b] Nayantara Mallesh and Matthew Wright. An Analysis of the Statistical Disclosure Attack and Receiver-bound Cover. *Computers and Security*, 30(8):597–612, November 2011.

[NS09] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing Social Networks. In *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, pages 173–187, Washington, DC, USA, May 2009.

[Oll07] Gunter Ollmann. The Phishing Guide-Understanding & Preventing Phishing Attacks. Technical Report, IBM Internet Security Systems, 2007.

[oPitIAC07] Committee on Privacy in the Information Age and National Research Council. *Engaging Privacy and Information Technology in a Digital Age*. National Academy Press, 2007.

[OTPG14] Simon Oya, Carmela Troncoso, and Fernando Pérez-González. Do Dummies Pay Off? Limits of Dummy Traffic Protection in Anonymous Communications. In *Privacy Enhancing Technologies*, pages 204–233, July 2014.

[PGMGVST+15] Javier Portela García-Miguel, Luis Javier García Villalba, Alejandra Guadalupe Silva Trujillo, Ana Lucila Sandoval Orozco, and Tai-Hoon Kim. Disclosing User Relationships in Email Networks. *Journal of Supercomputing (accepted)*, September 2015.

[PGT12] Fernando Pérez-González and Carmela Troncoso. Understanding Statistical Disclosure: A Least Squares Approach. In *Proceedings of the 12th International Conference on Privacy Enhancing Technologies*, pages 38–57, Berlin, Heidelberg, July 2012.

[PGTO14] Fernando Perez-Gonzalez, Carmela Troncoso, and Simon Oya. A Least Squares Approach to the Static Traffic Analysis of High-Latency Anonymous Communication Systems. *IEEE Transactions on Information Forensics and Security*, 9(9):1341–1355, September 2014.

[PGVST+15] Javier Portela, Luis Javier García Villalba, Alejandra Guadalupe Silva Trujillo, Ana Lucila Sandoval Orozco, and Tai-hoon Kim. Extracting Association Patterns in Network Communications. *Sensors*, 15(2):4052–4071, 2015.

[PH08] Andreas Pfitzmann and Marit Hansen. Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology. `http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.31.pdf`, 2008.

[Pos78a] Richard A. Posner. An Economic Theory of Privacy. *Regulation*, 2:19–26, June 1978.

[Pos78b] Richard A. Posner. Privacy, Secrecy, and Reputation. *Buffalo Law Review*, 28(1):1–55, 1978.

[PPW91] Andreas Pfitzmann, Birgit Pfitzmann, and Michael Waidner. ISDN-mixes: Untraceable Communication with Very Small Bandwidth Overhead. In *Proceedings of the GI/ITG Conference on Communication in Distributed Systems*, pages 451–463, 1991.

[PRI07] PRIME. Privacy and Identity Management for Europe: The IST PRIME Project. `http://www.prime-project.eu./`, 2007.

[PS09] Jung-Heum Park and Sang Hyuk Son. Conditional Matching Preclusion for Hypercube-Like Interconnection Networks. *Theoretical Compututer Science*, 410(27):2632–2640, June 2009.

[PWK10] Dang Vinh Pham, Joss Wright, and Dogan Kesdogan. The Reverse Statistical Disclosure Attack. In *Proceedings of the 16th European Conference on Research in Computer Security*, pages 508–527, Berlin, Heidelberg, June 2010.

[Rap03] Fabio Rapallo. Algebraic Markov Bases and MCMC for Two-Way Contingency Tables. *Scandinavian Journal of Statistics*, 30(2):385–397, 2003.

[Ray01a] Jean-François Raymond. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In *International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability*, pages 10–29, New York, NY, USA, January 2001.

[Ray01b] Jean-Francois Raymond. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In *Proceedings of the International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability*, pages 10–29, 2001.

[RP04] Marc Rennhard and Bernhard Plattner. Practical Anonymity for the Masses with MorphMix. In *Financial Cryptography*, pages 233–250, Key West, FL, USA, February 2004.

[RR98] Michael K. Reiter and Aviel D. Rubin. Crowds: Anonymity for Web Transactions. *ACM Transactions on Information and System Security (TISSEC)*, 1(1):66–92, November 1998.

[SBS02] Rob Sherwood, Bobby Bhattacharjee, and Aravind Srinivasan. P5: A Protocol for Scalable Anonymous Communication. In *Proceedings of the 2002 IEEE*

Symposium on Security and Privacy, pages 58–70, Washington, DC, USA, May 2002.

[SCM05] Len Sassaman, Bram Cohen, and Nick Mathewson. The Pynchon Gate: A Secure Method of Pseudonymous Mail Retrieval. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, pages 1–9, New York, NY, USA, November 2005.

[SD03] Andrei Serjantov and George Danezis. Towards an Information Theoretic Metric for Anonymity. In *Proceedings of the 2nd International Conference on Privacy Enhancing Technologies*, pages 41–53, Berlin, Heidelberg, April 2003.

[SG92] Calvin Streeter and David Gillespie. Social Network Analysis. *Journal of Social Service Research*, 16(1):201–222, 1992.

[SM08] Muhammad Sher and Thomas Magedanz. Secure Access to IP Multimedia Services Using Generic Bootstrapping Architecture (GBA) for 3G & Beyond Mobile Networks. In *Proceedings of the 2nd ACM international workshop on Quality of service & security for wireless and mobile networks*, pages 17–24, New York, NY, USA, Octuber 2008.

[Sol06] Daniel J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3):477–564, January 2006.

[Sol07] Daniel J. Solove. I've Got Nothing to Hide' and Other Misunderstandings of Privacy. *San Diego Law Review*, 44:745–768, 2007.

[SP07] Len Sassaman and Bart Preneel. The Byzantine Postman Problem: A Trivial Attack Against PIR-based Nym Servers. Technical Report, Katholieke Universiteit Leuven, 20007.

[Sta08] Illinois Statutes. Biometric Information Privacy Act. `http://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57`, 2008.

[Sta09] Texas Statutes. Chapter 503. Biometric Identifiers. `http://www.statutes.legis.state.tx.us/Docs/BC/htm/BC.503.htm`, 2009.

[STGV08] Alejandra Guadalupe Silva Trujillo and Luis Javier García Villalba. Redes Sociales: Retos, Oportunidades y Propuestas para Preservar la Privacidad. In *Actas del XXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2008)*, Madrid, España, September 2008.

[STGVD08] Alejandra Guadalupe Silva Trujillo, Luis Javier García Villalba, and Claudia Díaz. Construcción de Redes Sociales Anónimas. In *Actas de la X Reunión*

Española sobre Criptología y Seguridad de la Información (RECSI 2008), pages 647–652, Salamanca, España, September 2008.

[STPGMGV14] Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, and Luis Javier García Villalba. Refinamiento Probabilístico del Ataque de Revelación de Identidades. In *Actas de la XIII Reunión Española sobre Criptología y Seguridad de la Información (RECSI 2014)*, pages 297–302, Alicante, España, September 2014.

[STPGMGV15a] Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, and Luis Javier García Villalba. Ataque y Estimación de la Tasa de Envíos de Correo Electrónico mediante el Algoritmo EM. In *Actas del VIII Congreso Iberoamericano de Seguridad Informática (CIBSI 2015)*, Quito, Ecuador, November 2015.

[STPGMGV15b] Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, and Luis Javier García Villalba. Construcción de Redes Sociales Anónimas. In *Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2015)*, Pamplona, España, September 2015.

[STPGMGV15c] Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, and Luis Javier García Villalba. Extracción de Características de Redes Sociales Anónimas a través de un Ataque Estadístico. In *Actas del VIII Congreso Iberoamericano de Seguridad Informática (CIBSI 2015)*, Quito, Ecuador, November 2015.

[STRL01] Paul Syverson, Gene Tsudik, Michael Reed, and Carl Landwehr. Towards an Analysis of Onion Routing Security. In *International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability*, pages 96–114, New York, NY, USA, March 2001.

[SWY+03] Kent E. Seamons, Marianne Winslett, Ting Yu, Lina Yu, and Ryan Jarvis. Protecting Privacy During On-line Trust Negotiation. In *Proceedings of the 2nd International Conference on Privacy Enhancing Technologies*, pages 129–143, Berlin, Heidelberg, April 2003.

[TAKS07] Patrick P. Tsang, Man Ho Au, Apu Kapadia, and Sean W. Smith. Blacklistable Anonymous Credentials: Blocking Misbehaving Users Without Ttps. In *Proceedings of the 14th ACM Conference on Computer and Communications Security*, pages 72–81, New York, NY, USA, October 2007.

[TAKS08] Patrick P. Tsang, Man Ho Au, Apu Kapadia, and Sean W. Smith. Perea: Towards Practical Ttp-Free Revocation in Anonymous Authentication. In

Proceedings of the 15th ACM Conference on Computer and Communications Security, pages 333–344, New York, NY, USA, October 2008.

[TGPV08] Carmela Troncoso, Benedikt Gierlichs, Bart Preneel, and Ingrid Verbauwhede. Perfect Matching Disclosure Attacks. In *Proceedings of the 8th International Symposium on Privacy Enhancing Technologies*, pages 2–23, Berlin, Heidelberg, July 2008.

[TWH05] Joshua R. Tyler, Dennis M. Wilkinson, and Bernardo A. Huberman. E-mail as Spectroscopy: Automated Discovery of Community Structure within Organizations. *The Information Society*, 21(2):143–153, 2005.

[UMH10] Mohammed Uddin, Shahriar Tanvir Hasan Murshed, and Liaquat Hossain. Towards A Scale Free Network Approach to Study Organizational Communication Network. In *PACIS 2010 - 14th Pacific Asia Conference on Information Systems*, pages 1937–1944, 2010.

[VA05] Alejandro Velásquez and Norman Aguilar. Manual Introductorio al Análisis de Redes Sociales: Medidas de Centralidad. `http://revista-redes.rediris.es/webredes/talleres/Manual_ARS.pdf`, 2005.

[VWW05] Hal Varian, Fredrik Wallenberg, and Glenn Woroch. The Demographics of the Do-Not-Call List. *IEEE Security and Privacy*, 3(1):34–39, January 2005.

[Wes68] Alan F. Westin. Privacy and Freedom. *Washington and Lee Law Review*, 25(1):166, 1968.

[WF94] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[WM01] Marc Waldman and David Mazieres. Tangler: A Censorship-Resistant Publishing System Based on Document Entanglements. In *Proceedings of the 8th ACM Conference on Computer and Communications Security (CCS 2001)*, pages 126–135, November 2001.

[WRC00] Marc Waldman, Aviel Rubin, and Lorrie Cranor. Publius: A Robust, Tamper-Evident, Censorship-Resistant and Source-Anonymous Web Publishing System. In *Proceedings of the 9th USENIX Security Symposium*, pages 59–72, August 2000.

[WS98] Duncan Watts and Steven Strogatz. Collective Dynamics of 'Small-World' Networks. *Nature*, 393(6684):440–442, June 1998.

[WSSV15] Donghee Yvette Wohn, Jacob Solomon, Dan Sarkar, and Kami E. Vaniea. Factors Related to Privacy Concerns and Protection Behaviors Regarding

Behavioral Advertising. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1965–1970, New York, NY, USA, April 2015.

Part II

Papers Related to This Thesis

Appendix A

List of Papers

1. Alejandra Guadalupe Silva Trujillo, Luis Javier García Villalba, Claudia Diaz: "Construcción de Redes Sociales Anónimas". Actas de la X Reunión Española sobre Criptología y Seguridad de la Información (RECSI 2008), Salamanca, España, Septiembre 2 - 5, 2008.

2. Alejandra Guadalupe Silva Trujillo, Luis Javier García Villalba: "Redes Sociales: Retos, Oportunidades y Propuestas para Preservar la Privacidad". Actas del XXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2008). Madrid, España, Septiembre 22 - 24, 2008.

3. Javier Portela García-Miguel, Delfín Rupérez Cañas, Ana Lucila Sandoval Orozco, Alejandra Guadalupe Silva Trujillo, Luis Javier García Villalba: "Ataque de Revelación de Identidades en un Sistema Anónimo de Correo Electrónico". Actas de la XII Reunión Española sobre Criptología y Seguridad de la Información (RECSI 2012), San Sebastián-Donostia, España, Septiembre 4 - 5, 2014.

4. Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba: "Derivations of Traffic Data Analysis". Proceedings of the 6th International Conference on Information Technology (ICIT 2013). Amman, Jordan, May 8 - 10, 2013.

5. Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba: "Refinamiento Probabilístico del Ataque de Revelación de Identidades". Actas de la XIII Reunión Española sobre Criptología y Seguridad de la Información (RECSI 2014), Alicante, España, Septiembre 2 - 5, 2014, páginas 297 - 302.

6. Luis Javier García Villalba, Alejandra Guadalupe Silva Trujillo, Javier Portela: "Privacy in Data Centers: A Survey of Attacks and Countermeasures". Chapter 34 in Handbook on Data Centers (Samee U. Khan and Albert Y. Zomaya, Editors). Springer, USA, pp. 1029–1043, 2015.

7. Javier Portela, Luis Javier García Villalba, Alejandra Guadalupe Silva Trujillo, Ana Lucila Sandoval Orozco, Tai-Hoon Kim: "Extracting Association Patterns in Network Communications". Sensors, 2015, 15, pages 4052–4071, February 2015.

8. Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba: "Sistema para la Detección de Comunicaciones entre Usuarios de Correo Electrónico". Actas del XXX Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2015). Pamplona, España, Septiembre 2 - 4, 2015.

9. Javier Portela, Luis Javier García Villalba, Alejandra Guadalupe Silva Trujillo, Ana Lucila Sandoval Orozco, Tai-Hoon Kim: "Disclosing User Relationships in Email Networks". Journal of Supercomputing (aceptado, en proceso de publicación), September 2015.

10. Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba: "Ataque y Estimación de la Tasa de Envíos de Correo Electrónico mediante el Algoritmo EM". Actas del VIII Congreso Iberoamericano de Seguridad Informática (CIBSI 2015). Quito, Ecuador, Noviembre 10 - 12, 2015.

11. Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba: "Extracción de Características de Redes Sociales Anónimas a través de un Ataque Estadístico". Actas del VIII Congreso Iberoamericano de Seguridad Informática (CIBSI 2015). Quito, Ecuador, Noviembre 10 - 12, 2015.

X Reunión Española sobre Criptología y Seguridad de la Información

VNiVERSiDAD D SALAMANCA

FINAL DE LA INSCRIPCIÓN: 31/8/2008

Salamanca, del 2 al 5 de septiembre de 2008

Departamento de Matemática Aplicada

Universidad de Salamanca

VNiVERSiTAS STVDii SALAMANTiNi

Construcción de redes sociales anónimas

Alejandra Silva[1], L. Javier García[1], Claudia Diaz[2]

1. Grupo de Análisis, Seguridad y Sistemas (GASS)
Departamento de Sistemas Informáticos y Programación
Facultad de Informática
Universidad Complutense de Madrid (UCM)
E-mail: asilva@fdi.ucm.es, javiergv@sip.ucm.es

2. K.U.Leuven ESAT/COSIC
E-mail: claudia.diaz@esat.kuleuven.be

***Resumen*—Analizar una red social permite identificar sus líderes, roles y comunidades, así como su comportamiento, tamaño y heterogeneidad. Esta información es muy valiosa para optimizar o personalizar servicios, y para predecir el comportamiento de la red. Pero al mismo tiempo dichos análisis conllevan intrusiones a la privacidad de los individuos que la conforman. En el presente artículo se revisan las técnicas, algoritmos, y procedimientos que se han presentado recientemente en el campo de investigación para la anonimización de redes sociales, a fin de presentar un amplio panorama de lo que se ha propuesto, y los interrogantes que quedan aún por resolver.**

***Palabras clave*— Anonimato (*anonymity*), algoritmos para grafos (*graph algorithms*), análisis de redes sociales (*social networks analysis*), minería de datos (*data mining*), privacidad (*privacy*).**

I. Introducción

En los últimos años se han desarrollado tecnologías que permiten establecer comunidades sociales virtuales, así como trasladar al mundo virtual las comunidades existentes en el mundo real. Estas tecnologías están transformando la manera en que se desarrollan las relaciones sociales, y teniendo un gran impacto en nuestra sociedad.

Diversos investigadores han abordado este tema desde diferentes áreas de interés, entre ellas la mercadotecnia, epidemiología, sociología, criminalística, o terrorismo, entre otras. El análisis de redes sociales se ha facilitado en años recientes gracias al desarrollo de Internet y a la gran cantidad de información disponible.

Analizar la estructura de una red social revela información de los individuos que la componen, ya que al analizar las conexiones entre individuos podemos identificar los roles que tienen en su grupo o comunidad; así como las dinámicas de las relaciones entre individuos. Esta información es de gran utilidad para comprender mejor las dinámicas sociales. Por ejemplo, sociólogos e historiadores desean conocer la interrelación entre los actores sociales o políticos de una determinada red social para identificar agentes de cambio [1]. Otras investigaciones se han enfocado en analizar los envíos de correos electrónicos, con el objetivo de identificar comunidades y observar su comportamiento [2, 3, 4]. Para el análisis de las bitácoras en línea (*blogs)*, se emplean técnicas de inferencia colectiva que predicen el comportamiento de una entidad a través de sus conexiones. Y mediante técnicas de aprendizaje automático o modelos de lenguaje natural [5, 6], se pretende identificar al autor de un texto al realizar un análisis de su vocabulario y manera de escribir. Sin duda las redes sociales en Internet como *MySpace*, *Friendster*, *Match.com*, *FaceBook,* entre otras, han atraído la atención de millones de personas que participan en ellas activamente para establecer contacto con amigos, buscar empleo o pareja, compartir fotos, música, videos, etc. Diversas publicaciones han demostrado la sorprendente cantidad de información personal que usuarios de estos sitios publican, sin que parezca que sean conscientes de los riesgos que conlleva que esa información sea utilizada en otros contextos [12] [13]. Por ejemplo, cuando empresas encargadas de la contratación de personal realizan búsquedas en redes sociales en línea para investigar el perfil de sus candidatos [14].

A raíz de los eventos del 11 de septiembre se legitimó la aplicación de toda clase de herramientas tecnológicas para vigilar y monitorizar a las personas. Desde entonces, muchas naciones han reformado su legislación para permitir la recopilar información relativa al tráfico y la localización de dispositivos electrónicos como teléfonos fijos y móviles, servicios de mensajes cortos, faxes, *e-mails*, salas de conversación en línea, Internet, entre otros [14]. La justificación radica en prevenir, investigar y perseguir actividades ilícitas o delictivas que atenten contra el orden público, la salud o la seguridad nacional. Mientras tanto, la industria argumenta que el conocer los gustos y hábitos de sus clientes les permite mejorar y personalizar sus servicios.

Sin embargo, desde organizaciones que defienden y promocionan derechos relativos a la privacidad se han expresado recelos con respecto a los riesgos asociados a establecer estos mecanismos de vigilancia masiva. En particular, preocupa la falta de transparencia y responsabilidad con respecto al uso que se da a esta información, y los posibles abusos que se puedan derivar de ello. Un caso extremo que ilustra la importancia de proteger esta información lo ofrecen naciones con regímenes totalitarios, donde grupos de disidentes, periodistas, protestantes cívicos, líderes estudiantiles, organizaciones políticas de oposición o precursores de los derechos humanos quedan expuestos y en peligro para su integridad física [15].

Como podemos observar, es necesario establecer mecanismos para permitir a los individuos proteger la información relativa a las redes sociales a las que pertenecen. El objetivo de este trabajo es presentar un estudio del arte de las propuestas que se han desarrollado recientemente, en el campo de la construcción de redes sociales anónimas. En la sección 2 presentamos la relación de las redes sociales con el

anonimato. En la sección 3 se plantean tipos de ataques que deben ser tenidos en cuenta. En la sección 4 se analizan las técnicas, protocolos y algoritmos para construir redes sociales anónimas; y finalizamos este artículo en la sección 5, donde se presentan las conclusiones de este estudio así como los aspectos que requieren más investigación.

II. Formulación del problema

En esta sección formalizamos la definición y el modelado de redes sociales, definimos las brechas de privacidad que se desean impedir, y presentamos los supuestos y el planteamiento del problema.

A. Definición de red social.

Una red social puede representarse a través de un grafo donde los vértices representan a las personas y las aristas son las relaciones entre ellas. Formalmente, una red social se modela como un grafo $G = (V, E)$ donde:

- $V = (v_1,\ldots, v_n)$ es el conjunto de vértices o nodos que representan a entidades o individuos
- E es el conjunto de relaciones sociales entre ellos (representadas como aristas en el grafo) donde $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$.

B. Definición de anonimizar.

Definimos *anonimizar* como el proceso de transformar un grafo G en su equivalente anónimo AG.

C. Brechas de privacidad.

Las brechas de privacidad en la información de redes sociales pueden ser agrupadas en 3 categorías [11]: 1) revelación de identidad (*identity disclosure*): se descubre la identidad de los individuos asociados los vértices; 2) revelación de conexión (*link disclosure*): se descubren las conexiones entre dos vértices; 3) revelación de contenido (*content disclosure*): se compromete la privacidad de los datos coligados con cada vértice. El objetivo de anonimizar una red social es impedir que la identidad, conexiones, y contenido de los vértices sean revelados.

D. Planteamiento del problema.

En [8] se considera el siguiente planteamiento para definir el problema de preservar la privacidad de los datos en una red social publicada:

1) Se debe identificar la información que se desea proteger.
2) Se debe modelar el conocimiento y habilidades del adversario que trata de comprometer la privacidad.
3) Por último, se debe especificar el uso de la red social, de tal manera que se elija un método de anonimato adecuado que preserve la utilidad de la red y proteja su privacidad.

En este artículo consideramos ataques de revelación de identidad, donde el adversario intenta descubrir la correspondencia entre vértices del grafo anónimo y usuarios de la red social. Para los ataques activos y pasivos se asume que el adversario conoce al completo el grafo $G = (V, E)$ de la red social, la cuál es anonimizada a través de *naive anonimization*. Para los ataques de vecindario se asume que el adversario conoce sólo a los vecinos inmediatos de ciertos vértices y cómo están conectados. Las técnicas que se consideran en este trabajo para anonimizar la red son para prevenir la revelación de identidad. El recurso más utilizado para ocultar la correspondencia entre identidades y su correspondencia con los vértices en una red social es añadir y/o eliminar vértices y aristas.

III. Ataques

A. Anonimización inexperta (naive anonymization)

Primero revisamos el proceso de anonimización inexperto conocido en inglés como *naive anonymization* [7]. Consiste simplemente en renombrar los vértices de G con pseudónimos para prevenir la revelación de su identidad, sin modificar la estructura de la red. En este escenario, si el atacante cuenta por anticipado con información estructural de la red, podrá con alta probabilidad relacionar vértices del grafo anónimo con sus correspondientes. Por ejemplo: En la Fig. 1 vemos que Alicia está relacionada con Beto y Carlos, y que ambos tienen dos conexiones. Cuando el atacante observa el grafo anónimo, puede identificar al vértice 1 con Alicia, puesto que es el único con una estructura de conexiones que encaja con la esperada para Alicia.

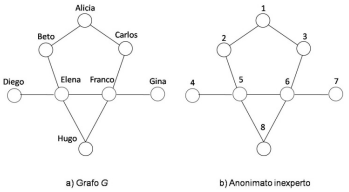


Fig. 1. Ejemplo de anonimato inexperto.

B. Ataques activos.

El objetivo del ataque activo es revelar las identidades de un conjunto de usuarios previamente elegidos. A estos usuarios se les conoce como usuarios víctima b.

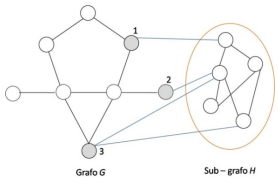


Fig. 2. Ejemplo de ataque activo con $k = 5$ y $b = 3$.

El adversario activo ejecuta los siguientes pasos antes de la anonimización de la red social G: 1) selecciona arbitrariamente un conjunto usuarios víctima $w_1,\ldots, w_b$; 2)

genera k vértices nuevos $X = \{x_1, \ldots, x_k\}$; y 3) crea las conexiones a los usuarios víctima $\{x_1, \ldots, w_1\}$ (por ejemplo enviando mensajes, creando entradas en libretas de direcciones, o alguna otra actividad dependiendo de la naturaleza de la red social). En la Fig. 2 se muestra un ejemplo donde un subgrafo H es insertado en G, con un valor $k = 5$ para un conjunto $\{x_1, \ldots, x_5\}$, y $b = 3$ usuarios víctima $\{w_1, w_2, w_3\}$.

Cuando G sea anonimizada con el modelo *naïve anonymization* el atacante podrá identificar H, analizando las aristas de AG, lo que le permite identificar a los usuarios víctima $w_1, \ldots, w_b$ y por lo tanto comprometer su privacidad. El adversario debe construir H con las siguientes propiedades para que sea efectiva: 1) debe ser única en G; 2) debe ser eficientemente localizable; y 3) no debe tener automorfismos

En base al principio de los ataques activos señalado, se presentan dos variantes cuyas diferencias radican en la construcción de H y en los algoritmos aplicados para la recuperación de H en AG.

1) Ataque basado en recorridos (Walk based attack)

Para el primer ataque se muestra que un subgrafo H creado aleatoriamente con $k = \Theta(\log n)$ vértices puede ser identificado en AG con alta probabilidad. Y si el máximo grado de los vértices en H es $\Theta(\log n)$, entonces H puede ser recuperada eficientemente. El algoritmo utilizado para la recuperación de H se llama *walk-based attack*. Consiste en localizar la ruta $x_1, x_2, \ldots, x_k$ en el grafo AG. El proceso inicia haciendo un recorrido vértice por vértice, y verificando la secuencia de grados de los vértices en la ruta para comprobar si coinciden con los esperados para el conjunto X.

Al aplicar este ataque en una red social de 4.4 millones de vértices y 77 millones de conexiones, se destaca que utilizando un valor $k = 7$ puede comprometerse en promedio la identidad de 70 vértices y cerca de 2,400 aristas.

2) Ataque basado en cortes (*cut-based attack*)

El segundo ataque activo llamado ataque basado en cortes también construye H aleatoriamente, pero es insertada en G utilizando muy pocas conexiones. El atacante recupera H a través de cálculos con alta complejidad basados en el modelo de corte de árboles Gumory-Hu. Este ataque utiliza $k = O(\sqrt{\log n})$ para revelar la identidad de $\Theta(\sqrt{\log n})$ vértices víctima. Se muestra que, en el peor de los casos se deben crear al menos de $\Omega(\sqrt{\log n})$ vértices nuevos en cualquier ataque activo que necesite un subgrafo H único e identificable con alta probabilidad.

En la construcción de H denotamos $\delta(H)$ el mínimo grado en H, y $\gamma(H)$ el valor del mínimo corte en H (número mínimo de conexiones cuya eliminación desconectan a H de G). Para encontrar H se utiliza el algoritmo de corte Gumory-Hu basado, en términos generales en seccionar el grafo iterativamente hasta encontrar un conjunto de vértices isomorfo a H. A diferencia del ataque basado en caminos, la aplicación del algoritmo para la recuperación de H en éste ataque, tiene un alto costo en términos computacionales. Para una red G con 100 millones de vértices, y un valor de $k = 12$, el adversario es capaz de identificar a $b = 3$ usuarios víctima con probabilidad de al menos 0.99.

C. Ataques pasivos.

El adversario considerado en este ataque es un conjunto de usuarios maliciosos que colaboran a fin de identificar a otros vértices en la red social anónima AG. Cuando AG se publica, los vértices maliciosos tratan de localizarse a sí mismos en la versión anónima de la red social. Esto les permite ubicar sus vértices vecinos y comprometer su identidad. Aplicando un ataque pasivo a la misma red social mencionada en el ataque activo basado en caminos, se obtuvo que si un usuario u es capaz de confabularse con otros $(k - 1)$ vértices vecinos, es capaz de identificar a todos los vértices conectados a ellos. Bajo este criterio se asume que: 1) una confabulación X de tamaño k es iniciada por un usuario que convence a $k - 1$ de sus vecinos; 2) los vecinos confabulados conocen con quién están relacionados dentro de X; 3) los vecinos confabulados conocen los nombres de las entidades con quien están relacionados fuera de X. Debido a que en este caso H no se construye aleatoriamente, no hay bases para considerar que sea única y fácilmente identificable.

El ataque se describe de la siguiente manera:

1) Un usuario x_1 selecciona $k - 1$ vecinos para formar una confabulación de usuarios maliciosos $X = \{x_1, \ldots, x_k\}$
2) Una vez que G es publicado, los usuarios maliciosos ejecutan el algoritmo del ataque basado en caminos con modificaciones mínimas.

Una vez que el grupo de usuarios maliciosos se encuentra en el grafo, les es posible determinar la identidad de algunos de sus vecinos en $G - X$.

D. Diferencias entre ataques pasivos y activos.

Los ataques activos tienen efectos más potentes en la red, ya que el adversario puede elegir los vértices que desea comprometer, siempre y cuando la naturaleza de la red le permita introducir sus vértices en las posiciones deseadas. En cambio, los ataques pasivos solamente pueden revelar la identidad de los vértices que están conectados al atacante (modelado como un grupo de usuarios maliciosos), hecho que garantiza su aplicación en casi cualquier tipo de red. Los ataques pasivos a diferencia de los activos no son fáciles de detectar, por el hecho de que el adversario pertenece a la red social y no genera evidencia de intromisiones externas.

E. Ataque de vecindario.

En [8] se identifica otro ataque a la privacidad en redes sociales llamado ataque de vecindario. En una red social $G = (V, E)$, el vecindario de un usuario $u \in V(G)$ es un subgrafo de vecinos de u que se denota como $Vecindario_G(u) = G(N_u)$ donde $N_u = \{v | (u, v) \in E(G)\}$. Si el atacante conoce a los vecinos de su vértice víctima y sus aristas a otros vértices, puede ser capaz de revelar varias identidades en una red social, aún cuando ésta haya sido modificada a través de técnicas de anonimato. Por ejemplo, supongamos que el atacante cuenta con información estructural del grafo G; sabe que Alicia tiene relación con Beto y Carlos, y que ellos a su vez tienen tres vecinos más; el atacante es capaz de identificar a Alicia y a sus vecinos en la red anónima buscando todos aquellos subgrafos con características similares al suyo.

Para proteger la privacidad satisfactoriamente se propone utilizar el modelo *k-vecindario* anonimato que se detallará en la siguiente sección.

IV. TÉCNICAS, ALGORITMOS Y PROTOCOLOS PARA LA CONSTRUCCIÓN DE REDES SOCIALES ANÓNIMAS

Uno de los problemas que se enfrentan en la construcción de redes sociales anónimas es que no se puede considerar cada vértice individualmente, sino que hay que tener en cuenta el grafo en su conjunto, ya que cualquier modificación en un vértice afectará las propiedades del grafo, tales como diámetro, centralidad, heterogeneidad; dando como resultado en la mayoría de los casos una red anónima tan diferente de la original que carece de utilidad.

A. Esquema de encriptación con llave pública.

En [9] presentan una propuesta considerando un escenario en el que múltiples partes tienen una pieza de la red, es decir considera la existencia de "autoridades" que conocen partes del grafo *G*. Se proponen una serie de protocolos criptográficos para transformar *G* en una versión anónima (*AG*), bajo la suposición de que la mayoría de las autoridades son honestas. Se considera que existen autoridades y entidades maliciosas, así como confabulaciones entre ellas. El resultado del proceso de anonimización es un grafo *AG* isomórfico a *G*, y su construcción se realiza a través del conjunto de autoridades de manera que ninguna de ellas es capaz de conocer la relación entre *AG* y *G*.

Para lograr este objetivo se recurre al esquema de encriptación con llave pública ElGamal, que sirve para encriptar la relación entre usuarios y pseudónimos.

Se asume que la red cuenta con conexiones dirigidas no etiquetadas, y se establecen una serie de medidas para evitar ataques por parte de vértices maliciosos que pudieran reportar conexiones específicas para facilitar ataques posteriores. Las medidas preventivas consisten en deshabilitar las etiquetas en las aristas; se eliminan las aristas que van de un vértice a si mismo; limitar las aristas salientes de los vértices (*out-degree*); limitar las aristas entrantes a los vértices (*in-degree*); agregar o eliminar a la red un número aleatorio de aristas y vértices. La selección del protocolo de anonimización se hace de acuerdo con la red, el escenario de aplicación y el objetivo de la transformación.

Para iniciar el proceso de construcción de la red anónima *AG* se permite a cada vértice reportar pseudónimamente sus conexiones en el grafo *G* a las autoridades y se forma una lista *L* de *n* textos cifrados. $E(1),\ldots, E(n)$. Las autoridades que actúan como servidores de mezcla (*mix servers*), ingresan la lista *L*, de la cual resulta una lista con permutaciones $E(\pi(1)),\ldots, E(\pi(n))$. Posteriormente, para ocultar el número de aristas que reportó cada vértice se añaden *n* elementos más a la lista con valor $E(-1)$ de tal manera que la lista *L'* resultante contenga *2n* elementos. Este protocolo sirve como base para la construcción del grafo *AG* tomando en consideración las medidas preventivas introducidas en el párrafo anterior.

B. K-Anonimato.

En [10] se describe el proceso de transformar una red *G* en una red *AG* con *naive anonymization*. Para evitar que el atacante reconozca los vértices en la red anónima a través de su grado o vecinos, se introduce el concepto de *k-candidato* anónimo. Una red satisface la condición de *k-candidato* anónimo si para cada vértice *v* en el grafo G hay al menos *k* vértices en *AG* que podrían corresponder con *v*.

Dos vértices son automórficamente equivalentes si su estructura dentro de la red es igual. Los vértices que cumplen esta condición pueden permanecer ocultos fácilmente ya que son indistinguibles estructuralmente. Para llevar a cabo un ataque sobre este tipo de vértices, el adversario necesita tener un amplio conocimiento de la red, lo cual puede no ser factible en ciertos tipos de redes.

El grado de conocimiento de la red por parte de un adversario proviene de dos tipos de consultas:

1) Consultas de requerimiento de vértices: proporcionan la información estructural de un vértice en la red.
   a. $\mathcal{H}_0(v)$ proporciona el nombre de *v*,
   b. $\mathcal{H}_1(v)$ proporciona el grado,
   c. $\mathcal{H}_2(v)$ proporciona una lista con el grado de cada vecino del vértice *v*;
2) Consultas para el conocimiento del subgrafo: verifican la existencia de un subgrafo específico en torno al vértice *v*.

A través de este tipo de consultas, en [10] se plantean ataques a tres redes diferentes, y los resultados muestran que gran parte de la información del grafo queda al descubierto para el atacante.

La técnica propuesta para construir la red anónima establece una secuencia de *m* relaciones eliminadas seguidas de *m* relaciones agregadas en el grafo *G*. Las relaciones eliminadas se eligen aleatoriamente (uniformemente) del conjunto de relaciones del grafo original. En este modelo se asume que el atacante sólo ataca un vértice a la vez, y que realiza el análisis estructural para identificar ese vértice a través de sus relaciones. De este estudio se derivan diversas interrogantes concernientes a las estrategias que podrían utilizar los atacantes para la óptima recolección de información: Dado un número limitado de tiempo y recursos, ¿podría el adversario obtener información estructural acerca de un vértice o información de sus atributos? Cuando se está recolectando información estructural, ¿cómo selecciona el atacante el siguiente vértice a explorar? De las conclusiones de [10] se destaca la relación inversamente proporcional entre el grado de anonimato y la utilidad de grafo *AG* obtenido tras modificación aleatoria de *G*.

C. K-Vecindario Anonimato.

En [8] se ejemplifica el llamado ataque de vecindario. Esta idea está relacionada con las consultas para conocimiento de subgrafos descritas en el apartado anterior donde se busca el conjunto de vecinos de un vértice *v* en *G*, para identificarlo posteriormente en la red anónima *AG*. Recordemos que se pretende proteger la privacidad de un grafo con la técnica *k-anónima*, de forma que para cada vértice *v* existen por lo menos $k-1$ vértices con igual grado. Decimos que un grafo cumple con la condición de *k-vecindario* si todos sus vértices cumplen la condición de *k-anónima*.

Se define un grafo simple como:

$$G = (V, E, L, \mathcal{L}),$$

donde *V* es el conjunto de vértices, *E* corresponde al conjunto de aristas en $V \times V$, *L* es el conjunto de etiquetas, y la función de etiquetado que asigna a cada vértice su etiqueta correspondiente es $\mathcal{L}: V \rightarrow L$.

Un vértice es *k-anónimo* en un grafo G si existen al menos otros $(k - 1)$ vértices $v_1,\ldots, v_{k-1} \in V_G$ tal que todos los subgrafos construidos por los vecinos de $v_1,\ldots, v_{k-1}$ tienen la misma estructura.

Dado un grafo $G = (V_G, E_G)$ y un entero k, el objetivo es construir un nuevo grafo $AG = (V_{AG}, E_{AG})$ tal que AG sea *k-vecindario* anónimo, y donde $V_{AG} = V_G$, $E_{AG} \supseteq E_G$.

Existen dos formas de anonimizar los vecindarios de vértices: generalizando las etiquetas de vértices y agregando aristas. Por ejemplo, si tenemos una red social donde cada vértice representa a un autor y las aristas ligadas a dos vértices indican que han sido coautores por lo menos en un artículo. Para generalizar las etiquetas de los vértices sería necesario quitar los nombres de los autores y utilizar en su lugar, por ejemplo, el nombre de la institución a la que pertenecen. Añadir aristas permite cumplir con la condición de *k-vecindario*. Ambos métodos generan un costo de anonimización y se elige el que menor costo deriva su aplicación. El costo de anonimización en dos vértices u y v mide la semejanza entre $Vecindario_G(u)$ y $Vecindario_G(v)$. Cuanto menor sea el costo, más similitudes tendrán ambos vecindarios.

En este escenario no se añaden vértices falsos para mantener la estructura global de la red social, y las aristas del grafo G se mantienen en su versión anónima AG. El método para construir AG consiste en dos pasos. Primero, se extraen los vecindarios de todos los vértices en G (para facilitar la comparación entre vecindarios de diferentes vértices se propone una técnica de codificación de componentes de vecindario). El segundo paso consiste en organizar los vértices en grupos, iniciando la anonimización con los de grado mayor. Al anonimizar vecindarios similares se minimiza la pérdida de información en la transformación de G a AG y se preserva cierta similitud entre la red original y anónima.

Las conclusiones derivadas de la aplicación de este algoritmo a un conjunto de datos sintéticos destacan que el costo del anonimato se incrementa con el número de vértices en el grafo y con el parámetro k (dado que el proceso de anonimización de un vértice requiere que haya otros k vértices con idéntica estructura de conexiones). Por último, cuando la conectividad de los vértices se incrementa, el costo de anonimato crece también. Como trabajo futuro se plantea resolver d-vecindarios donde (d >1), ya que sólo se modeló el problema de 1-vecindarios.

D. K- grado anonimato.

En [11], se propone armar un grafo anónimo AG con el mínimo de modificaciones sobre el grafo original G, a fin de preservar su utilidad como representación de G. Considérese un grafo simple G (V, E), donde V es el conjunto de vértices y E el conjunto de aristas en G; dado un grafo G y un entero k, modifique G para construir AG con *k-grado* anónimo, en donde por cada vértice v existan al menos $k - 1$ vértices de igual grado. La *secuencia de grados* de G se denota como d_G, y es un vector de tamaño $n = |V|$ que contiene los grados de cada vértice en G.

Un grafo G (V, E) cumple con la condición de *k-grado* anónimo si la secuencia de grados del grafo G, llamada d_G, es *k-anónimo*. El costo que conlleva el proceso de hacer un grafo anónimo se define como $G_A(AG, G) = |E_{AG}| - |E_G|$.

Los objetivos planteados son: 1) encontrar el *k-grado* anónimo para el grafo; 2) minimizar el costo G_A; 3) mantener una estructura similar de G en AG ($V_{AG} = V_G$).

Se asume el manejo de grafos simples, es decir sin dirección, peso, y donde no se permiten las conexiones de un vértice a sí mismo, ni múltiples conexiones entre un par de vértices. Para minimizar el número de conexiones adicionales se persigue minimizar la distancia L_1 entre la secuencia de grados de G y AG; donde $L_1(d_{AG} - d_G) = \sum_i |d_{AG}(i) - d_G(i)|$, ya que $|E_{AG}| - |E_G| = \frac{1}{2} L_1(d_{AG} - d_G)$. El modelo permite cierta flexibilidad al formar el grafo anónimo, llamada versión "relajada", la cual no cumple estrictamente la condición de igualdad en la estructura, y se conforma con que sea similar. De esta manera la intersección del conjunto de conexiones es: a) $E_{AG} \cap E_G = E_G$ en la versión estricta; y b) $E_{AG} \cap E_G \approx E_G$ para la versión relajada. De esta observación se deriva el siguiente algoritmo:

1. A partir de la secuencia de grados original d_G, se construye una nueva secuencia de grados d' que sea *k*-anónima, minimizando al mismo tiempo el costo $L_1(d' - d_G)$
2. Dada la nueva secuencia de grados d', se construye un grafo AG (V, E) tal que $d_{AG} = d'$, $V_{AG} = V_G$ y $E_{AG} \cap E_G = E_G$ ($E_{AG} \cap E_G = \approx E_G$ para la versión relajada)

El primer paso es resuelto por un algoritmo de programación dinámica de tiempo lineal, mientras que en el segundo se aplica un conjunto de algoritmos de construcción de grafos [11]. En pruebas realizadas en redes sociales con datos reales y sintéticos, se demuestra que los algoritmos son eficientes y preservan la utilidad del grafo mientras satisfacen la condición de *k-grado* anónimo. En las conclusiones también se enfatiza lo complicado de medir con exactitud el grado de información perdida puesto que no existen métricas efectivas para tal problema.

V. CONCLUSIONES

En este documento describimos los métodos y algoritmos que han sido propuestos para anonimizar redes sociales, y los ataques con los que se puede descubrir la identidad de los usuarios que la componen. De acuerdo con los resultados revisados, podemos concluir que los protocolos de anonimizacion propuestos hasta ahora consideran escenarios muy específicos, y que por tanto no pueden aplicarse para la protección de redes sociales genéricas.

Pudimos notar que para efectuar un ataque activo es necesario que la red permita agregar vértices en los lugares escogidos por el adversario, algo que puede no ser realista en muchos casos prácticos.

Una tarea prioritaria es sin duda, desarrollar métricas que permitan evaluar el grado de información perdida en el proceso de anonimización, de forma que se puedan establecer compromisos entre el grado de protección de los vértices y la utilidad de la versión anónima de la red.

REFERENCIAS

[1] J. Imizcoz, *Introducción actores sociales y redes de relaciones: reflexiones para una historia global.* Bilbao: Universidad del Pais Vasco, 2001, pp. 19-30.

[2] Joshua R. Tyler, Dennis M. Wilkinson, Bernardo A. Huberman, Email as spectroscopy: automated discovery of community structure within organizations, in *Proceedings of Communities and technologies*, 2003, pp. 81-96.

[3] Culotta, R. Bekkerman, A. McCallum. Extracting social networks and contact information from email and the web,in *Proceedings of CEAS-1*, 2004.

[4] Van Alstyne, M. and Zhang, J. 2003. "EmailNet: automatically mining social networks from organizational email communications", in *Proceedings of Annual Conference of the North American Association for Computational Social and Organizational Sciences* (NAACSOS'03), Pittsburg, PA, 2003.

[5] A. Anderson, M. Corney, O. de Vel, and G. Mohay. *Identifying the Authors of Suspect E-mail*, Communications of the ACM, 2001

[6] A. McCallum, X. Wang, and A. Corrada-Emmanuel. *Topic and Role Discovery in Social Networks*, Journal of Artificial Intelligence Research 30, 2007, pp. 249-272.

[7] L. Backstrom, C. Dwork, and J. Kleinberg. Wherefore art thou R3579X?: Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*, pp. 181–190, Alberta, Canada, May 2007.

[8] B. Zhou and J. Pei. Preserving privacy in social networks against neighborhood attacks, in *Proceedings of the 24th International Conference on Data Engineering (ICDE'08)*, Cancun, Mexico, April 2008.

[9] K. Frikken and P. Golle. Private social network analysis: How to assemble pieces of a graph privately. In *Proceedings of the 5th, ACM Workshop on Privacy in Electronic Society (WPES'06)*, pp. 89–98, Alexandria, VA, 2006.

[10] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava. "Anonymizing social networks". Technical report, University of Massachusetts Amherst, 2007.

[11] K. Liu and E. Terzi. Towards identity anonymization on graphs, in *Proceedings of ACM SIGMOD*, Vancouver, Canada, June 2008.

[12] A. Acquisti, and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the Facebook, in *Proceedings of 6th Workshop on Privacy Enhancing Technologies* (pp. 36-58).

[13] Ed, Giles Hogben, *Security Issues and Recommendations in Online Social Networks*, ENISA Position Paper, Oct 2007.

[14] Privacy and Human Rights 2006. An International Survey of Privacy Laws and Developments. EPIC. (http://www.privacyinternational.org/)

[15] http://www.rsf.org/article.php3?id_article=10615

URSI 2008

XXIII Simposium Nacional de la Unión Científica Internacional de Radio

Universidad Complutense de Madrid

22 - 24 Septiembre 2008

Artículos

Seguridad en las Comunicaciones I

Sesión VIII: Miércoles 24, 15:30 - 17:15 h Aula 7

Cifrado basado en la identidad con tarjetas de circuito integrado

José de Jesús Ángel Ángel, Guillermo Morales Luna

Smartcards are able to provide secure communications in environments in which a central agent, say a Treasury entity, establishes a cryptographic platform for secure communications, say citizens community. Here we review Shailaja protocols for ciphering and authentication using smartcards and IBE, and we recall the main characteristics for elliptic curves suggested by Scott in order to obtain robust and efficient implementations of IBE on smartcards. We detail an elementary extension of the protocols for joint access to secure communications.

Sistemas anónimos en escenarios globales

Rodolfo Leonardo Sumoza Matos, Luis Javier García Villalba

The anonymous systems' implementation, from a practical point of view, still possesses a set of unresolved issues. The zones (countries or organizations) that censor and block the communications is one of them. This work mentions the implications related to the worldwide anonymous systems' implementation, and it proposes to use the reputation's systems and trust to manage and reach anonymous comunications in a large or global level.

Redes sociales: retos, oportunidades y propuestas para preservar la privacidad

Alejandra Silva Trujillo, Luis Javier García Villalba

Social networks sites are one of the biggest technological phenomena. The importance of these sites is because industry, entities and individuals have adopted them to share their emotions, feelings, interests, ideologies. There are several risks to expose a huge amount of private information. The combination of data mining methods and analysis social networks could lead to perform networking viral strategies. In this paper first we describe the privacy importance in social networks and then we present the state of the art of several proposals released recently in this topic.

Redes sociales: retos, oportunidades y propuestas para preservar la privacidad

Alejandra Silva, Luis Javier García Villalba

Grupo de Análisis, Seguridad y Sistemas (GASS)
Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática, Despacho 431
Universidad Complutense de Madrid (UCM)
C/ Profesor José García Santesmases s/n
Ciudad Universitaria, 28040 Madrid
E-mail: {asilva, javiergv}@fdi.ucm.es
URL: http://gass.ucm.es

***Abstract*- Social networks sites are one of the biggest technological phenomena. The importance of these sites is because industry, entities and individuals have adopted them to share their emotions, feelings, interests, ideologies. There are several risks to expose a huge amount of private information. The combination of data mining methods and analysis social networks could lead to perform networking viral strategies. In this paper first we describe the privacy importance in social networks and then we present the state of the art of several proposals released recently in this topic.**

I. INTRODUCCIÓN

El ser humano es una especie que gusta de compartir emociones, sentimientos, ideologías, las formas para hacerlo han variado a lo largo de la historia. Las formas de comunicación a través de símbolos, imágenes, señas, lenguaje hablado, etc., no son más que una manera de expresión. Actualmente en la sociedad digital en que vivimos existen muchas herramientas que nos permiten interactuar con nuestros similares en cuestión de segundos no importando la localización geográfica de nuestros interlocutores. Las redes sociales en línea se han convertido en una herramienta vital para la industria, entidades ó individuos que gustan de mantener contacto con los miembros de la red social a la que pertenecen. El número de usuarios en plataformas como Facebook, MySpace, Friendster, por mencionar algunos, se ha incrementado a niveles insospechados. La industria del mercadeo o la publicidad parecen ser el mayor disparador de estos sitios dada la importante derrama económica que proporcionan. Ante este fenómeno, existe un gran riesgo al dejar expuesta la gran cantidad de datos personales que estos sitios contienen, tales como: fecha de nacimiento, número de teléfono, domicilio, nombre de la escuela a la que asisten, nombres de amigos, familiares, entre otros. Ya sea por la vigilancia de propios y extraños, ó de la industria de la mercadotecnia, con los sitios de redes sociales podemos observar un deterioro masivo en la privacidad de los usuarios que las conforman.

El presente trabajo tiene como objetivo contribuir con la discusión acerca de la privacidad de los sitios *web* de redes sociales, tema que sin lugar a dudas ha estado ausente en las mesas de trabajo de investigadores. Es por ello que presentamos un estudio del estado del arte de las propuestas que se han planteado recientemente en este ámbito. En la siguiente sección describimos la importancia de la privacidad en los sitios de redes sociales. La sección III abordamos las aplicaciones en redes sociales, cuáles son los riesgos de privacidad y describimos una forma de clasificar sus áreas sensibles. En la sección IV se introducen varias propuestas de solución, y finalmente en la sección V se exhiben las conclusiones.

II. IMPORTANCIA DE LA PRIVACIDAD EN LAS REDES SOCIALES

Los sitios de redes sociales se han convertido en el blanco perfecto para diferentes áreas de estudio al sacar provecho de las características que proporcionan [1], las cuáles son: a) los usuarios otorgan voluntariamente información acerca de ellos mismos; b) se puede confiar en la veracidad de los datos (por ejemplo en redes cuyo propósito es buscar empleo); y c) las redes son visibles al ejecutar un análisis simple de interacciones en la red. Por lo que, al combinar técnicas de minería de datos y de análisis de redes se pueden ejecutar estrategias de mercadeo o publicidad viral [2], [3], [4], cuyo resultado deriva en la identificación de aquellos grupos o individuos idóneos, que puedan ser agentes potenciales para dispersar información de productos a sus amigos y conocidos. De esta forma, se puede aplicar una de las primeras técnicas del mercadeo: la publicidad de boca en boca.

Compañías de entretenimiento, telefonía móvil, bebidas, marcas comerciales para jóvenes están notando la oportunidad de mercado que ofrecen las redes sociales en linea; por ejemplo Wal-Mart lanzó en Facebook el grupo *Roommate Style Match* [5] esperando atrapar la atención de estudiantes universitarios. Otros productos comerciales se caracterizan por tener sus propios grupos con información de sus productos, eventos especiales y promociones. Las campañas presidenciales en los Estados Unidos de América también han hecho uso de las redes sociales al intentar reclutar voluntarios, recaudar fondos y atraer la atención y preferencia de jóvenes votantes. Cada diez minutos sitios como [6] buscan párrafos publicados en redes sociales para

ubicar las frases como "*I feel*" o "*I am feeling*", de ahí extraen el nombre del autor de dicha frase, e identifican su página de perfil, donde se describe la edad, género, país, estado, ciudad; con estos tres últimos datos y basándose en la hora que se publicó la frase pueden observar las condiciones del clima para esa ciudad. Al ejecutar la búsqueda de frases que revelen sentimientos en redes sociales como MySpace, MSN, blogger, entre otros, se generan registros a una extensa base de datos donde se incluyen los datos asociados al sentimiento, y la foto que fue publicada. Y, aunque en las políticas de privacidad de dicho sitio se menciona que no se asocia el nombre de los individuos con los sentimientos que expresan, sí se coloca el URL al blog desde el cuál se obtuvo la frase.

Por otro lado algunos estudios recientes [7] presentaron que la mayoría de los usuarios en sitios de redes sociales son jóvenes y niños (entre 6 y 17 años). De aquí se desprende la problemática en relación a la supervisión de padres de familia en las actividades en línea, la cuál tiende a ser limitada por diversos factores, entre ellos, la dificultad de entender las políticas de privacidad empleadas en estos sitios, así como también la barrera generacional que implica que los jóvenes y niños se les facilita el uso de las tecnologías al haber crecido con ellas.

III. Aplicaciones en redes sociales

Facebook es un sitio web de redes sociales muy popular. Los creadores de este singular sitio, cuyo origen fue cultivar las relaciones entre la comunidad estudiantil de Harvard no sospecharon que su creación iba a convertirse en el gigante que es actualmente. En mayo del 2007 Facebook lanzó su plataforma de aplicaciones gratis. A partir de este lanzamiento se invitó a la comunidad de desarrolladores a crear todo tipo de aplicaciones. Posteriormente, unos meses después se presentó OpenSocial, formado a partir de una alianza entre Google, MySpace, Bebo y muchas otras redes sociales, con la intención de promover un conjunto de estándares común para desarrolladores de software que permitiera compartir características sociales y portables a aplicaciones y redes sociales dentro de cualquier sitio web.

A. Riesgos de privacidad en aplicaciones

Tanto Facebook como OpenSocial, las más grandes plataformas en redes sociales, otorgan acceso a los desarrolladores a las funciones base y a la información que poseen, como: 1) Información del perfil (datos del usuario), 2) Información del segundo nivel de conexión (amigos), 3) Flujos de actividades. La información sensible que circula por estas aplicaciones está siendo reutilizada y compartida como parte de Interfaces de programación de aplicaciones (API) y plataformas sociales. Debido a la expansión de audiencia y a la disponibilidad de recursos tecnológicos se está incrementando el número de aplicaciones en redes sociales que ofrecen servicios variados, y cuya similitud radica en sacar el máximo provecho de la información personal que recolectan. Las plataformas de aplicaciones dan a los desarrolladores acceso a datos que de otra manera no estaría disponible a ellos a través de la interfaz de usuario. Los usuarios no pueden añadir una aplicación a su perfil sin a la vez, otorgar el permiso para el acceso a sus datos.

B. Áreas sensibles en redes sociales

El éxito de las aplicaciones sociales ha generado diversos planteamientos con el propósito de reconsiderar cómo deberían ser configuradas, propagadas, compartidas y reutilizadas las propiedades de datos y privacidad en las comunidades virtuales. A continuación se muestran diferentes áreas que pueden comprometer la privacidad de los usuarios de redes sociales [8] [1]: i) Falta de control sobre los flujos de actividad (*activity stream*); ii) Conexión no deseada; iii) Desanonimización a través de la combinación de redes sociales; iv) Revelación de información por otro usuario.

Un flujo de actividades es una colección de eventos asociados con un solo usuario. Estos eventos podrían incluir cambios que el usuario hizo a su perfil, si el usuario agrega o ejecuta una aplicación en particular, si comparte nuevos asuntos, o si se comunica con uno de sus amigos. Los flujos de actividad de un usuario son visibles por lo regular por sus amigos, aunque varía dependiendo de la red social. Lo anterior impacta en la privacidad del usuario puesto que no conoce con exactitud a quiénes se les da a conocer sus actividades, ni cuáles son los eventos que se registran en los flujos.

La conexión no deseada ocurre cuando existen vínculos en Internet de un usuario con una entidad o persona que con la que no desea ser relacionado. Obviamente esta área no está limitada exclusivamente a redes sociales.

Por otro lado, es posible descubrir la identidad de un usuario al comparar la información disponible en diversos sitios de redes sociales, aún cuando la información es parcialmente modificada. Al comparar ciertos datos como fecha de nacimiento, libros y películas favoritas, etc., se puede adivinar la identidad de un usuario que gusta de proporcionar información en diversos sitios.

En las redes sociales en internet el concepto de amistad no es simétrico. Por ejemplo, Alicia puede publicar que es amiga de Beto o peor aún, sus preferencias, intereses u otra información personal, sin la autorización de Beto. Por lo que está información puede ser utilizada para revelar su identidad.

IV. Soluciones propuestas

En la búsqueda de soluciones debemos identificar: Primero, los conflictos potenciales de privacidad que surgen de la interacción de la red social. Para ello, es necesario un análisis de requerimientos, que incluye métodos de resolución de conflictos. Segundo, las preferencias de privacidad y los requerimientos deben ser formalizados de tal manera que la aplicación pueda detectar problemas, alertar y ayudar al usuario. Esto debe realizarse a través de la adopción y concepción de la infraestructura tecnológica y legal de las Tecnologías que incrementan la Privacidad (PET´s por sus siglas en inglés), los protocolos de privacidad (P3P y APPEL/XPref) y la legislación existente relacionada.

A. Análisis de requerimientos

Para identificar los conflictos en los sitios de redes sociales en se propone utilizar el método de Análisis de Requerimientos Multilaterales de Seguridad (MRSA, por sus siglas en inglés) [1]. El objetivo de este método es considerar los intereses ó necesidades de seguridad y privacidad de todos los *stakeholders* relacionados al sistema y desarrollar

mecanismos para negociar entre ellos. De esta manera se forma una lista compleja donde se describen los requerimientos de privacidad desde el punto de vista de cada *stakeholder*. Los resultados pueden derivar a inconsistencias, repeticiones y conflictos, para ello es necesaria una adecuada administración de requerimientos. Finalmente para llegar a una negociación entre requerimientos se sugieren los siguientes métodos: Relajación, Refinamiento, Mutua concesión, Reestructuración, entre otros. Una vez que se han analizado los requerimientos e identificado los conflictos entre ellos, los desarrolladores puedan establecer mecanismos que les ayuden a satisfacer los requerimientos operacionales en los sitios de redes sociales.

B. Requerimientos de diseño de redes sociales y sus aplicaciones

Para resolver de la falta de control en los flujos de actividades el usuario debe conocer explícitamente y tener control de cada evento que se registrará en su flujo de actividades. Así como también de la lista de usuarios que pueden ver sus flujos de actividad. Hay que considerar las expectativas del usuario y construir las aplicaciones de tal manera que actúen como el usuario espera [8].

Se propone extender la clasificación común de niveles de confidencialidad, como se describe a continuación:

- Datos privados. No se pueden revelar a menos que haya un consentimiento explícito por el usuario.
- Datos de grupo. Están disponibles a los que están en el mismo grupo que el usuario.
- Datos de comunidad. Están disponibles a los usuarios en línea y registrados. No se permiten visitantes anónimos.
- Datos públicos. Tienen acceso todos los visitantes de la red social, incluyendo visitantes anónimos.

C. Herramientas de detección de inferencias

Para la segunda y tercer área se propone la construcción de herramientas similares. En el caso de las conexiones no deseadas, se propone desarrollar una aplicación de descubrimiento automático de enlaces, es decir, cuando el usuario cree un contenido en internet, la herramienta deberá indicar qué información podría relacionar al usuario con los perfiles de otros sitios. El usuario en base a dicha información puede decidir si publicarlo o no. En el caso de la combinación de redes sociales se propone la utilización de métodos de detección de inferencias. Por ejemplo, una herramienta de detección de inferencias basada en web podría enviar un mensaje de alerta al usuario que la combinación de ciertas palabras que está por publicar podría dar lugar a aparecer en búsquedas que dejarían al descubierto su identidad.

D. Plataforma de preferencias de privacidad (P3P)

Al considerar el problema de la información personal de un usuario revelada por otros usuarios de la misma red, se debe reforzar las políticas de privacidad. Para ello se utiliza el concepto de red semántica que se basa en la idea de añadir metadatos semánticos, para que sea posible evaluarlos automáticamente por máquinas de procesamiento y que permitan informar al usuario de los efectos de publicar determinada información. P3P es un protocolo diseñado para informar a los usuarios que datos serán almacenados, cómo serán usados, por cuánto tiempo serán almacenados en los sitios que el usuario visita. Cuando las políticas del usuario establecidas en P3P no coinciden con las que el servidor del sitio donde navega, P3P informa al usuario y pregunta si desea continuar en el sitio bajo el entendido de que existe un riesgo al permanecer en él. Sin embargo, el aplicar P3P en redes sociales puede dejar varios huecos respecto a la privacidad deseada, por lo que se recomienda ampliar el protocolo P3P que permita codificar la inferencia de datos que pueda resultar en brechas a nivel de confidencialidad.

E. Interfaz de programación de aplicaciones

En [9] se detalla el desarrollo de una Interfaz de Programación de Aplicaciones (API) llamada *Privacy-by-proxy* que intenta preservar la privacidad de los usuarios de redes sociales, así como también mostrar la información de amigos de tal manera que no se degrade la funcionalidad de las aplicaciones. Después de analizar 150 aplicaciones de Facebook, se pudo determinar que la mayoría podrían mantener su funcionalidad utilizando una interfaz limitada que solo proporcionara acceso a una red social anónima. Para llevar a cabo el modelo *privacy-by-proxy* se requiere que el número ID que es necesario para identificar al usuario en cada aplicación este encriptado. A través de una función de encriptación simétrica donde las llaves son el número ID de la aplicación y una llave secreta almacenada en el servidor. Para solicitar el perfil de usuario y los datos de sus amigos, todos los usuarios primero son encriptados, por lo tanto la aplicación solo tiene acceso a una red social anónima. Para prevenir la desanonimización se limita que la aplicación solo puede mostrar información pública de aquellos IDs que pertenecen al la lista de contactos del usuario actual. Una de las ventajas es que esta al tener controlada la salida de las aplicaciones de terceras partes, *privacy-by-proxy* utiliza nuevas etiquetas y transforma los datos sin necesidad de cambiar su arquitectura.

V. CONCLUSIONES

Las áreas de oportunidad alrededor de los sitios de redes sociales son bastas, dado que es un tema de reciente interés las herramientas que se describieron a lo largo del presente trabajo son sólo el comienzo de un conjunto de herramientas que deben considerar diversas arquitecturas de redes sociales. Es necesario establecer mejores mecanismos para proteger la privacidad de los usuarios en los sitios de redes sociales tomando en consideración los requerimientos de los múltiples *stakeholders*. Usuarios, operadores de redes sociales y empresas (industria de publicidad ó mercadeo), que están detrás de estos sitios, deben considerar la importancia de direccionar adecuadamente los requerimientos de privacidad presentes, ya que esto permitirá garantizar la permanencia y preferencia al incrementar el nivel de confianza en dichos sitios. Sin lugar a dudas uno de los mayores retos para las técnicas de detección de inferencias es construir reglas semánticas considerando el lenguaje coloquial y abreviaciones y códigos de escritura. Por otro lado para la Interfaz de programación de aplicaciones propuesta podemos notar que está desarrollada para un escenario específico, por lo que su aplicación en otros sitios de redes sociales pueda no encajar con lo esperado.

No podemos negar que aún quedan muchos retos por resolver, y que presenciaremos en el futuro mejoras y abusos en las tecnologías de ésta área de interés.

AGRADECIMIENTOS

Los autores agradecen la financiación que les brinda el Ministerio de Ciencia e Innovación a través del Proyecto TEC2007-67129/TCM y el Ministerio de Industria, Turismo y Comercio a través del Proyecto Avanza I+D TSI-020100-2008-365.

REFERENCES

[1] S. Preibusch, B. Hoser, S. Gürses and B. Berendt, "Ubiquitous social networks – opportunities and challenges for privacy aware user modelling", in *Proc. of the Data mining for user modelling workshop*, 2007.

[2] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing", in *Proc. of the Eighth Intl. Conf. on Knowledge Discovery and Data mining (SIGKDD'02)*, 2002.

[3] J. Leskovec, L.A. Adamic and B. Huberman, *The Dynamics of viral marketing*, ACM Transactions on the Web, 2007.

[4] S. Hill, F. Provost and C. Volinsky, "Learning and Inference in Massive Social Networks", in *Proc. of the 5th International Workshop on Mining and Learning with Graphs*, August 2007.

[5] http://www.foxnews.com/story/0,2933,292757,00.html

[6] We feel fine Project. http://www.wefeelfine.org

[7] J. Bailey, "A Report on ´Terra incognita': The 29th International Conference of Data Protection and Privacy Commissioners", September, 2007. http://www.privacyconference2007.gc.ca/workbooks/Terra_Incognita_summary_E.html

[8] M. Chew, D. Balfanz and B. Laurie, "(Under)mining Privacy in Social Networks", in *Proc. of Web 2.0 Security and Privacy 2008*, May 2008.

[9] A. Felt and D. Evans, "Privacy Protection for Social Networking Platforms", in *Proc. of Web 2.0 Security and Privacy 2008*, May 2008.

Actas

XII Reunión Española sobre Criptología y Seguridad de la Información

Donostia-San Sebastian
2012

Editores:
U. Zurutuza
R. Uribeetxeberria
I. Arenaza-Nuño

4–7 Septiembre, 2012

Edita:
Servicio Editorial de Mondragon Unibertsitatea
http://recsi2012.mondragon.edu
Mondragon Unibertsitatea
Loramendi, 4. Apartado 23
20500 Arrasate - Mondragon

ISBN: 978-84-615-9933-2

1^a Edición: Julio de 2012

Ataque de Revelación de Identidades en un Sistema Anónimo de Correo Electrónico

Javier Portela García-Miguel[1], Delfín Rupérez Cañas[2], Ana Lucila Sandoval Orozco[2],
Alejandra Guadalupe Silva Trujillo[2], Luis Javier García Villalba[2]

[1] Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Estadística e Investigación Operativa III
Escuela Universitaria de Estadística, Despacho 721, Universidad Complutense de Madrid (UCM)
Avenida Puerta de Hierro s/n, 28040 Madrid
E-mail: jportela@estad.ucm.es

[2] Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM)
Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid
Email: {delfinrc, asandoval, asilva, javiergv}@fdi.ucm.es

Resumen*—El objetivo de nuestro trabajo es desarrollar un ataque global de tipo SDA** ***(Statistical Disclosure Attack) **que permita identificar las relaciones entre usuarios de una red basada en un sistema anónimo de** ***mixes*****. Nuestros escenarios son más generales en relación a otros ataques SDA. Asimismo presentamos un nuevo esquema teórico de modelado basado en tablas de contingencia. Proporcionamos soluciones para todos los usuarios simultáneamente, debido a que la dependencia de los datos no posibilita centrarse en usuarios específicos sin tener en cuenta las posibilidades combinatorias. A diferencia de simulaciones desarrolladas sobre este mismo tema, este trabajo ha sido desarrollado con datos reales de una aplicación de correos electrónicos, tomando en consideración las propiedades especiales de las redes de comunicación establecidas entre usuarios reales.**

I. Introducción

En las redes de comunicación los *mixes* proporcionan protección contra potenciales observadores al ocultar la apariencia de mensajes, patrones, longitud y relación entre emisores y receptores. Chaum [1] propuso ocultar la correspondencia entre emisores y receptores cifrando mensajes y reordenándolos a través de un camino de *mixes* antes de enviarlos a su destino. Se han propuesto muchos otros diseños, incluidos Babel [2], Mixmaster [3] o Mixminion [4]. Las diferencias entre estos sistemas no serán abordadas en nuestro trabajo: la información que usamos sólo se relaciona a emisores y receptores que están activos en un período de tiempo y la manera con la cual se reordenan los mensajes no afecta al ataque. Otra clase de diseños de anonimato, como *Onion routing* [5] son de baja latencia y están orientados a *Web browsing* y otros servicios interactivos. Nuestro método no se enfoca en estos diseños, los cuales pueden ser tratados efectivamente con ataques con períodos cortos de tiempo o de conteo de paquetes [6]. Los ataques contra las redes de *mixes* pretenden reducir el anonimato al relacionar cada emisor y receptor con sus correspondientes mensajes enviados o recibidos, o bien relacionar emisores con receptores. Al observar la red los atacantes pueden deducir la frecuencia de las relaciones, comprometiendo los *mixes* o llaves, alterando o retrasando los mensajes. Pueden ser capaces de deducir el destino más probable de los mensajes a través de falsos mensajes enviados a la red, y utilizar esta técnica para aislar y conocer las propiedades de ciertos mensajes previamente definidos. En [7] se muestra un resumen de ataques basados en análisis de tráfico. En [8], [9], [10], [11] se trata el tema del k anonimato, situado en el contexto multidimensional. Agrawal y Kesdogan [12] presentaron el *disclosure attack*, un ataque centrado en un mix de lotes simple, cuyo objetivo es obtener información de un emisor particular Alicia. El ataque es global, en el sentido de que recaba información sobre el número de mensajes enviados por Alicia y recibidos por otros usuarios; y pasivo, ya que el atacante no puede alterar la red, por ejemplo, enviando falsos mensajes o retrasándolos. Se asume que Alicia tiene exactamente m receptores, que envía mensajes con la misma probabilidad a cada uno de sus receptores, y además, que envía un mensaje en cada lote de b mensajes. Se podrían identificar a los receptores de Alicia clasificados en conjuntos disjuntos a través de algoritmos numéricos. Danezis [13] presenta el *Statistical Disclosure Attack* (SDA), considerando las hipótesis de [12]. En el SDA los receptores se ordenan en términos de probabilidad. Alicia debe demostrar patrones de envío consistentes a largo plazo para obtener buenos resultados. En [14] se describe el SDA cuando se usa *threshold mix* o *pool mix*, considerando las hipótesis de artículos previos donde se conoce el número de receptores de Alicia, o se enfoca en un solo usuario de Alicia. El SDA de doble orientación [15] usa las posibilidades de réplicas entre usuarios. El *Perfect Matching Disclosure Attack* [16] pretende utilizar información simultánea de todos los usuarios para obtener mejores resultados en la revelación de los receptores de Alicia. Este trabajo se enfoca en el problema de obtener información de las relaciones o la comunicación entre usuarios de una red, donde se obtiene información parcial. El enfoque de modelado del algoritmo y esquema de solución

Tabla I
EJEMPLO DE TABLA DE CONTINGENCIA

Receptores	Emisores			
	U3	U4	U5	Total enviados
U1	4	0	0	4
U2	0	1	0	1
U3	0	0	2	2
Total recibidos	4	1	2	7

Tabla II
EJEMPLO DE TABLA DE CONTINGENCIA CON INFORMACIÓN DE MARGINALES

Receptores	Emisores			
	U3	U4	U5	Total enviados
U1				4
U2				1
U3				2
Total recibidos	4	1	2	7

Tabla III
EJEMPLO DE TABLA CON COTAS OBTENIDAS

Receptores	Emisores			
	U3	U4	U5	Total enviados
U1	(1,4)	(0,1)	(0,2)	4
U2	(0,1)	(0,1)	(0,1)	1
U3	(0,2)	(0,1)	(0,2)	2
Total recibidos	4	1	2	7

son aplicados en datos de correos electrónicos. Como las soluciones individuales son interdependientes, nuestro ataque no se centra en un usuario en concreto, sino que pretende obtener la máxima información de todos los usuarios. La información utilizada es el número de mensajes enviado y recibido por cada usuario. Esta información es obtenida en rondas que pueden ser determinadas por intervalos de tiempo de una longitud determinada, o alternativamente en lotes de mensajes de igual tamaño. El marco base y supuestos necesarios para desarrollar nuestro algoritmo son los siguientes:

- El atacante conoce el número de mensajes enviados y recibidos por cada usuario en cada ronda.
- La ronda puede ser determinada por el sistema (lotes) o puede basarse en intervalos regulares de tiempo donde el atacante obtiene la información adicional de los mensajes enviados y recibidos. Hemos utilizado ambos métodos en nuestras aplicaciones obteniendo ligeramente mejores resultados al utilizar lotes (batches).
- El método está restringido, por el momento, a un sistema mix simple (sin considerar los *threshold mix* o *pool mix*).
- No se plantean restricciones sobre el número de amigos de cada usuario, ni sobre el número de mensajes a enviar. Ambos se consideran desconocidos de antemano.
- El atacante controla todos los usuarios del sistema. En nuestra aplicación nos centramos en los correos electrónicos que los usuarios de un dominio envían y reciben en este dominio.

Este artículo se compone de 5 secciones, siendo la primera la presente introducción. En la sección II plantea el problema, formulándolo con un nuevo enfoque a través de tablas de contingencia. Se presentan cotas y otras técnicas básicas de obtener información sobre el número de mensajes que envía cada usuario. La sección III explica el algoritmo propuesto, y detalla cómo puede obtenerse información relevante de las relaciones existentes (o no existentes) entre usuarios. La sección IV presenta la aplicación del algoritmo a datos reales. Finalmente, en la sección V se presentan las conclusiones sobre los resultados, y se plantean limitaciones y trabajos futuros a desarrollar sobre este ataque.

II. EL PROBLEMA

El atacante obtiene información de cuántos mensajes envía y recibe cada usuario en cada ronda. Normalmente el conjunto de emisores y receptores no es el mismo, aún cuando algunos usuarios puedan ser emisores y receptores en alguna ronda en particular. Además, el número total de usuarios en el sistema N no está presente en cada ronda, pues solo una fracción de ellos está recibiendo o enviando mensajes. En la Figura 1 se muestra una posible ronda, que por razones pedagógicas se compone de un mínimo de usuarios.

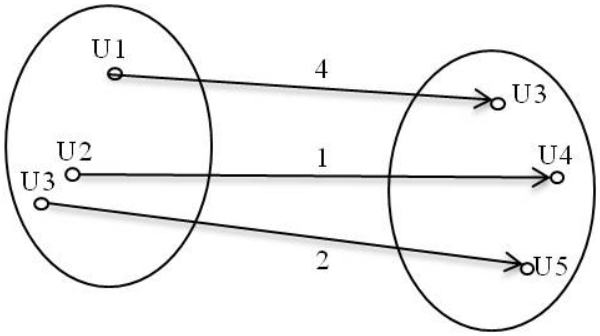


Figura 1. Relación entre emisores y receptores.

La información de esta ronda se puede representar en una tabla de contingencia (vea la Tabla I), donde el elemento (i, j) representa el número de mensajes enviados del usuario i al usuario j. El atacante solamente ve la información presente en las marginales agregadas donde, las filas representan el número de mensajes enviados por cada usuario, y las columnas, el número de mensajes recibido por cada usuario, según aparece en la Tabla II. Por medio de los valores marginales es posible obtener información importante. Las cotas de los elementos pueden ser útiles, ya que nos pueden proporcionar relaciones directas entre usuarios. Las cotas de Fréchet sobre tablas de contingencia son muy conocidas en estudios de revelación [19]. Denotando con n_{ij} el contenido del elemento (*i*,*j*), n_{i+} el valor marginal de la fila *i*, n_{+j} el valor marginal de la columna *j* y n el total. Las cotas de Fréchet se establecen como se muestra en la ecuación 1. Por ejemplo, partiendo de la Tabla II, se obtienen las cotas presentadas en la Tabla III.

$$max(n_{i+} + n_{+j} - n, 0) \leq n_{ij} \leq \text{mín}(n_{i+}, n_{+j}) \quad (1)$$

III. EL ALGORITMO

El objetivo principal del algoritmo que proponemos es extraer información relevante sobre las relaciones (o no relaciones) entre cada par de usuarios. Esta información puede ser obtenida en forma de reglas (0=relación, 1=no relación) o como probabilidades estimadas de relación, Otras fuentes de información obtenidas que pueden ser utilizadas

son la distribución estimada de mensajes del usuario *i* al *j* por unidad de tiempo, y la media estimada de mensajes de *i* a *j* por unidad de tiempo. La información obtenida por el atacante son las sumas marginales, por fila y columna, de cada una de las rondas $1, ..., T$ donde T es el número total de rondas. Hay que observar que en cada ronda la dimensión de la tabla es diferente, pues no tomamos en cuenta a usuarios que no envían (marginal de la fila=0) ni reciben (marginal de la columna=0) mensajes. Decimos que un elemento (*i*, *j*) está "presente" en una ronda si las marginales correspondientes no son cero. Esto significa que el usuario *i* está presente como emisor y el *j* como receptor. Se puede construir una tabla final A resumiendo todas las rondas y obteniendo una tabla con todos los mensajes enviados y recibidos por cada usuario en el intervalo de tiempo total considerado para el ataque. Cada elemento (*i*, *j*) de esta tabla final representaría el número total de mensajes enviados de *i* a *j*. Aunque la información obtenida en cada ronda es más precisa y relevante, un estimado exacto de la tabla A sería el principal objetivo ya que por ejemplo, un cero en la celda (*i*, *j*) y en la celda (*j*, *i*) significaría no relación entre los usuarios *i* y *j*. Mientras que un valor positivo indicaría que algún mensaje ha sido enviado en alguna ronda. Se presenta un algoritmo para generar tablas factibles (tablas cuyas sumas marginales en cada fila y columna coinciden con los valores marginales conocidos por el atacante).

Algoritmo 1

1. Comenzar con la columna 1, fila 1: generar n_{11} de una distribución uniforme entera en las cotas de la ecuación 1 donde $i = 1, j = 1$.
2. Para cada elemento n_{k1} en esta columna, si los elementos del renglón hasta $k-1$ se han obtenido, se calculan nuevas cotas para n_{k1} a partir de la ecuación 2.

$$\max\Big((0, (n_{+1} - \sum_{i=1}^{k-1} n_{i1}) - \sum_{f=k+1}^{r} n_{i+}\Big) \leq n_{ij} \leq \min(n_{k+}, n_{+j} - \sum_{f=1}^{k-1} n_{i1}) \qquad (2)$$

   El elemento n_{k1} se genera entonces según un entero uniforme.
3. EL último elemento de la fila se rellena automáticamente dado que las cotas superior e inferior coinciden, haciendo $n_{(k+1)+} = 0$ por conveniencia.
4. Una vez que una columna está rellena, las marginales por fila n_{i+} y el valor N se actualizan por substracción de los elementos ya calculados, y el resto de la tabla se trata como una tabla nueva con una columna menos.
   El algoritmo calcula columna a columna hasta tener toda la tabla llena.

El tiempo empleado depende de la complejidad del problema (número de elementos, número promedio de mensajes). Para tablas grandes, toma menos de 3 minutos obtener un millón de tablas factibles, es decir aún cuando el numero de datos de correos electrónicos sea alto, no representa problema alguno. Repetir el algoritmo como está escrito para cada tabla no proporciona soluciones uniformes, porque algunas tablas son más probables que otras debido al orden utilizado al rellenar filas y columnas. Como debemos considerar a priori todas las soluciones igualmente posibles para una ronda determinada, se realizan dos modificaciones adicionales: i) Antes de generar soluciones se reordenan aleatoriamente las filas y columnas de la tabla; ii) Una vez que se generan todas las tablas deseadas, solo se conservan aquellas que son diferentes entre sí. Estas dos modificaciones han significado una mejora muy importante en los resultados de nuestro ataque. Decidir el número de tablas a generar plantea un problema interesante. Calcular el número de tablas factibles distintas en una tabla de contingencia con marginales fijos es todavía un problema abierto, que ha sido abordado a través de: métodos algebraicos, que son poco prácticos incluso para dimensiones moderadas, y por aproximaciones normales, que dan malos resultados con matrices dispersas, con muchos ceros y valores bajos, que es justo el tipo de matriz en nuestras aplicaciones. Hasta ahora la mejor aproximación para estimar el número de tablas factibles es utilizar las tablas generadas. Un estimado del número de tablas puede ser obtenido al promediar sobre las tablas generadas el valor $\frac{1}{q(T)}$ [20]. El número de tablas factibles va desde valores moderados que son fácilmente abordados como 100,000 obteniendo todas las tablas por simulación, hasta números tan altos como 10^{13}. Generar todas las tablas posibles para este último caso llevaría, con el ordenador que hemos usado, al menos 51 días. La razón principal por la que se complica llevar a cabo un ataque determinístico de intersección es la cantidad de tablas factibles, aún cuando las dimensiones de usuarios sean bajas o moderadas. Además, los ataques estadísticos centrados en un único usuario sin tener en cuenta las relaciones entre todos los usuarios son muy optimistas, pues la dimensión de las posibilidades es tan grande que llevaría años de comportamiento consistente de un usuario en particular para alcanzar convergencia débil. La información obtenida finalmente consiste en un número fijo de tablas factibles generadas para cada ronda. Considerando la información obtenida sobre todas las rondas, la media de cada elemento sobre todas las tablas para todas las rondas es un estimado del valor real de este elemento. La media obtenida en cada elemento y ronda se agrega sobre todas las rondas para obtener un estimado de la tabla agregada, $\widehat{A}$. Regularmente, los elementos en $\widehat{A}$ son estrictamente positivos excepto para casos triviales, debido a que es muy probable que para cada elemento exista una ronda al menos en la que el estimado sea positivo, generando con ello una media final positiva. Además, los elementos finales generados no son buenos estimados debido a que son valores medios obtenidos a partir de cotas. Es posible reescalar la matriz $\widehat{A}$ para resumir el numero total de mensajes pero los estimados siguen sin ser precisos. Por otro lado, hemos encontrado que existe una relación lineal entre los elementos estimados y los valores reales.

Para obtener información relevante sobre las relaciones es necesario fijar los elementos cero más probables. Para cada

elemento, se estima la probabilidad de cero. Esto se hace calculando el porcentaje de tablas con ese elemento cero para cada ronda que el elemento está presente, y multiplicando las probabilidades obtenidas para todas esas rondas (el elemento será cero en la tabla final si es cero en todas las rondas). Se utilizan las siguientes expresiones si calculamos las probabilidades para el elemento (*i*, *j*) y se generan M tablas por ronda:

$$log\Big(p\Big(el(i,j)=0\Big)\Big) = -N_p \log(M) + \sum_{t=1,(i,j)p}^{T} \log(n_t^{(i,j)})$$

$n_t^{(i,j)}$= N^o de tablas con elemento *(i, j)* = 0 en la ronda *t*.

N_p= N^o de rondas con elemento *(i, j)* presente.

Los elementos de la tabla final se ordenan por su probabilidad de cero, a excepción de los elementos que ya son ceros triviales (elementos que representan pares de usuarios que nunca han coincidido en ninguna ronda). Los elementos cero menos probables son considerados candidatos a "relación existente". El objetivo principal del método es detectar con precisión: 1. Celdas que son cero con alta probabilidad (no relación $i \rightarrow j$). 2. Celdas que son positivas con alta probabilidad (relación $i \rightarrow j$).

Nuestro método de clasificación consiste en seleccionar la probabilidad de un punto de corte p (valores cercanos a 0,85 han dado buenos resultados en nuestras aplicaciones) y considerar clasificadas como "celdas cero" aquellas con probabilidad de cero $> p$, en tanto se considerarán clasificadas como "celdas positivas" aquellas con probabilidad de cero $< 1 - p$. El resto de celdas se considerarán como "no clasificadas". Este es un enfoque conservador del problema de clasificación, que se utiliza cuando es importante detectar elementos que pertenecen a ciertas clases con alta probabilidad, aunque el método conlleve a elementos no clasificados. Nuestro método es simétrico debido a que el intervalo de rechazo es determinado por un único valor p (puede también ser asimétrico, si el investigador lo desea, fijando diferentes puntos de corte en cada extremo). Por lo regular, en nuestras aplicaciones el porcentaje de celdas no clasificadas es menor del 15 %.

El algoritmo lleva a un test de clasificación binaria para los elementos diagnosticados, donde 0 en un elemento (*i*, *j*) significa no relación emisor-receptor de *i* a *j*, y 1 significa relación positiva emisor-receptor de *i* a *j*. Algunas métricas características para los tests de clasificación binaria son la sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo. Consideramos TP a los verdaderos positivos, FP a los falsos positivos, TN a los verdaderos negativos y FN a los falsos negativos:

Sensibilidad = $\frac{TN}{TN+FP}$ mide la capacidad del test para reconocer valores negativos verdaderos.

Especificidad = $\frac{TP}{TP+FN}$ mide la capacidad del test para reconocer valores positivos verdaderos.

Valor predictivo positivo = $\frac{TP}{TP+FP}$ mide la precisión del test en predecir valores positivos.

Valor predictivo negativo = $\frac{TN}{TN+FN}$ mide la precisión del test en predecir valores negativos.

No hay una manera perfecta de describir esta información con solo número. Para nuestro caso, donde el tamaño de las clases difiere, debido a que la tasa de negativos (valores 0) es muy alta comparada con la de positivos, se puede utilizar el coeficiente de correlación de Matthews para evaluar el desempeño del test, MCC, que se define así:

$$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (3)$$

Este coeficiente regresa un valor entre -1 y 1. Un coeficiente de $+1$ representa una predicción perfecta, 0 una predicción aleatoria y -1 una predicción inversa.

IV. Aplicación a datos de correo electrónico

Se realizaron un gran número de simulaciones a medida que el método era desarrollado, pero las especiales singularidades de los datos de correos electrónicos eran más apropiadas par medir la confiabilidad del método. Se utilizaron como base datos proporcionados por el Centro de Computación de la Universidad Complutense de Madrid, a fin de evaluar el funcionamiento del ataque. Se obtuvo el tiempo del envío, emisores y receptores (anónimos) para cada mensaje enviado durante un lapso de 12 meses, en un dominio restringido a una Facultad. Se determinaron la longitud de las rondas y tamaño de los lotes para evaluar el método. Fueron eliminados aquellos mensajes enviados de manera evidente a listas, mensajes institucionales y mensajes que provenían o eran enviados fuera del dominio.

Como un primer ejemplo, la Tabla IV presenta los resultados obtenidos para 4 meses. Consta de un total 97 emisores y 103 receptores. La complejidad de las rondas (el número de usuarios o dimensión de las tablas en las rondas) crece a medida que crece la longitud del intervalo de ronda o tamaño del lote. La tabla final agregada A tiene 9909 elementos. El punto de corte utilizado fue $p = 0,85$.

Tabla IV
Resultados de la Simulación

Tamaño del lote	% Falsos Negativos	% Falsos Positivos	Sensibilidad	Especificidad	MCC	% No clasificado
7	5	12	0,45	0,96	0,67	21
10	10	20	0,34	0,95	0,58	22
20	4	30	0,34	0,98	0,43	14
40	4	59	0,34	0,98	0,44	13
60	3	63	0,33	0,98	0,42	13
80	3	65	0,33	0,98	0,41	13

Los resultados empeoran cuando crece el tamaño del lote, y el atacante tendrá que obtener más información (rondas) para disminuir este efecto. Mientras que, los resultados son muy buenos para tamaños de lotes pequeños y permiten revelar algunas relaciones positivas así como un gran número de relaciones no existentes con un MCC = 0,67. Está claro que aumentar el tamaño del lote obliga al investigador a aumentar el punto de corte p, para evitar un alto porcentaje de falsos positivos, que es intolerable más allá de 50 %. Un punto

de corte más alto significa como consecuencia un mayor porcentaje de celdas no clasificadas. Cuando el tamaño del lote es grande se puede utilizar un punto de corte asimétrico.

Las siguientes figuras son presentadas en orden para estudiar la sensibilidad del método a variaciones en el punto de corte p, tamaño del lote y horizonte de datos recogidos. Las Figuras 2, 3, 4 se realizan para un horizonte de 4 meses, p = 0,20 y tamaño de lote 20. El número de tablas generadas por ronda afecta la precisión del método, pero menos de lo que intuitivamente se podría sospechar. Aún con tamaños grandes de lote, que dan lugar usualmente a espacios grandes de soluciones factibles, el generar más de 50000 tablas no mejora significativamente el método (Figuras 2 y 3). Las Figuras 5, 6 y 7 se realizan para p = 0,20 y tamaño de lote 20. A medida que la información obtenida crece en número de meses, la precisión del método mejora (Figura 5). La Figura 8 se realiza para un horizonte de 4 meses, p = 0,20. El tamaño del lote afecta de manera significativa la precisión del método. Cuanto más alto sea el tamaño, los resultados son peores pues la dimensión de las tablas es mayor y por lo tanto la complejidad del problema crece (Figura 8). El método de clasificación con opción de rechazo presentado es simétrico respecto a $\alpha = 1 - p$, y por lo tanto una curva ROC no es apropiada: la sensibilidad y especificidad crecen a medida que α decrece. Pero a la vez el porcentaje de celdas no clasificadas también se incrementa. El investigador debe decidir un punto de corte adecuado que no derive en un número alto de celdas no clasificadas. La Figura 9 muestra que trazando una línea vertical en $\alpha = 0,20$ $(p = 0,80)$ nos arroja un razonable 20 % de celdas no clasificadas, con sensibilidad de 0,45 y especificidad cerca de 0,98. La Figura 9 se realiza para un horizonte de 4 meses, tamaño de lote 20.

V. Conclusiones y Trabajo Futuro

Este trabajo presenta un método para detectar relaciones (o no relaciones) entre usuarios en un entorno de comunicaciones, donde la información obtenida es incompleta. Es el primer enfoque práctico al problema de revelación de datos de correos electrónicos, y, en nuestro conocimiento, es el primer trabajo en el cual se utilizan datos reales no simulados, para evaluar el rendimiento de ataque de revelación. Los resultados son alentadores pues se obtiene una alta especificidad y una moderada o alta sensibilidad, con un rango de celdas no diagnosticadas relativamente bajo. El método puede ser aplicarse a otras escenarios, como *pool mixes*, o situaciones donde se puede utilizar información adicional. Se ha utilizado también computación paralela con buenos resultados para acelerar el método. El ataque también puede ser utilizado en otros entornos de comunicaciones como redes sociales o protocolos *peer to peer*, y a problemas reales de de-anonimización que no tienen por qué ser del dominio de las comunicaciones, como revelar tablas públicas o investigación forense. Se necesita profundizar en la investigación, en los aspectos de la selección de puntos de corte p, el número óptimo de tablas a generar o incluir mejoras en la solución final, quizá rellenando celdas iterativamente y ciclando el algoritmo.

Agradecimientos

Los autores agradecen la financiación que les brinda el Subprograma AVANZA COMPETITIVIDAD I+D+I del Ministerio de Industria, Turismo y Comercio (MITyC) a través del Proyecto TSI-020100-2011-165. Asimismo, los autores agradecen la financiación que les brinda el Programa de Cooperación Interuniversitaria de la Agencia Española de Cooperación Internacional para el Desarrollo (AECID), Programa PCI-AECID, a través de la Acción Integrada MAEC-AECID MEDITERRÁNEO A1/037528/11.

Referencias

[1] D. Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms" *Communications of ACM*, Vol. 24 No. 2, pp. 84-88, 1981.

[2] C. Gulcu and G. Tsudik, "Mixing E-mail with Babel", *in Proceedings of the Network and Distributed Security Symposium (NDSS 96)*, pp. 2-16, February 1996.

[3] U. Moller, L. Cottrell, P. Palfrader, and L. Sassaman, "Mixmaster Protocol - version 2", *IETF Internet Draft*, July 2003, http://www.abditum.com/mixmasterspec.txt.

[4] G. Danezis, R. Dingledine, and N. Mathewson, "Mixminion: Design of a Type III Anonymous Remailer Protocol", *in Proceedings of the IEEE Symposium on Security and Privacy*, pp. 2-15, May 2003.

[5] TorStatus *Tor Network Status*, 2010. http://torstatus.cyberphunk.org.

[6] A. Serjantov and P. Sewell, "Passive Attack Analysis for Connection Based Anonymity Systems", *Computer Security - ESORICS 2003*, Springer-Verlag, LNCS 2808, pp. 116-131, October 2003.

[7] J. F. Raymond, "Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems", *in Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability*, Springer-Verlag New York, Inc. pp. 10-29, 2001.

[8] M. Ercan, C. Clifton and E. Nergiz, "Multirelational *k-Anonymity", IEEE Transaction on Knowledge and Data Engineering*, Vol. 21, No. 8, pp. 1417-1421, 2009.

[9] D. Sacharidis, K. Mouratidis and D. Papadias, "*k*-Anonymity in Presence of External Databases", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 22, No. 3, pp. 392-403, 2010.

[10] S. Kisilevich, L. Rokach, Y. Elovici and B. Shapira, "Efficient Multidimensional Suppression for K-Anonymity", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 22, No. 3, pp. 334-347, 2010.

[11] G. Ghinita, P. Kalnis and Y. Tao, "Anonymous Publication of Sensitive Transactional Data", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 23, No. 2, pp. 161-174, 2011.

[12] D. Agrawal and D. Kesdogan, "Measuring Anonymity: the Disclosure Attack". *IEEE Security & Privacy*, Vol. 1, No. 6, pp. 27-34, Nov.-Dec. 2003.

[13] G. Danezis, "Statistical Disclosure Attacks: Traffic Confirmation in Open Environments", *in Proceedings of the Security and Privacy in the Age of Uncertainty, (SEC2003)*, Kluwer, pp. 421-426, 2003.

[14] G. Danezis and A. Serjantov, "Statistical Disclosure or Intersection Attacks on Anonymity Systems", *Lecture Notes in Computer Science 3200*, pp. 293-308, 2005.

[15] G. Danezis, C. Diaz and C. Troncoso, "Two-sided Statistical Disclosure Attack", *in Proceedings of the 7th International Conference on Privacy Enhancing Technologies (PET' 07) LNCS 4776*, pp. 30-44, 2007.

[16] C. Troncoso, B. Gierlichs, B. Preneel and I. Verbauwhede, "Perfect Matching Disclosure Attacks", *in Proceedings of the 8th International Symposium on Privacy Enhancing Technologies (PET' 08), LNCS 5134*, pp. 223, 2008.

[17] A. Pfitzmann and M. Kihntopp, "Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology", *LNCS 2009*, pp. 1-9, 2001.

[18] L. Willenborg and T. Waal, "Elements of Statistical Disclosure Control", *Lecture Notes in Statistics*, Vol. 155, No. 15, pp. 261, 2001.

[19] A. Dobra and S. E. Fienberg, "Bounds for Element Entries in Contingency Tables Given Marginal Totals and Decomposable Graphs", *in Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97 No. 22, pp. 11885-11892, 2000.

[20] Y. Chen, P. Diaconis, S. P. Holmes and J. S. Liu, "Sequential Monte Carlo Methods for Statistical Analysis of Tables", *Journal of the American Statistical Association*, Vol. 100, pp. 109-120, 2005.

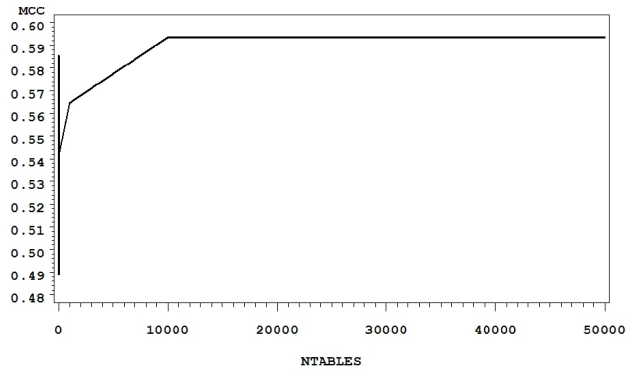


Figura 2. Coeficiente MCC vs N^o Tablas / Ronda.

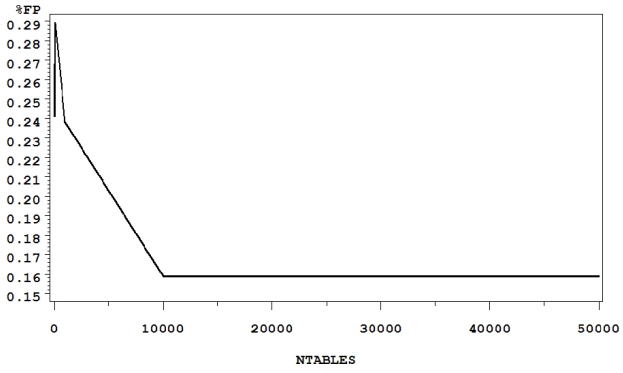


Figura 3. Tasa de Falsos Positivos vs número de tablas / Ronda.

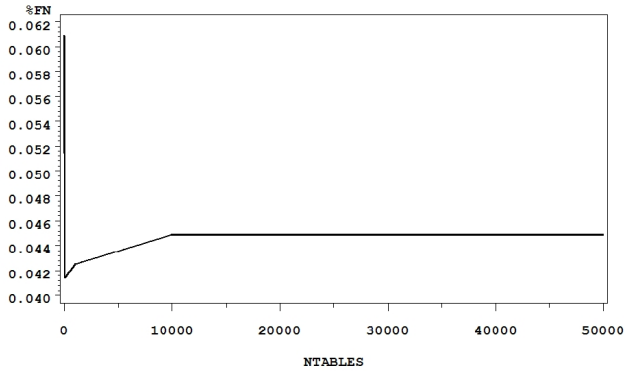


Figura 4. Tasa de Falsos Negativos vs N^o Tablas / Ronda.

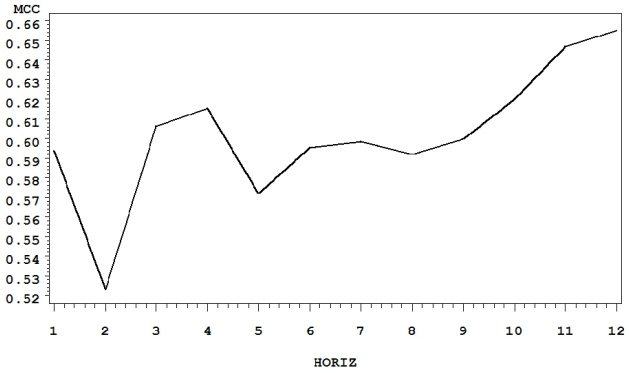


Figura 5. Coeficiente MCC vs Horizonte del Ataque.

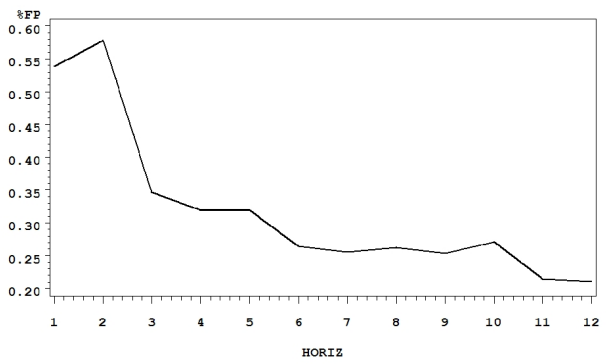


Figura 6. Tasa de Falsos Positivos vs Horizonte del Ataque.

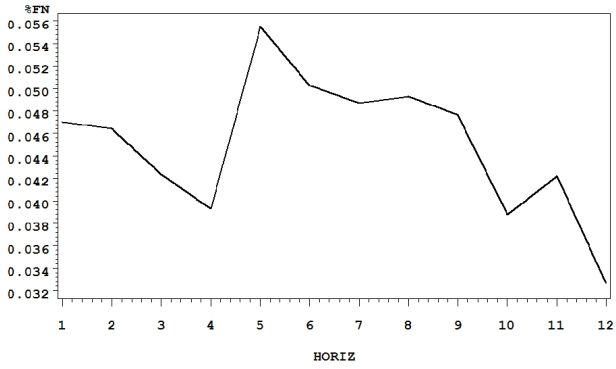


Figura 7. Tasa de Falsos Negativos vs Horizonte del Ataque.

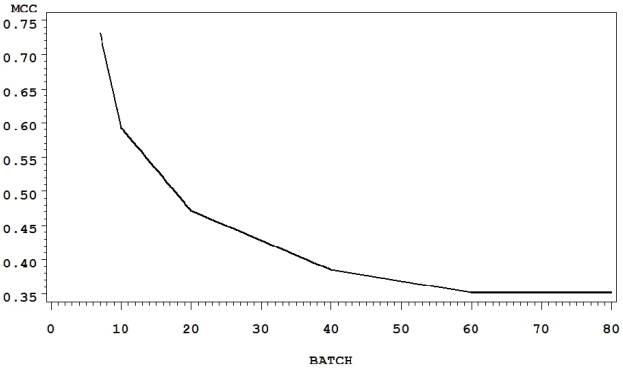


Figura 8. Coeficiente MCC Coefficient vs Tamaño de Lote.

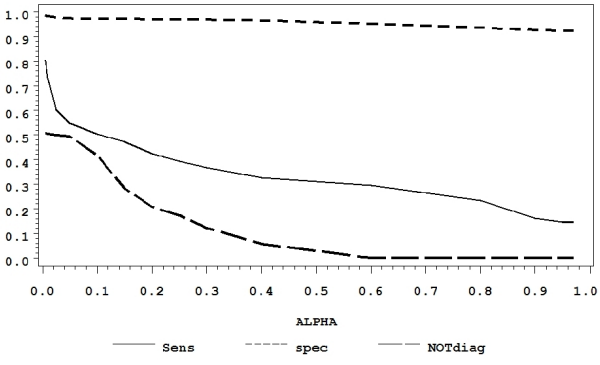


Figura 9. Sensibilidad, Especificidad y Celdas no Clasificadas vs $\alpha = 1 - p$.

DERIVATIONS OF TRAFFIC DATA ANALYSIS

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba

Group of Analysis, Security and Systems (GASS)
Department of Software Engineering and Artificial Intelligence (DISIA)
School of Computer Science, Office 431, Universidad Complutense de Madrid (UCM)
Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain
Email: {asilva, javiergv}@fdi.ucm.es, jportela@estad.ucm.es

Abstract

Public networks such as Internet do not provide a secure communications between subjects. Communication over such networks is susceptible to being compromised by unauthorized third parties. There are specific scenarios where data encryption is required: to help to protect data from being viewed, providing ways to detect whether data has been modified and offering a secure channel to communicate. In order to ensure privacy and anonymity communication researchers have developed several techniques which make possible anonymous web surfing, e-voting, report emailing and others. The aim of this paper is to present an overview of how large amounts of traffic that has been routed through an anonymous communication system can find communication relationships.

Keywords - Privacy, anonymous communications, statistical disclosure attack.

1 INTRODUCTION

Nowadays technology is an important key for our lives. Internet has become a useful tool for people to communicate and exchange data with each other. Public networks such as Internet do not provide a secure communications between subjects. Communication over such networks is susceptible to being compromised by unauthorized third parties. There are specific scenarios where data encryption is required: tohelp to protect data from being viewed, providing ways to detect whether data has been modified and offering a secure channel to communicate.In order to ensure privacy and anonymity communication researchers have developed several techniques which make possible anonymous web surfing, e-voting, report emailing and others.

In order to show the classic situation where cryptography is used, consider two subjects Alice and Bob communicate over a simple and unprotected channel. Alice and Bob want to ensure their communication will be incomprehensible to anyone who might be listening. Also, they must ensure that the message has not been altered by a third party during transmission. And, both must ensure that message comes really from Alice and not someone who is supplanting her identity.

Cryptography is used to achieve: i) Confidentiality: To help protect a user's identity or data from being read; ii) Data integrity: To help protect data from being changed; iii) Authentication: To ensure that data originates from a particular subject; iv) Non-repudiation: To prevent a particular subject from denying that he have sent a message.

The aim of this paper is to provide an overview of how large amounts of traffic that has been routed through an anonymous communication system can be mined in order to find communication relationships.

2 MIX NETWORKS

The field of anonymous communications started in the 80's when Chaum [1] introducedthe concept of anonymous emails. He suggested hiding the sender – receiver linking, taking the messages on cipher layers using a public key.A mix network aims is to hide the correspondences between the items in its input and those in its output. It collects a number of packets from distinct users called anonymity set, and then it changes the incoming packets appearance through cryptographic operations. This make impossible to link inputs and outputs taking to account timing information, see Fig. 1.

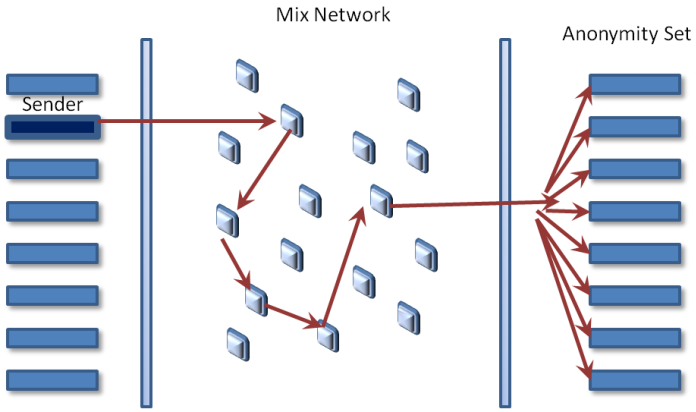


Fig.1. Basic model of a mix network

The mixing technique is called threshold mix. Anonymity properties get proportionally stronger when anonymity set increases, and these are based on uniform distribution of the actions execution of the set of subjects. A mix is a go-between relay agent that hides a message's appearance, including its bit pattern and length. For example, say Alice generates a message to Bob with a constant length. A sender protocol executes several cryptographic operations through Bob and mix public keys. Thenthemix hides the message's appearance by decoding it with itscorrespondent private key.

The initial process in order to Alice sends a message to Bob using a Mix system is to prepare the message. The first phase is to choose the message transmission path; this path must have a specific order for iteratively sending it before the message gets its final destiny. It is recommended to use more than one mix in every path for improving the system security. The next phase is to utilize the public keys of the chosen mixes for encrypting the message in the inverse order that they were chosen at first. In other words, the public key of the last mix encrypts initially the message, then the next one before the last one and finally the public key of the first mix will be used for encrypt the message.A layer is built every time the message is encrypted and the next node address is included. This way when the first mix gets a message prepared, this will be decrypted with its corresponding private key and will get the next node address.An observer, or an active attacker, should not be able to find the link between the bit pattern of the encoded messages arriving at the mix and the decoded messages departing from it. Appends a block of random bits at the end of the message have the purpose to make messages size uniform.

3 MIX SYSTEMS ATTACKS

The attacks against mix systems are intersection attacks. They take into account a message sequence through the same path in a network, it means performing traffic analysis. The set of most likely receivers is calculated for each message in the sequence and the intersection of the sets will make possible to know who the receiver of the stream is. Intersection attacks are designed based on correlating the times when senders and receivers are active. By observing the recipients that received packets during the rounds when Alice is sending, the attacker can create a set of Alice's most frequent recipients, this way diminishing her anonymity.

A. The disclosure attack

The disclosure attack was presented by Agrawal and Kesdogan in [2]. They model the attack by considering a bipartite graph $G = (A \cup B, E)$. The set of edges E represents the relationship between senders and recipients$A$ and B. Mixes assume that all networks links are observable. So, the attacker can determine anonymity sets by observing the messages to and from an anonymity network; the problem arises at asking for how long it is necessary the observation.The attack is global, in the sense that it retrieves information about the number of messages sent by Alice and received by other users; and passive, in the sense that attackercannot alter the network (sending false messages or delaying existent ones). Authors assume a particular user, Alice,sends messages to a limited $m$recipients. A disclosure attack has a learning phase and an excluding phase. The attacker should find m disjoint recipients set by observing Alice's incoming and outgoing messages. In thisattack, authors make

several strategies in order to estimate the average number of observations for achieve the disclosure attack. They assume that: *i*) Alice participates in all batches;*ii*) only one of Alice's peer partners is in the recipient sets of all batches. In conclusion, this kind of attack is very expensive because it takes an exponential time taking into account the number of messages to be analyzed trying to identify mutually disjoint set of recipients. This is the main bottleneck for the attacker, and it derives an NP-complete problem. Test and simulations showed it works well just in very small networks. A more efficient approach to get an exact solution was proposed in [3].

B. The Statistical Disclosure Attack (SDA)

This attack proposed by Danezis in [4] is based in the Disclosure Attack. It requires less computational effort by the attacker and gets the same results. The method tries to reveal the most likely set of Alice's friends using statistical operations and approximations. It means that the attacks applies statistical properties on the observations and recognize potential recipients, but it does not solve the NP-complete problem presented in previous attack. Consider as $\vec{v}$ the vector with *N* elements corresponding to each potential recipient of the messages in the system. Assume Alice has *m*recipients as the attack above, so $\frac{1}{m}$ might receive messages by her, always that $|\vec{v}| = 1$. The author also defines$\vec{u}$ as the uniform distribution over all potential recipients *N*. In each round the probability distribution is calculated, sorecipients are ordered according its probability. The information provided to the attacker is a series of vectors representing the anonymity set observed according to the t messages sent by Alice. The attacker will use this information to deduce $\vec{v}$. The highest probability elements will be the most likely recipients of Alice. Variance on the signal and the noise introduced by other senders are used in order to calculate how many observations are necessary.Alice must demonstrate consistent behaviour patterns in the long term to obtain good results, but this attack can be generalized and applied against other anonymous communication network systems. A simulation over pool mixes are in [5].Distinct to the predecessor attack, SDA just show likely recipients and does not identify Alice's recipients with certainty.

C. Extending and Resisting Statistical Disclosure

One of the main characteristics in Intersection Attacks counts on a fairly consistent sending pattern or a specific behaviour for users in an anonymity network. Mathewson and Dingledine in [6] make an extension of the original SDA. One of the more significant differences is they consider that real social networks has a scale-free network behaviour, and also consider this behaviour changes slowly over time. They do not simulate these kinds of attacks.

In order to model the sender behaviour, authors assume Alice sends *n*messages with a probability $P_m(n)$; and the probability of Alice sending to each recipient is represented in a vector $\vec{v}$. First the attacker gets a vector $\vec{u}$ whose elements are: 1/*b* the recipients that have received a message in the batch, and 0 for recipients that have not received anything. For each round*i* in which Alice sent a message, the attacker observes the number of messages $m_i$sent by Alice and calculate the arithmetic mean.

Simulations on pool mixes are presented taking into account that each mix retains the messages in its pool with the same probability every round. The results show that increase the variability in messages makes the attack slower by increasing the number of output messages. Finally they examine the degree to which a non-global adversary can execute a SDA. Assuming each sender chooses with the same probability all mixes as entry and exit points and attacker is a partial observer of the mixes. The results suggest that the attacker can succeed on a long-term intersection attack even when he observes partially the network. When most of the network is observed the attack can be done, and if more of the network is hidden then attacker will have fewer possibilities to succeed.

D. Two Sided Statistical Disclosure Attack (TS-SDA)

In [7] Danezis*et al.* provide an abstract model of an anonymity system considering that users send messages to his contacts, and some messages sent by a particular user are replies. This attack assumes a more realistic scenario regardingthe user behaviour on an email system; its aim is to estimate the distribution of contacts of Alice, and to deduce the receivers of all the messages sent by her.

The model consider *N* as the number of users in the system that send and receive messages. Each user *n* has a probability distribution $D_n$of sending a message to other users. For example the target user Alice has a distribution $D_A$of sending messages to a subset of her *k* contacts. At first the target of the attack, Alice, is the only user that will be model as replying to messages with a probability *r*. The

reply delay is the time between a messageis received and sent again. The probability of a reply *r* and the reply delay rate are assumed to be known for the attacker, just as *N* and the probability that Alice initiates messages. Based on this information the attacker estimates: i)the expected number of replies for a unit of time; ii) The expected volume of discussion initiations for each unit of time; iii) The expected volume of replies of a particular message.

Finally authors show a comparative performance of the Statistical Disclosure Attack (SDA) and the Two Sided Disclosure Attack (TS-SDA). It shows that TS-SDA obtains better results than SDA. The main advantage of the TS-SDA is its ability to uncover the recipient of replies. And SDA vaguely performs better on reveal discussion initiations. Inconvenient details for application on real data is the assumption all users have the same number of friends to which they send messages with uniform probability.

E. Perfect Matching Disclosure Attack (PMDA)

The PMDA [8] is based on graph theory, it considers all users in a round at once, instead one particular user iteratively. No assumption on the users' behaviour is required to reveal relationships between them.Comparing with previous attacks where Alice sends exactly one message per round, this model permits users to send or receive more than one message in each round. Bipartite graphs are employed to model a threshold mix, and through this show how weighted bipartite graphs can be used to disclosure users' communication. A bipartite graph $G = (S \cup R, E)$ considers nodes divided in two distinct sets *S*(senders) and *R*(receivers) such that every edge *E* links one member in *S* and one member in *R*. It is required that every node is incident to exactly one edge. In order to build a threshold mix is considered *t* messages sent during one round of the mix form the set *S,* and each node $s \in S$ is labelled with the sender's identity *sen(s)*. Equally, the *t* messages received during one round form the set *R* where each node *r* is labelled with the receiver's identity *rec(r)*. A perfect matching *M* on *G* links all *t* sent and received messages. Additionally P' is $t \times t$matrix containing weights $w_{s,r}$, representing probabilities for all possible edges in *G*.

The procedure for one round is: i) sent messages are nodded in *S*, and marked with their senders' identities; ii) received messages are nodes in *R*, and marked with their receivers' identities; iii) derive the $t \times t$matrix: first estimating user profiles when SDA and then de-anonymize mixing round with $P'(s,r) \coloneqq \tilde{P}_{sen(s),SDA}(rec(r)), s \in S_i, r \in R_i$; iv) replace each element of the matrix$P'(s,r)$ with $\log_{10}(P'(s,r))$; v) having each edge associated with a log-probability, a maximum weighted bipartite matching on the graph $G = (S \cup R, E)$ outputs the most likely sender-receiver combination. This work shows is not enough to take the perspective of just one user of the system

Results of experimentation show that this attack does not consider the possibility that users send messages with different frequencies. An extension of the proposal considers a Normalized SDA. Another related work concerning perfect matchings is perfect matching preclusion [9, 10] where Hamiltonian cycles on the hypercube are used.

F. Vida: How to Use Bayesian Inference to De-anonymizePersistent Communications

A generalisation of the disclosure attack model of an anonymity system applying Bayesian techniques is introduced by Danezis*et al*[11].Authors built a model to represent long term attacks against anonymity systems, which are represented as N_{user} users that send N_{msg}messages to each other. Assume each user has a sending profile, sampled when a message is to be sent to determine the most likely receiver. The main contributions are two models: 1) Vida Black-box modelrepresents long term attacks against any anonymity systems; 2) Vida Red-Blue allows an adversary to execute inference on a selected target through traffic analysis.

Vida Black Box model describes how messages are generated and sent in the anonymity system. In order to perform inference on the unknown entities they use Bayesian methods. The anonymity system is represented by a bipartite graph linking input messages i_x with its correspondent output messageso_i without taking into account their identities. The edges are labelled with its weight that is the probability of the input message being output. Senders are associated with multinomial profiles,which are used to choose their correspondent receivers. Through Dirichlet distribution these profiles are sampled. Applying the proposed algorithm it will throw a set of samples that will be used for attackers to estimate the marginal distributions linking senders with their respective receivers.

Vida Red-Blue model tries to answer needs of a real-world adversary, considering that he is interested in particular senders and receivers previously chosen.The adversary chooses Bob as target receiver, it will be called "Red" and all other receivers will be tagged as "Blue". The bipartite graph is divided in

two sub-graphs: one containing all edges ending on the Red target and one containing all edges ending on a Blue receiver. Techniques Bayesian are used to select the candidate sender of each Red message: the sender with the highest a-posterior probability is chosen as the best candidate.

The evaluation includes a very specific scenario where consider: i) messages sent by up to 1000 senders to up to 1000 receivers; ii) each sender is assigned 5 contacts randomly; iii) everyone sends messages with the same probability; iv) messages are anonymized using a threshold mix with a batch of 100 messages.

G. SDA with Two Heads (SDA-2H)

One of the most used strategies to attempt against SDA is sending cover traffic which consists of fake or dummy messages mixing with real ones that can hide Alice's true sending behaviour. SDA-2H [12] is an extension of SDA [3] and takes its predecessor as a baseline to improve it at consider background traffic volumes in order to estimate the amount of dummy traffic that Alice sends.Dummy traffic serve as a useful tool to increase anonymity and they are classified based on their origin: i) user cover, generated by the user Alice; ii) background cover, generated by senders other than Alice in the system; iii) receiver-bound cover, generated by the mix. This work is centred on background cover which is created when users generated false messages along with their real ones. The objective for the attacker is to estimate how much of Alice's traffic is false based on the observations between the volume of incoming and outgoing traffic. Authors make several simulations and they found that for a specific number of total recipients, the increase in the background messages makes harder for the attacker to succeed considering that total recipients and Alice's recipients are unchanged. They find also that when Alice's recipients stay and the number of total recipients increases, the attacker would need few rounds to observe for finding Alice's recipients. A comparative between SDA and SDA-2H shows that SDA-2H may not be better than SDA in all the cases, but SDA-2H take into account the effect of background cover to achieve a successful attack.

H. A Least Squares Approach to Disclosure Attack (LSDA)

Derived of an algorithm based on the Maximum Likelihood the least squares approach is proposed by Pérez-González and Troncoso [13], this attack estimates the communication partners of user in a mix network. The aim is to be able to estimate the probabilities that Alice sends a message to Bob; this will derive to a sender and receiver profiles applicable for all users. They makethe following assumptions to model the attack: the probability of sending a message from a user to a specific receiver is independent of previous messages, the behaviour of all users are independent from the others, any incoming message to the mix is considered a priori sent by any user with uniform probability, and parameters used to model statistical behaviour do not change over time. The LSDA is improved to minimize the Mean Squared Error between actual transition probabilities $p_{i,j}$ and adversary's estimated$\hat{p}_{j,i}$. In order to show the profiling accuracy of the attack they propose two metrics: i) MSE_pThe Mean Squared Error per transition probability, represents the average squared between the elements of the estimated matrix $\hat{p}$ and the elements of the matrix p (which describes the real behaviour of the users); ii)$MSE_{qi.}$ The Mean Squared Error per sender profile, which measures the average squared error between the probability of the estimated $\hat{q}_i$ and the actual $q_i$user *i*'s sender profile. The smaller the MSE, the better is the estimation. Authors claim LSDA estimates sender and receiver profiles simultaneously through executing LSDA in the reverse direction; considering the receivers and senders and so on. In their results they found out that LS coincides with SDA estimations of unknown probabilities, and concludes that LSDA is better than its predecessor's statistical attacks.

4 CONCLUSIONS

Statistical disclosure attacks are known as a powerful long-term tool against mix network whose aim is to make possible anonymous communication between senders and receives belonging to it. We have presented several attacks by adversaries on mix- based anonymity systems, their mechanisms, strengths and weakness. Each work has assumed very specific scenarios but any of them solve the problems that are presented on real-world data.In order to develop an effective attack, it must be taking into account the special properties of network human communications.

Researchers have hypothesized that some of these attacks can be extremely effective in many real-world contexts. Nevertheless it is still an open problem to approach under which circumstances and for how long of observations these attacks would be successful. More work can be done on develop

new modelling frameworks to provide solutions to all users simultaneously. Focus the simulations on real applications such as email data and social networks would be an interesting topic.

ACKNOWLEDGMENTS

This work was supported by the Agencia Española de CooperaciónInternacionalpara el Desarrollo (AECID, Spain) through AcciónIntegrada MAEC-AECID MEDITERRÁNEO A1/037528/11.

References

[1] David L. Chaum. Untraceable Electronic Mail, return addresses, and digital pseudonyms. *Communications of the ACM*, Vol. 24, No. 2, pp. 84-90, February 1981.

[2] DakshiAgrawal, DoganKesdogan. Measuring anonymity: The disclosure attack. *IEEE Security & Privacy*, Vol. 1, No. 6, pp. 27–34, November – December 2003.

[3] George Danezis. Statistical Disclosure Attacks: Traffic Confirmation in Open Environments. Security and Privacy in the Age of Uncertainty, *IFIP Advances in Information and Communication Technology* (Sabrina de Capitani di Vimercati, PierangelaSamarati, SokratisKatsikas, Eds.), pp. 421-426, April 2003.

[4] DoganKesdogan, LexiPimenidis. The hitting set attack on anonymity protocols. *InProceedings of the 6th Workshop on Information Hiding (IH)*, Toronto, Canada. Lecture Notes in Computer Science, Vol. 3200, pp. 326–339, 2004.

[5] Nick Mathewson, Roger Dingledine. Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. *InProceedings of the 4th Workshop on Privacy Enhancing Technologies (PET)*, Toronto, Canada. Lecture Notes in Computer Science, Vol. 3424, pp. 17-34, 2005.

[6] George Danezis, Andrei Serjantov. Statistical Disclosure or Intersection Attacks on Anonymity Systems. *In Proceedings of the 6th Workshop on Information Hiding (IH)*, Toronto, Canada. Lecture Notes in Computer Science, Vol. 3200, pp. 293-308, May 2004.

[7] George Danezis, Claudia Diaz, Carmela Troncoso. Two-Sided Statistical Disclosure Attack. *In Proceedings of the 7th International Workshop on Privacy Enhancing Technologies (PET)*, Ottawa, Canada. Lecture Notes in Computer Science, Vol. 4776, pp. 30-44, 2007.

[8] Carmela Troncoso, BenediktGierlichs, Bart Preneel, Ingrid Verbauwhede. Perfect Matching Disclosure Attacks. *In Proceedings of the 8th Privacy Enhancing Technologies Symposium (PETS)*, Leuven, Belgium. Lecture Notes in Computer Science, Vol. 5134, pp. 2-23, 2008.

[9] Robert Brigham, Frank Harary, Elizabeth Violin, Jay Yellen. Perfect-Matching Preclusion. *CongressusNumerantium,*UtilitasMathematica Publishing, Inc., 174, pp. 185–192, 2005.

[10] Jung-Heum Park, Sang HyukSon. Conditional Matching Preclusion for Hypercube-Like Interconnection Networks. *Theoretical Computer Science*, Vol. 410, pp. 2632-2640, June, 2009.

[11] George Danezis, Carmela Troncoso. Vida: How to use Bayesian Inference to De-Anonymize Persistent Communications. *In*Proceedings of the 9th International Symposium on Privacy Enhancing Technologies (PET), Seattle, WA, USA. *Lecture Notes in Computer Science*, Vol. 5672, pp. 56-72, 2009.

[12] Mahdi N. Al-Ameen, Charles Gatz, Matthew Wright. SDA-2H: Understanding the Value of Background Cover against Statistical Disclosure. *Journal of Networks*, Vol. 7, No. 12, pp. 1943-1951, 2012.

[13] Fernando Perez-Gonzalez, Carmela Troncoso. Understanding Statistical Disclosure: A Least Squares approach. Proceedings of the 12th Privacy Enhancing Technologies Symposium (PETS), Vigo, Spain. *Lecture Notes in ComputerScience*, Vol. 7384, pp. 38-57, 2012.

Rafael Álvarez · Joan Josep Climent · Francisco Ferrández · Francisco M. Martínez
Leandro Tortosa · José Francisco Vicent · Antonio Zamora
(editores)

Actas de la XIII Reunión Española sobre Criptología y Seguridad de la Información

RECSI XIII

Alicante, 2-5 de septiembre de 2014

PUBLICACIONES UNIVERSIDAD DE ALICANTE

Actas de la XIII Reunión Española sobre Criptología y Seguridad de la Información

RECSI XIII

Alicante, 2-5 de septiembre de 2014

Rafael Álvarez · Joan Josep Climent · Francisco Ferrández · Francisco M. Martínez
Leandro Tortosa · José Francisco Vicent · Antonio Zamora
(editores)

Publicaciones de la Universidad de Alicante

Campus de San Vicente, s/n
03690 San Vicente del Raspeig
Publicaciones@ua.es - http://publicaciones.ua.es
Teléfono: 965 903 480

ISBN: 978-84-9717-323-0

Universitat d'Alacant
Universidad de Alicante

Refinamiento Probabilístico del Ataque de Revelación de Identidades

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba
Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM)
Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España
Email: {asilva, javiergv}@fdi.ucm.es, jportela@estad.ucm.es

***Resumen*—En la actualidad muy pocas empresas reconocen que se encuentran continuamente en riesgo al estar expuestos a ataques informáticos tanto internos como externos. Más allá de simplemente instalar herramientas de protección contra hackers y células del crimen organizado tales como antivirus y firewalls, deben incluir mecanismos adecuados de seguridad en TI que brinden protección a los ataques que son cada vez más complejos. Existen diversos estudios que muestran que aún cuando se aplique el cifrado de datos en un sistema de comunicación, es posible deducir el comportamiento de los participantes a través de técnicas de análisis de tráfico. En este artículo presentamos un ataque a un sistema de comunicación anónimo basado en el ataque de revelación de identidades. El refinamiento probabilístico presenta una mejora sustancial respecto al ataque previo.**

***Palabras clave*—Análisis de tráfico, ataques estadísticos de revelación, comunicaciones anónimas, privacidad. *(Traffic analysis, statistical disclosure attacks, anonymous communications, privacy)*.**

I. Introducción

Empresas, organizaciones y sociedad generan millones de datos diariamente desde diferentes fuentes tales como: operaciones comerciales y mercantiles, redes sociales, dispositivos móviles, documentos, entre otros. La mayor parte de esta información se almacena en bases de datos altamente sensibles. Se consideran datos sensibles aquellos que puedan revelar aspectos como origen racial o étnico, estado de salud presente y futuro, información genética, creencias religiosas, filosóficas y morales, afiliación sindical, opiniones políticas, preferencia sexual y cualquier otro que pueda utilizarse para generar un daño, llámese robo de identidad, extorsión ó fraude por mencionar algunos.

La seguridad en los *data centers* se ha vuelto una de las grandes prioridades ya que tanto los ladrones de datos y células del crimen organizado buscan insistentemente infiltrarse en el perímetro de defensas a través de complejos ataques con un éxito alarmante, derivando en efectos devastadores. Hoy en día estamos inmersos en una sociedad digital donde podemos organizar un evento y enviar una invitación por Facebook; compartir fotos con amigos por medio de Instagram; escuchar música a través de Spotify; preguntar la ubicación de una calle utilizando Google Maps. La información personal es protegida por medio de la legislación y aunque no en todos los países se aplique efectivamente, en el ámbito de la sociedad digital funciona de manera diferente [1]. Toda la información disponible acerca de una persona puede ser referenciada con otra y dar lugar a prácticas de violación de la intimidad.

Cada persona tiene el derecho de controlar su información personal y proporcionarla a ciertas terceras partes. Desde la década pasada se observa una mayor preocupación por cómo se maneja la información privada de los usuarios en el ámbito gubernamental y de las empresas. Y recientemente, después de la filtración de información de un técnico estadunidense de la CIA al mundo, aumentaron las mesas de diálogo, investigaciones y fundamentalmente se creó toda una polémica en torno a la privacidad de los datos y lo expuesto que estamos a ser objetos de monitorización.

Las organizaciones privadas y públicas, así como las personas deben incluir la protección de la privacidad más allá de los típicos aspectos de integridad confidencialidad y disponibilidad de los datos. Aplicaciones utilizadas para garantizar la protección de la privacidad son por ejemplo los sistemas de resistencia a la censura, espionaje, entre otros; algunos de ellos utilzados para ofrecer seguridad a disidentes o periodistas viviendo en países con regímenes represores. Dentro de la misma rama de tecnologías, también existen mecanismos utilizados para acelerar la transición de cifrado como un servicio, que incluye cifrado basado en hardware con almacenamiento de llaves, esquemas de protección centralizada de datos para aplicaciones, bases de datos, ambientes virtuales de almacenamiento, y controles de acceso basados en roles.

Los ataques en las redes de comunicación son un serio problema en cualquier organización. Las nuevas tecnologías tienen un gran reto al buscar mejorar soluciones de seguridad para centros de datos. Se ha probado que el análisis de tráfico y la topología de una red, no proporcionan suficiente protección en la privacidad de los usuarios aún cuando se apliquen mecanismos de anonimato, ya que a través de información auxiliar, un atacante puede ser capaz de menguar sus propiedades. En el contexto de las redes de comunicación, con el análisis del tráfico se puede deducir información a partir las características observables de los datos que circulan por la red tales como: el tamaño de los paquetes, su origen y destino, tamaño, frecuencia, temporizacion, entre otros.

En este artículo nos enfocamos en mostrar cómo el análisis de tráfico de datos puede comprometer el anonimato de un sistema de comunicación anónima a través de técnicas y métodos que arrojen como resultado los patrones de comunicación de

los elementos que la componen.

La composición del presente artículo es de la siguiente manera, en primer lugar la introducción. En la sección II abordamos el estado del arte. La siguiente sección describe el algoritmo utilizado, haciendo énfasis en el refinamiento probabilístico. En la sección IV presentamos la aplicación del algoritmo. Y finalmente en la sección V mostramos las conclusiones sobre los resultados y trabajos futuros

II. Estado del arte

II-A. Privacidad

La definición de privacidad de acuerdo a [2] es el derecho de un individuo a decidir qué información acerca de él mismo puede ser comunicada a otro y bajo qué circunstancias.

Economistas, sociólogos, historiadores, abogados, ingenieros en sistemas informáticos, por mencionar algunos, han adoptado su propia definición de privacidad, tal como su valor, alcance, prioridad y curso de estudio. Detalles relacionados a los antecedentes, legislación e historia de la privacidad se muestran en [3]. De acuerdo a los expertos, privacidad e intimidad son conceptos difíciles de definir; consideramos parte de ello: las condiciones de salud, identidad, orientación sexual, comunicaciones personales, preferencias religiosas, estados financieros, además de muchas otras características. Trabajos relacionados en cómo las PETs se han aplicado desde áreas del entorno económico, social y técnico [4].

Las bases de la legislación respecto a la privacidad datan del año 1948, en la Declaración Universal de Derechos Humanos donde se estableció que ninguna persona debía ser sujeta a interferencias arbitrarias en su privacidad, familia, hogar o correspondencia, así como a su honor y reputación. Pero, a pesar de los avances políticos y legales que se han dado, no ha sido posible resolver algunos de los problemas fundamentales para evitar los abusos que se dan todos los días. La falta de claridad y precisión en los derechos a la libertad de expresión y los límites de información son un problema latente.

El desarrollo e los medios de comunicación digital, el auge de las redes sociales, la facilidad de acceso a dispositivos tecnológicos, está permeando la tranquilidad de miles de personas en su vida pública y privada. Ejemplos abundan, como el caso de una funcionaria de una localidad belga, quien fue sorprendida y videograbada mientras mantenía relaciones sexuales en las oficinas del Ayuntamiento. La grabación fue realizada y subida a Internet por un grupo de jóvenes. Otro escándalo se dio cuando el presidente del Instituto de Seguridad Social de Guatemala quién fue filmado en su oficina cuando realizaba actos poco legales. A diferencia del primer caso, en éste último sí existía un crimen que perseguir y la acción se justificaba para dar a conocer los hechos públicamente.

Como éstos, muchos más casos son parte del material disponible en internet y en los medios convencionales, como los videos que se filtraron de la Viceministra de Cultura y Juventud de Costa Rica, y del concejal del PSOE en Yébenes, España. A nadie parece importar los efectos que continúan afectando vidas, donde la indiferencia parece ser la constante. La participación de los derechos humanos nacionales e internacionales, el gobierno, los medios de comunicación así como la sociedad parecen estar lejanos de este problema. El escándalo a expensas de la intrusión y diseminación de la vida privada e íntima de las personas es inaceptable. Es un círculo vicioso que tiene su origen en la violación de un derecho, pero más cuando se lleva a las redes sociales y de ahí a la mayoría de los medios de comunicación con el pretexto de ser noticia.

II-B. Privacy Enhancing Technologies

La Comisión Europea define las Tecnologías que mejoran la privacidad [5] como “El uso de los PETs puede ayudar a diseñar sistemas de comunicación y servicios de forma que minimiza la recolección y uso de datos personales y facilita el cumplimiento con la regulación de protección de datos”. No hay una definición aceptada por completo de las PETs, así como tampoco existe una clasificación. La literatura relacionada a las categorías de los PETs de acuerdo a sus principales funciones, administración de privacidad y herramientas de protección de privacidad [6] [7] [8]. En general las PETs son observadas como tecnologías que se enfocan en:

a. Reducir el riesgo de romper principios de privacidad y cumplimiento legal.
b. Reducir al mínimo la cantidad de datos que se tienen sobre los individuos.
c. Permitir a los individuos a mantener siempre el control de su información.

Varios investigadores se han centrado en proteger la privacidad y los datos personales por medio de técnicas criptográficas. Las aplicaciones PETs tales como seguros digitales individuales o administradores virtuales de identidad se han desarrollado para plataformas confiables de cómputo. Tradicionalmente las PETs han estado limitadas para proporcionar pseudononimato [9]. En contraste a los datos totalmente anónimos, el pseudononimato permite que datos futuros o adicionales sean relacionados a datos actuales. Este tipo de herramientas son programas que permiten a individuos negar su verdadera identidad desde sistemas electrónicos que operan dicha información y sólo la revelan cuando sea absolutamente necesario. Ejemplos incluyen: navegadores web anónimos, servicios email y dinero electrónico. Para dar un mejor enfoque acerca de las PETs, consideremos la taxonomía de Solove [10] utilizada para categorizar la variedad de actividades que afectan la privacidad. Para mayor información respecto a las propiedades de privacidad en escenarios de comunicación anónimos vea [9].

- Recolección de información: Vigilancia, Interrogatorio.
- Procesamiento de la Información: Agregación, Identificación, Inseguridad, Uso secundario, Exclusión.
- Difusión de la Información: Violación de la confidencialidad, Divulgación, Exposición, Aumento de la accesibilidad, Chantaje, Apropiación, Distorsión.
- Invasión: Intrusiones, Interferencia en la toma de decisiones.

La recolección de la información puede ser una actividad dañina, aunque no toda la información es sensible, ciertos

datos definitivamente lo son. Cuando la información es manipulada, utilizada, combinada y almacenada. se etiqueta a dichas actividades como Procesamiento de la información; cuando la información es liberada, encaja en las actividades conocidas como Difusión de la información. Finalmente, el último grupo de las actividades es la Invasión que incluye violaciones directamente a individuos. Todas estas actividades son parte de las prácticas comunes de las compañías que se dedican a recolectar información, como la preferencia de compras, hábitos, nivel educativo, entre otros. Todo ello por medio de múltiples fuentes para propósitos de venta.

En otras sub-disciplinas de las ciencias computacionales, la privacidad también ha sido motivo de investigación principalmente en como las soluciones de privacidad se pueden aplicar en contextos específicos. En otras palabras, definir el proceso de cuándo y cómo deben aplicarse las soluciones de privacidad. Antes de elegir una tecnología de la protección de privacidad surgen varias preguntas que deben responderse dado que no existe la certeza de que una tecnología soluciona un problema en específico. Una de las preguntas a considerar es quién define qué es la privacidad, el diseñador de tecnologías, los lineamientos de la organización, o los usuarios [11].

II-C. Comunicaciones anónimas

Las comunicaciones anónimas tienen como objetivo ocultar las relaciones en la comunicación. Dado que el anonimato es el estado de ausencia de identidad, las comunicaciones anónimas se pueden lograr removiendo todas las características identificables de una red anónima. Consideremos a un sistema donde se concentra un conjunto de actores en una red de comunicación, tales como clientes, servidor y nodos. Estos actores intercambian mensajes por medio de canales públicos de comunicación. Pitfzmann y Hansen [9] definieron el anonimato como el estado de ser no identificable dentro de un conjunto de sujetos, conocido como el conjunto anónimo. Una de las principales características del conjunto anónimo es su variación en el tiempo. La probabilidad que un atacante puede efectivamente revelar quién es el receptor de un mensaje es exactamente de $1/n$, siendo n el número de miembros en el conjunto anónimo. La investigación en esta área se enfoca en desarrollar, analizar y llevar a cabo ataques de redes de comunicación anónimas. La infraestructura del Internet fue inicialmente planteado para ser un canal anónimo, pero ahora sabemos que cualquiera puede espiar la red. Los atacantes tienen diferentes perfiles tales como su área de acción, rango de usuarios, heterogeneidad, distribución y localización. Un atacante externo puede identificar patrones de tráfico para deducir quiénes se comunican, cuándo y con qué frecuencia.

En la literatura se ha clasificado a los sistemas de comunicación anónima en dos categorías: sistemas de alta latencia y baja latencia. Las primeras tienen como objetivo proporcionar un fuerte nivel de anonimato pero son aplicables a sistemas con actividad limitada que no demandan atención rápida tal como el correo electrónico. Por otro lado, los sistemas de baja latencia ofrecen mejor ejecución y son utilizados en sistemas de tiempo real, como por ejemplo aplicaciones web, mensajería instantánea entre otros. Ambos tipos de sistemas se basan en la propuesta de Chaum [12], quién introdujo el concepto de *mix*. El objetivo de una red de *mixes* es ocultar la correspondencia entre elementos de entrada con los de salida, es decir encubrir quien se comunica con quien. Una red de *mixes* reúne un cierto número de paquetes de usuarios diferentes llamado el conjunto anónimo, y entonces a través de operaciones criptográficas cambia la apariencia de los paquetes de entrada, por lo que resulta complicado para el atacante conocer quiénes se comunican. Los *mixes* son el bloque base para construir todos los sistemas de comunicación de alta latencia [12]. Por otro lado en los últimos años, se han desarrollado también sistemas de baja latencia, como por ejemplo: Crowds [13], Hordes [14], Babel [15], AN.ON [16], Onion routing [17], Freedom [18] and Tor [19]. Actualmente, la red de comunicación anónima más utilizado es Tor, que permite navegar de manera anónima en la web. En [20] se muestra un comparativo de la ejecución de sistemas de comunicación de alta y baja latencia.

II-D. Redes mixes

En 1981, Chaum introduce el concepto de las redes *mixes* cuyo propósito es ocultar la correspondencia entre elementos de entrada con los de salida. Una red de mixes recolecta un número de paquetes desde diferentes usuarios llamado el conjunto anónimo, y entonces cambia la apariencia de los paquetes de entrada a través de operaciones criptrográficas. Lo anterior hace imposible relacionar entradas y salidas. Las propiedades de anonimato serán más fuertes en tanto el conjunto anónimo sea mayor. Un *mix* es un agente intermediario que oculta la apariencia de un mensaje, incluyendo su longitud. Por ejemplo, supongamos que Alice genera un mensaje para Bob con una longitud constante. Un protocolo emisor ejecuta varias operaciones criptográficas a través de las llaves públicas de Bob. Después, la red *mix* oculta la apariencia del mensaje al decodificarlo con la llave privada del *mix*.

El proceso inicial para que Alice envíe un mensaje a Bob utilizando un sistema de *mixes* es preparar el mensaje. La primera fase es elegir la ruta de transmisión del mensaje; dicha ruta debe tener un orden específico para enviar iterativamente antes de que el mensaje llegue a su destino final. La siguiente fase es utilizar las llaves públicas de los *mixes* elegidos para cifrar el mensaje, en el orden inverso en que fueron elegidos. En otras palabras la llave pública del último *mix* cifra inicialmente el mensaje, después el penúltimo y finalmente la llave pública del primer *mix* es usada. Cada vez que se cifra el mensaje una capa se construye y la dirección del siguiente nodo es incluida. De esta manera cuando el primer *mix* obtiene un mensaje preparado, dicho mensaje será descifrado a través de la llave privada correspondiente y será direccionado al siguiente nodo.

Los ataques externos se ejecutan desde fuera de la red, mientras que los internos son desde nodos comprometidos los cuales son de hecho parte de la misma red. Las redes de *mixes* son una herramienta poderosa para mitigar los ataques externos al cifrar la ruta emisor- receptor. Los nodos

participantes de una red *mix* transmiten y retardan los mensajes con el fin de ocultar su ruta. Pero es posible que puedan estar comprometidos y llevar a cabo ataques internos. Este tipo de problema se trata en [13] al ocultar el emisor o receptor de los nodos de transmisión.

II-E. Análisis de tráfico

El análisis de tráfico pertenece a la familia de técnicas utilizada para deducir información de los patrones de un sistema de comunicación. Se ha demostrado que el cifrado por sí mismo no garantiza el anonimato. Aún cuando el contenido de las comunicaciones sean cifradas, la información de enrutamiento debe enviarse claramente ya que los ruteadores deben determinar el siguiente punto de la red a dónde se direccionará el paquete. En [21] se muestran algunos de las técnicas de análisis de tráfico utilizadas para revelar las identidades en una red de comunicación anónima.

II-F. Ataques estadísticos

La familia de ataques estadísticos fue iniciada por Danezis en [22] donde introdujo el ataque estadístico de revelación (*Statistical Disclosure Attack, SDA*). En dicho trabajo se nota que llevando a cabo un amplio número de observaciones por cierto período de tiempo en una red de *mixes*, se puede calcular la probabilidad de distribuciones de envío/recepción de mensajes y con ello menguar la identidad de los participantes en un sistema de comunicación anónimo. A partir de éste ataque se desarrollaron muchos más tomando como base el análisis de tráfico para deducir cierta información a partir de los patrones de comportamiento en un sistema de comunicación.

Los ataques contra redes de *mixes* son conocidos también como ataques de intersección [23]. Se toma en cuenta la secuencia de un mensaje a través de una misma ruta en la red, esto quiere decir que se analiza el tráfico. El conjunto de los receptores más probables se calcula para cada mensaje en la secuencia e intersección de los conjuntos lo que permite conocer quién es el receptor de un determinado mensaje. Los ataques de intersección se diseñan basándose en la correlación de los tiempos donde emisores y receptores se encuentran activos. Al observar los elementos que reciben paquetes durante las rondas en las que Alice está enviando un mensaje, el atacante puede crear un conjunto de receptores más frecuentes de Alice. La información proporcionada a los atacantes es una serie de vectores representando los conjuntos de anonimato observados de acuerdo a los t mensajes enviados por Alice. Dentro de la familia de ataques estadísticos, cada uno de ellos se modela con un escenario muy específico; y en algunos casos poco semejantes al comportamiento de un sistema de comunicación real. Algunos asumen que Alice tiene exactamente m receptores y que envía mensajes a cada uno de ellos con la misma probabilidad, o bien son ataques que se enfocan en un solo usuario como soluciones individuales que son interdependientes, cuando la realidad indica cuestiones diferentes.

III. ALGORITMO

El objetivo de nuestro algoritmo es extraer información relevante sobre las relaciones entre cada par de usuarios. En [24] se describe el problema, así como el marco base y supuestos. Las tablas de las rondas donde se muestran los patrones de comunicación entre usuarios se representan con valores de 1 si existe relación y 0 en caso contrario. El atacante es capaz de observar cuántos mensajes son enviados y recibidos, es decir las sumas marginales por fila y columna de cada ronda 1, ...,T donde T es el número total de rondas. En cada ronda sólo consideramos usuarios que reciben y envían mensajes. Por lo tanto, decimos que un elemento (i, j) está presente en una ronda si las marginales correspondientes son diferentes a 0.

Hemos adoptado el término "cero trivial", que son los elementos que representan pares de usuarios que nunca han coincidido en ninguna ronda, Denotando n_{ij} el contenido del elemento (i, j), n_{i+} el valor marginal de la fila i, n_{+j} el valor marginal de la columna j, n la suma de los elementos y r el número de filas.

Algoritmo 1: Descripción del algoritmo

1. Generar n_{11} de una distribución uniforme entera donde $i = 1, j = 1$;
2. Iniciar un recorrido por columnas, para cada elemento n_{k1} en esta columna hasta $k-1$, se calculan nuevas cotas para n_{k1} a partir de la siguiente ecuación:

$$\text{máx}\big((0, (n_{+1} - \sum_{i=1}^{k-1} n_{i1}) - \sum_{i=k+1}^{r} n_{i+}\big) \leq$$

$$n_{ij} \leq \text{mín}(n_{k+}, n_{+j} - \sum_{i=1}^{k-1} n_{i1})$$

   n_{k1} se genera según un entero uniforme;
3. El último elemento de la fila se rellena automáticamente al coincidir las cotas superior e inferior coinciden, haciendo $n_{(k+1)+} = 0$ por conveniencia;
4. Cuando se completa la columna ésta se elimina de la tabla y se recalculan las marginales por fila n_{i+} y el valor n;
5. La tabla tiene ahora una columna menos y se repite el proceso hasta llenar todos los elementos;

Al final lo que obtenemos son una serie de tablas factibles generadas para cada ronda. Por lo que la media de cada elemento sobre todas las tablas para todas las rondas es una estimación de su valor real. La media obtenida por elemento y ronda se agrega sobre todas las rondas la cual representa un estimado de la tabla agregada $\widehat{A}$. Para cada elemento, se estima la probabilidad de cero, calculando el porcentaje de tablas con elemento cero para cada ronda en que el elemento está presente y multiplicando las probabilidades obtenidas para todas esas rondas. En la tabla resultante los elementos se ordenan por

su probabilidad de cero a excepción de los elementos que son cero triviales. De esta manera, los elementos con menor probabilidad de ser cero son los que se consideran candidatos a tener una relación. Para llevar a cabo la clasificación seleccionamos un punto de corte p y consideramos "celdas cero" si probabilidad de cero $> p$, en tanto las "celdas positivas" son aquellas donde la probabilidad de cero$< 1 - p$. Aquellas celdas que no entran en estas dos categorías se les llama "no clasificadas".

El algoritmo utilizado en [24] presupone inicialmente equiprobabilidad de las tablas extraídas. Al desarrollarlo se obtienen, al margen de una primera clasificación de las celdas (*i*, *j*) en 1 ó 0 según exista comunicación o no entre ese par de usuarios, estimaciones para la tasa de mensajes enviados por ronda para cada celda. A partir de estas estimaciones iniciales, puede volver a desarrollarse el algoritmo en un segundo ciclo, en el cual las tablas no se generan con equiprobabilidad. En el primer ciclo del algoritmo el valor de cada celda en cada tabla-ronda era generado según una distribución uniforme manteniendo las restricciones dadas por la información marginal conocida. En este segundo ciclo existen varias posibilidades teniendo en cuenta las primeras estimaciones:

a. Generar el valor de cada celda en cada tabla-ronda según una distribución de Poisson cuyo parámetro lambda es la tasa estimada de mensajes por ronda para esa celda.
b. Generar el valor de cada celda en cada tabla-ronda según la distribución de probabilidad discreta del número de mensajes por ronda en esa celda. Esta distribución es construida a partir de los resultados del primer ciclo del algoritmo, estimando probabilidades de 0, 1, 2, ... mensajes según su porcentaje relativo de ocurrencias.

Este segundo ciclo puede volver a servir de base para ciclos sucesivos en un proceso iterativo. En los resultados siguientes se ha utilizado la opción b). Para llevar a cabo nuestro ataque, primero por cuestiones pedagógicas, simulamos los datos de un sistema de correo electrónico. Para la generación de rondas definimos el número de usuarios participantes *N*, lambda que es el promedio de mensajes enviados por ronda en la celda (*i*, *j*) y el número de rondas *NR* que se desea generar.

1. Con las rondas simuladas se ejecuta el Algoritmo 1 y se obtienen las tablas factibles de cada ronda. Posteriormente se lleva a cabo un test de clasificación binaria para los elementos calculados, donde 0 en la celda (*i*, *j*) significa que no existe relación entre el emisor *i* y el receptor *j*, en tanto 1 significa que sí hay comunicación entre ellos.
2. Generar métricas características para los tests de clasificación binaria (sensibilidad, especificidad, valor predictivo negativo, valor predictivo positivo).
3. Con la información de las tablas factibles para cada ronda se calculan las frecuencias relativas de 0, 1, 2, ... mensajes para cada celda y se obtiene una aproximación a la distribución de probabilidad del número de mensajes por ronda, a partir de la normalización de esas frecuencias relativas.
4. Se vuelve a ejecutar el algoritmo utilizando las probabilidades estimadas para cada celda, normalizadas en cada caso a sus restricciones, en lugar de la distribución uniforme.
5. Se generan métricas de clasificación binaria y se vuelven a estimar las probabilidades.
6. Se itera el proceso a partir del punto 4.

IV. Aplicación del Algoritmo

Llevamos a cabo un gran número de simulaciones luego de generar rondas. El algoritmo no proporciona soluciones uniformes, dado que algunas tablas son más probables que otras debido al orden utilizado al ir llenando filas y columnas. No nos enfocamos en encontrar soluciones para un solo usuario, por lo que: i) Reordenamos aleatoriamente filas y columnas antes de calcular tablas factibles; ii) Conservamos solo las tablas factibles diferentes.

La Tabla I presenta los resultados obtenidos aplicando los algoritmos anteriormente descritos. Los resultados de la iteración 1 corresponden a la aplicación de lo que llamamos primer ciclo [24]; a partir de la iteración 2 se ejecuta el segundo ciclo y de acuerdo a los resultados que obtuvimos pudimos observar que tres iteraciones nos proporcionaban mejores resultados en la mayoría de los casos. Se puede observar también que la complejidad de las rondas crece cuando el número de usuarios y el número de rondas es mayor.

Tabla I
Resultados de la Simulación

No. de usuarios	Iteración	Sensibilidad	Especificidad	VPP	VPN	% de clasificación
10	1	0.9876	0.5789	0.9166	0.9090	0.91
	2	0.9876	0.9473	0.9473	0.9876	0.98
	3	0.9876	0.9473	0.9473	0.9876	0.98
	4	0.9473	0.9876	0.9473	0.9876	0.98
15	1	0.3225	0.9948	0.9090	0.9018	0.90
	2	0.6774	0.9948	0.9545	0.9507	0.95
	3	0.8387	0.9948	0.9629	0.9747	0.97
	4	0.8064	0.9948	0.9615	0.9698	0.96
20	1	0.1818	0.9857	0.6666	0.8846	0.87
	2	0.7272	0.9857	0.8888	0.9583	0.95
	3	0.8181	0.9857	0.9	0.9718	0.96
	4	0.7272	0.9857	0.8888	0.9583	0.95
25	1	0.2297	0.9969	90.9444	0.8507	0.85
	2	0.4324	1	1	0.8858	0.89
	3	0.5540	1	1	0.9080	0.91
	4	0.6486	1	1	0.9261	0.93
30	1	0.1058	0.9981	0.90	0.8764	0.8768
	2	0.2235	0.9981	0.95	0.8909	0.8928
	3	0.3764	0.9981	0.96	0.9104	0.9136
	4	0.3058	0.9981	0.96	0.9013	0.904
35	1	0.0441	0.9986	0.8571	0.8544	0.85
	2	0.2205	0.9986	0.9677	0.8780	0.88
	3	0.2720	0.9986	0.9736	0.8851	0.89
	4	0.2941	0.9986	0.9756	0.8882	0.89

En la Figura 1 se modela la tasa de clasificación respecto a las veces que se ha iterado el algoritmo. Se puede observar una mejora en el porcentaje de clasificación en todos los casos, en relación a la iteración 1.

V. Conclusiones

En las redes de comunicación, los *mixes* ofrecen protección contra observadores al ocultar la apariencia de los mensajes,

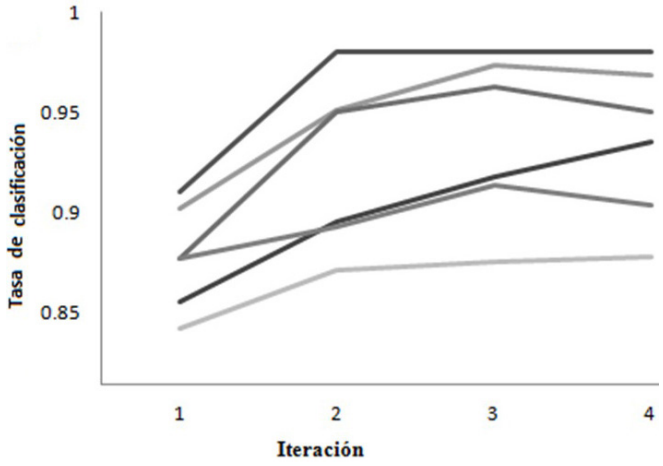


Figura 1. Tasa de clasificación vs. Número de iteración

sus patrones, longitud y enlaces entre emisores y receptores. El objetivo de este trabajo es desarrollar un ataque estadístico global para revelar la identidad de emisores y receptores en una red de comunicaciones que está protegida por técnicas estándar basadas en *mixes*. Para efecto de refinar nuestro ataque tomamos en cuenta las tablas factibles no repetidas, calculamos las frecuencias relativas para cada celda y obtuvimos una aproximación a la distribución de probabilidad del número de mensajes. El método puede ser aplicado en otro tipo de sistemas de comunicación como por ejemplo en redes sociales y protocolos punto a punto; asimismo puede ser implementado fuera del dominio de las comunicaciones como la revelación estadística de tablas públicas y la investigación forense. Nuestro método es afectado por muchos factores como el número de usuarios y el número promedio de mensajes por ronda lo que deriva a una alta complejidad de las tablas que influye de manera negativa en el ataque. El alcance en la tasa de clasificación muestra que entre mayor es el número de rondas se obtienen mejores resultados. Finalmente iteramos el algoritmo. Es necesaria mayor investigación para definir con cuántas iteraciones se pueden ver mejores resultados. De acuerdo a la literatura revisada, podemos concluir que los protocolos de anonimización propuestos hasta ahora consideran escenarios muy específicos. Los ataques estadísticos de intersección se centran en un usuario solamente, sin considerar las relaciones entre todos los usuarios.

Agradecimientos

El Grupo de Investigación GASS agradece la infraestructura proporcionada por el Campus de Excelencia Internacional (CEI) Campus Moncloa Clúster de Cambio Global y Nuevas Energías (y, más concretamente, el sistema EOLO como recurso de computación de alto rendimiento HPC - High Performance Computing), infraestructura financiada por el Ministerio de Educación, Cultura y Deporte (MECD) y por el Ministerio de Economía y Competitividad (MINECO).

Referencias

[1] B. Krishnamurthy. “Privacy and Online Social Networks: can color less green ideas sleep furiosly?” IEEE Security and Privacy, Vol. 11, No. 3, pp. 14–20, May 2013.
[2] A. Westin. “Privacy and Freedom”, Vol. 25, New York: Atheneum: Washington and Lee Law Review, 1968.
[3] R. Gellman y P. Dixon. “Online Privacy: A Reference Handbook”, Santa Barbara, CA.: ABC - CLIO, September, 2011.
[4] R. Gross and A. Acquisti. “Information revelation and privacy in online social networks”, Proceedings of the 2005 ACM workshop on Privacy in the electronic society, Alexandria, VA, USA, pp. 71–80, November 2005.
[5] European Commission. “Press release: Privacy Enhancing Technologies (PETs)”, May 2, 2007.
[6] L. Fritsch. “State or the art of privacy-enhancing technology (PET)”, Norwegian Computing Center Report, Oslo, Norway, 2007.
[7] The META Group, “State of the art of privacy-enhancing technology (PET)”, Danish Ministry of Science, Technology and Innovation, Denmark, March, 2005.
[8] C. Adams. “A Classification for Privacy Techniques”, University of Ottawa Law and Technology Journal, Vol. 3, No. 1, pp. 35–52, 2006.
[9] A. Pfitzmann y M. Hansen. “Anonymity, unlinkability, unobservability, pseudonymity, and identity management: a consolidated proposal for terminology”, TU Dresden, February 2008.
[10] D. Solove. “A Taxonomy of Privacy”, University of Pennsylvania Law Review, Vol. 154, No. 3, January, 2006.
[11] C. Diaz y S. Gurses. “Understanding the landscape of privacy technologies”, Proc. of the Information Security Summit, pp. 58–63, Prague, Czech Republic, May, 2012.
[12] D. Chaum. “Untraceable electronic mail, return addresses, and digital pseudonyms”, Communications ACM, Vol. 24, No. 2, pp. 84–90, February 1981.
[13] M. K. Reiter y A. D. Rubin. “Crowds: anonymity for Web transactions”, ACM Transactions on Information Security and System Security (TISSEC), Vol. 1, No. 1, pp. 66-92, November 1998.
[14] B. Levine y C. Shields. “Hordes: a multicast based protocol for anonymity”, Journal of Computer Security, Vol. 10, No. 3, pp. 213-240, September 2002.
[15] C. Gulcu y G. Tsudik. “Mixing Email BABEL”, in Proceedings of the 1996 Symposium on Network and Distributed System Security, pp. 2-16, San Diego, CA, USA., February 1996.
[16] O. Berthold, H. Federrath y S. Kospel. “Web MIXes: A system for anonymous and unobservable Internet access”, in Proceedings of the International workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability, pp. 115-129, Berkeley, CA, USA., July 2000.
[17] D. Goldschlag, M. Reed y P. Syverson. “Hiding Routing Infomation”, in Proceedings of the the First Workshop on Information Hiding, pp. 137-150, London, UK, 1996.
[18] A. Back, I. Goldberg y A. Shostack. “Freedom systems 2.1. security issues and analysis”, Zero Knowledge Systems, May 2001.
[19] R. Dingledine, N. Mathewson y P. Syverson. “Tor: The second-generation onion router”, in Proceedings of the the 13th USENIX Security Syposium, pp. 303-320, San Diego, CA, USA, August 2004.
[20] K. Loesing. “Privacy-enhancing Technologies for Private Services”, University of Bamberg, 2009.
[21] M. Edman y B. Yener. “On Anonymity in an Electronic Society: A Survey of Anonymous Communication Systems”, ACM Computing Surveys, Vol. 42, No. 1, pp. 1–35, December 2009.
[22] G. Danezis. “Statistical disclosure attacks: Traffic confirmation in open environments”, in Proceedings of the Security and Privacy in the Age of Uncertainty Conference, (SEC2003), Kluwer, pp. 421–426, May 2003.
[23] J. F. Raymond. “Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems”, in Proceedings of the International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability”, New York, NY, USA, 2001.
[24] J. Portela García-Miguel, D. Rupérez Cañas, A. L. Sandoval Orozco, A. G. Silva Trujillo y L. J. García Villalba. “Ataque de Revelación de Identidades en un Sistema de Correo Electrónico”, Actas de la XII Reunión Española sobre Criptología y Seguridad de la Información (RECSI 2012), Donostia-San Sebastián, España, Septiembre 2012.

Samee U. Khan · Albert Y. Zomaya
Editors

Handbook on Data Centers

Springer

Editors
Samee U. Khan
Department of Electrical
and Computer Engineering
North Dakota State University
Fargo
North Dakota
USA

Albert Y. Zomaya
School of Information Technologies
The University of Sydney
Sydney
New South Wales
Australia

ISBN 978-1-4939-2091-4
ISBN 978-1-4939-2092-1 (eBook)
DOI 10.1007/978-1-4939-2092-1

Library of Congress Control Number: 2014959415

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Privacy in Data Centers: A Survey of Attacks and Countermeasures

Luis Javier García Villalba, Alejandra Guadalupe Silva Trujillo and Javier Portela

1 Introduction

A Data Center collects, stores, and transmits huge dimensions of sensitive information of many types. Data Center security has become one of the highest network priorities as data thieves and crime cells look to infiltrate perimeter defenses through increasingly complex attack vectors with alarming success and devastating effects.

Today, organizations are placing a tremendous amount of collected data into massive repositories from various sources, such as: transactional data from enterprise applications and databases, social media data, mobile device data, documents, and machine-generated data. Much of the data contained in these data stores is of a highly sensitive nature and would trigger regulatory consequences as well as significant reputation and financial damage. This may include social security numbers, banking information, passport numbers, credit reports, health details, political opinions and anything that can be used to facilitate identity theft.

Our daily activities are developed in a digital society where the interactions between individuals and other entities are through technology. Now, we can organize an event and send the invitation using a social network like Facebook, sharing photos with friends using Instagram, listening to music through Spotify, asking for an address using Google Maps; all of these activities are just some of the ways in which many people are already working on the Internet every day. Personal information in real world is protected from strangers but it is different in the online world, where people disclose it [1]. All available information about a person gets cross-referenced,

L. J. García Villalba (✉) · A. G. Silva Trujillo · J. Portela
Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), Faculty of Information Technology and Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Madrid, Spain
Calle Profesor José García Santesmases 9, Ciudad Universitaria, 28040 Madrid, Spain
e-mail: javiergv@fdi.ucm.es

S. U. Khan, A. Y. Zomaya (eds.), *Handbook on Data Centers,*
DOI 10.1007/978-1-4939-2092-1_34

and the resulting dossier ends up being used for many purposes, lawful and otherwise. This practice has expanded over the years; the companies that compile and sell these dossiers are known as data brokers.

The communication systems behaviour has changed and it has been forced to improve its management in order to protect users privacy and satisfy the new requirements. Data centers provide a unique choice, rather than collecting data on network devices with limited capabilities for measurement, it offers measurements at the servers, even commodity versions of which have multiple cores besides other facilities. The essence of a data center is not based on concentration of data but rather the capacity to provide particular data or combinations of data upon request.

Governments and industry take advantage of sophisticated data storage tools and are using it to profile their users for financial, marketing, or just statistical purposes; organizations are able to acquire and maintain massive infrastructure at bargain prices and this derives to multiple benefits.

Individuals have the right to control their private information and only provide it to certain third parties. In the last decade users privacy concerns have grown [2–4] and since then several technologies have been developed to enhance privacy. Privacy enhancing technologies (PETs) are designed to offer mechanisms to protect personal information, and can be used with high level policy definition, human processes and training in the use of computer and communication systems [5–7]. PETs have been proposed to defend users privacy in user, network and server areas. Private and public organizations, as well as individuals should include the protection of privacy besides the typical aspects like integrity, confidentiality and availability of data. Privacy protection must avoid the disclosure of identities in a communication system. Motivations of these issues include censorship resistance, spies or law enforcement, whistleblowers, dissidents and journalists living under repressive regimes.

There are some technologies used to accelerate the transition to encryption as a service including hardware-based encryption key storage, centralized data protection schemes for applications, databases, storage and virtualized environments, as well as role-based access controls. Despite significant investment in security technology, organizations have a great hole in security effectiveness. This is due to the fact that conventional defenses rely on IP addresses and digital signatures. Signatures used in antivirus and intrusion prevention systems are effective at detecting known attacks at the time attacks are launched. They are not effective, however at detecting new attacks and are incapable of detecting hackers who are still in the reconnaissance phase, probing for weakness to attack. IP reputation databases, meanwhile, rely on the notion that attackers can be identified by their IP addresses, and so share this information across systems. Unfortunately, this is as ineffective method as it uses a postal address to identify someone. Network attacks are a serious threat to an organization. Next generation technologies are encouraged to improve the encryption solutions available at data center level. However, it has been proved that traffic and network topology analysis do not provide enough users privacy protection, even when anonymization mechanisms are applied. Using auxiliary information, adversaries can diminish anonymity properties.

In this chapter we focus on how the analysis of traffic data can compromise anonymity, showing the methods and techniques of how large amounts of traffic that has been routed through an anonymous communication system can establish communication relationships. In terms of information retrieved and considering these as leakages, designers in data centers will take them to build better capabilities to prevent attacks. Cloud computing and data centers have revolutionized the industrial world but have data protection implications which should be seriously looked into by all stakeholders to avoid putting people's privacy at risk. The solution to the previously mentioned privacy problems could be the adoption of appropriate privacy enhancing technologies.

2 Privacy

The definition of privacy according to [8] is "the right of the individual to decide what information about himself should be communicated to others and under what circumstances".

Economists, sociologists, historians, lawyers, computer scientists, and others have adopted their own privacy definitions, just as the value, scope, priority and proper course of study of privacy. Details about the background, law and history of privacy are showed in [9]. According to experts, privacy and intimacy are difficult concepts to define. However, we may consider personal health conditions, identity, sexual orientation, personal communications, financial or religious choices, along with many other characteristics. References from literature on how privacy solutions are applied from economic, social and technical areas are in [4, 10, 11].

Respect for privacy as a right includes undesirable interference, the abusive indiscretions and invasion of privacy, by any means, documents, images or recording. The legal foundations date back to 1948. In that year, the Universal Declaration of Human Rights was released, in which it was established that no person "shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation". However, despite the legal and political developments that have taken place since then, it has not been possible to solve a fundamental problem to curb abuses every day. The lack of clarity and precision in the right to freedom of expression and information limits is an open issue; cases that threaten these rights are increasing.

The development of digital media, the increasing use of social networks, the easier access to modern technological devices, is perturbing thousands of people in their public and private lives. Examples abound, the most recent was the deputy mayor of a Flemish town, who was caught and recorded on a video while having sex with a man in the Town Hall offices. The recording was made and released for an unknown group of young boys. Another scandal was the president of the Guatemalan Institute of Social Security, who was shot in his office committing "lewd acts". Unlike the previous one, in this case there was a crime and the action given was justified publicly. All of this stuff is available on the Internet and traditional media, the videos

that were leaked to the Vice Minister of Culture and Youth of Costa Rica, and the PSOE councilor in the Yébenes, Spain. Nobody seems to care about the effects which it continues to have on their lives. Indifference seems to be the constant. Participation of national and international human rights, government, media, and even the civil society organizations, seems to be far from this problem. However, the situation should be of concern. The scandal at the expense of the intrusion and dissemination of the private and intimate lives of people is unacceptable. It is a vicious circle that has its origin in the violation of a right, but when it is the social networks and hence most of the national and international media, on the pretext of being "news".

3 Privacy Enhancing Technologies

The European Commission define Privacy enhancing technologies [12] as "The use of PETS can help to design information and communication systems and services in a way that minimizes the collection and use of personal data and facilitates compliance with data protection rules. The use of PETs should result in making breaches of certain data protection rules more difficult and / or helping to detect them".

There is no widely accepted definition of the term PETs nor does there a distinguished classification exist. Literature about categorized PETs according to their main functions, privacy management and privacy protection tools [13–15].

In general PETs are observed as technologies that focus on:

- Reducing the risk of breaking privacy principles and legal compliance.
- Minimizing the amount of data held about individuals.
- Allowing individuals to maintain control of their information at all times.

Several researchers are centered on protection of privacy and personal data through sophisticated cryptology techniques. PET's applications such as individual digital safes or virtual identity managers have been proposed for trusted computing platforms.

PETs have traditionally been restricted to provide "pseudonymisation" [16]. In contrast to fully anonymized data, pseudonymisation allows future or additional data to be linked to the current data. These kind of tools are software that allow individuals to deny their true identity from those operating electronic systems or providing services through them, and only disclose it when absolutely necessary. Examples include: anonymous web browsers, email services and digital cash.

In order to give a better explanation about PETs applied in a data center, consider the Solove's Taxonomy [17] used to categorize the variety of activities to infringe privacy. We refer to [16] for further definitions of privacy properties in anonymous communication scenarios.

- *Information Collection*: Surveillance, Interrogation.
- *Information Processing*: Aggregation, Identification, Insecurity, Secondary Use, Exclusion.

- *Information Dissemination*: Breach of Confidentiality, Disclosure, Exposure, Increased Accessibility, Blackmail, Appropriation, Distortion.
- *Invasion*: Intrusion, Decisional Interference.

Collecting information can be a damaging activity, not all the information is sensitive but certain kinds definitely are. All this information is manipulated, used, combined and stored. These activities are labeled as Information Processing. When the information is released, this group of activities is called Information dissemination. Finally, the last group of activities is Invasion that includes direct violations of individuals. Data brokers are companies that collect information, including personal information about consumers, from an extensive range of sources for the purpose of reselling such information to their customers, which include private and public sector entities. Data brokers activities can fit in all of the categories above.

In other sub-disciplines of computer science, privacy has also been the focus of research, concerned mainly with how the privacy solutions are to be applied in specific contexts. In simple terms, they are concerned with defining the process of when and how to apply privacy solutions. Before choosing a technology for privacy protection, several questions have to be answered because there is no certainty that one type of technology solves one specific problem. One of the questions to consider is who defines what privacy is? (The technology designer, the organization's guidelines, or the users) [18].

4 Anonymous Communications

Anonymous communications aim to hide communications links. Since anonymity is the state of absent identity, anonymous communication can only be achieved by removing all the identifying characteristics from the anonymized network. Let's consider a system as a collection of actors, such as clients, servers, or peers, in a communication network. These actors exchange messages via public communication channels. Pitfzmann and Hansen [16] defined anonymity as "the state of being not identifiable within a set of subjects, the anonymity set".

One of the main characteristics of the anonymity set is its variation over time. The probability that an attacker can effectively disclose the message's sender is exactly $1/n$, with n as the number of members in the anonymity set. The research on this area has been focused on developing, analyzing and attacking anonymous communication networks. The Internet infrastructure was initially supposed to be an anonymous channel, but now we know that anyone can be spying in the network to reveal our data. Attackers have different profiles such as their action area, users volume capacity, heterogeneity, distribution and location. An outside attacker may identify traffic patterns to deduce who has communication with whom, when, and its frequency.

There are three different perspectives on anonymous communication: (i) Sender anonymity: Sender can contact receiver without revealing its identity; (ii) Receiver anonymity: Sender can contact receiver without knowing who the receiver is; (iii)

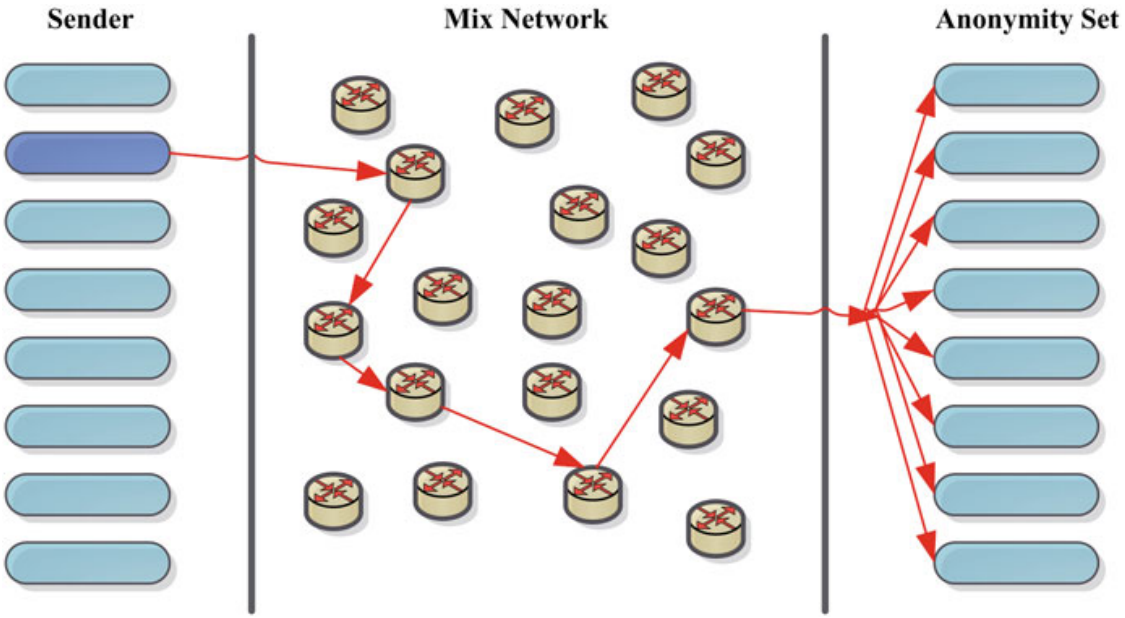


Fig. 1 Anonymous communications network

Unlinkability: Hide your relationships from third parties. According to [16] unlinkability between two items of interest occurs when an attacker of the system cannot distinguish if the two items of interest (in a system) are related or not.

Over the past years, anonymous communications has been classified by two categories: high latency systems and low latency systems. The first ones aim to provide a strong level of anonymity but are just applicable for limited activity systems that do not demand quick responses, such as email systems. On the other hand, low latency systems offer a better performance and are used in real-time systems. Examples include web applications, secure shell and instant messenger. Both systems are built on a reflection of Chaum's proposal [19]. Unlinkability is provided in a similar way in both cases using a sequence of nodes between a sender and its receiver, and using encryption to hide the message content. An intermediate node knows only its predecessor and its successor.

The mix networks systems are the basic building blocks of all modern high latency anonymous communication systems [19]; On the other hand, several designs have been developed to provide anonymity in recent years with for low latency systems, such as Crowds [20], Hordes [21], Babel [22], AN.ON [23], Onion routing [24], Freedom [25], I2P [26] and Tor [27]. Nowadays, the most widely used anonymous communication network is Tor; allowing anonymous navigation on the web. A comparison of the performance of high latency and low latency anonymous communication systems is showed in [28].

5 Mix Networks

In 1981, Chaum [19] introduced the concept of Mix networks whose purpose is to hide the correspondences between the items in its input and those in its output. A mix network collects a number of packets from distinct users called anonymity set, and then it changes the incoming packets appearance through cryptographic operations. This makes it impossible to link inputs and outputs taking into account timing information. Anonymity properties are strongest as well as the anonymity set is bigger, and these are based on uniform distribution of the actions execution of the set subjects. A mix is a go-between relay agent that hides a message's appearance, including its bit pattern and length. For example, say Alice generates a message to Bob with a constant length, a sender protocol executes several cryptographic operations through Bob and Mix public keys. After that, a mix hides the message's appearance by decoding it with the Mix private key.

The initial process for Alice to be able to send a message to Bob using a Mix system is to prepare the message. The first phase is to choose the path of the message transmission; this path must have a specific order for iteratively sending before the message gets its final destination. It is recommended to use more than one mix in every path to improve the security of the system. The next phase is to use the public keys of the chosen mixes for encrypting the message, in the inverse order that they were chosen. In other words, the public key of the last mix initially encrypts the message, then the next one before the last one and finally the public key of the first mix will be used. Every time that the message is encrypted, a layer is built and the next node address is included. This way when the first mix gets a message prepared, this will be decrypted with his correspondent private key and will get the next node address.

External attacks are executed outside the network, while internal attacks are from compromised nodes, which are actually part of the network. Mix networks are a powerful tool to mitigate outside attacks by making the sender and receiver path untraceable. The participant nodes in a mix network relay and delay messages in order to hide the route of the individual messages through the mix. However, they can be corrupted nodes that perform inside attacks. This kind of problem is addressed [20] by hiding the sender or the receiver from the relay nodes.

6 Traffic Analysis

Traffic analysis belongs to a family of techniques used to deduce information from patterns in a communication system. It has been demonstrated that encryption by itself does not provide proper anonymity; different works utilize traffic analysis techniques to uniquely identified encrypted entities. Even if communication content is encrypted, routing information has to be clearly sent because routers must determine

the next network point to which a packet should be forwarded. For example, various traffic analysis techniques have been used to disclose identities in an anonymity communication network [29].

However, there is very little information about network-level traffic characteristics of recent data centers. A data center refers to any large, dedicated cluster of computers that is owned and operated by a single organization. Data center of various sizes are being built and employed for a diverse set of purposes today. On the one hand, large universities and private enterprises are increasingly consolidating their IT services within on-site data centers containing a few hundred to a few thousand servers. Furthermore, large online service providers, such as Microsoft, Google and Amazon, are rapidly building data centers to accomplish their requirements.

Very few studies of data center traffic have been published since the challenge of instrumentation and the confidentiality of the data create significant obstacles for researchers. According to literature, there are a few that contain traffic data from corporate data centers [30]. An overview of enterprise and Internet traffic based on traces captured at Lawrence Berkeley National Laboratory appears in [31]. How using the data collected from end hosts to assess the number of unsuccessful connection attempts in an enterprise network has been applied is found in [32]. A survey showing data center components and management challenges, including: power, servers, networking and software is presented in [33]. Finally, [34] examines congestion in a data center network, but only [35] focused on the design and implementation of protocols to provide reliable communication on data centers, but recognizes that more work need to be done in order to protect privacy.

7 Mix Systems Attacks

The attacks against mix systems are intersection attacks [36]. They take into account a message sequence through the same path in a network, it means performing traffic analysis. The set of most likely receivers is calculated for each message in the sequence and the intersection of the sets will make it possible to know who the receiver of the stream is. Intersection attacks are designed based on correlating the times when senders and receivers are active. By observing the recipients that received packets during the rounds when Alice is sending, the attacker can create a set of Alice's most frequent recipients, this way diminishing her anonymity.

Next, we present the family of statistical disclosure attack, which is based in executing traffic analysis techniques.

8 The Disclosure Attack

The beginning of this family is the disclosure attack [37, 38]. The attack was modeled by considering a bipartite graph $G = (A \bigcup B, E)$. The set of edges E represents the relationship between senders and recipients A and B. Mixes assume that all networks

links are observable. So, the attacker can determine anonymity sets by observing messages to and from an anonymity network; the problem arises for how long the observation is necessary. The attack is global, in the sense that it retrieves information about the number of messages sent by Alice and received by other users, and passive, in the sense that the attacker cannot alter the network (sending false messages or delaying existent ones). Authors assume a particular user, Alice, sends messages to limited m recipients. A disclosure attack has a learning phase and an excluding phase. The attacker should find m disjoint recipients set by observing Alice's incoming and outgoing messages. In this attack, authors make several strategies in order to estimate the average number of observations for achieve the disclosure attack. They assume that: i) Alice participates in all batches; ii) only one of Alice's peer partners is in the recipient sets of all batches. In conclusion, this kind of attack is very expensive because it takes an exponential time taking into account the number of messages to be analyzed trying to identify mutually disjoint sets of recipients. This is the main bottleneck for the attacker, and it derives from an NP-complete problem. Test and simulations showed it only works well in very small networks.

9 The Statistical Disclosure Attack (SDA)

The SDA proposed by Danezis [39] is based on the previous attack. It requires less computational effort by the attacker and gets the same results. The method tries to reveal the most likely set of Alice's friends using statistical operations and approximations. It means that the attacks applies statistical properties on the observations and recognize potential recipients, but it does not solve the NP-complete problem presented in previous attack. Consider $\overrightarrow{v}$ as the vector with N elements corresponding to each potential recipient of the messages in the system. Assume Alice has m recipients as the attack above, so $\frac{1}{m}$ might receive messages by her and it's always $|\overrightarrow{v}| = 1$. The author also defines $\overrightarrow{u}$ as the uniform distribution over all potential recipients N. In each round the probability distribution is calculated, so recipients are ordered according to its probability. The information provided to the attacker is a series of vectors representing the anonymity sets The highest probability elements will be the most likely recipients of Alice. Variance on the signal and the noise introduced by other senders is used in order to calculate how many observations are necessary. Alice must demonstrate consistent behaviour patterns in the long term to obtain good results, but this attack can be generalized and applied against other anonymous communication network systems. A simulation over pool mixes are in [40]. Distinct to the predecessor attack, SDA only show likely recipients and does not identify Alice's recipients with certainty.

10 Extending and Resisting Statistical Disclosure

One of the main characteristics in Intersection Attacks relies on a fairly consistent sending pattern or a specific behaviour for users in an anonymity network. Mathewson and Dingledine in [41] make an extension of the original SDA. One of the more significant differences is that they consider that a real social network has a scale-free network behaviour, and also such behaviour changes slowly over time. They do not simulate these kinds of attacks.

In order to model the sender behaviour, authors assume Alice sends n messages with a a probability $Pm(n)$; and the probability of Alice sending to each recipient is represented in a vector $\overrightarrow{v}$. First the attacker gets a vector $\overrightarrow{u}$ whose elements are: $\frac{1}{b}$ the the recipients that have received a message in the batch, and 0 for recipients that have not. For each round i in which Alice sent a message, the attacker observes the number of messages m_i sent by Alice and calculates the arithmetic mean.

Simulations on pool mixes are presented, taking into account that each mix retains the messages in its pool with the same probability every round. The results show that increasing variability in the message makes the attack slower by increasing the number of output messages. Finally they examine the degree to which a non-global adversary can execute a SDA. Assuming all senders choose with the same probability all mixes as entry and exit points and attacker is a partial observer of the mixes. The results suggest that the attacker can succeed on a long-term intersection attack even when it partially observes the network. When most of the network is observed the attack can be made, and if more of the network is hidden then the attacker will have fewer possibilities to succeed.

11 Two Sided Statistical Disclosure Attack (TS-SDA)

[42] Danezis et al. provide an abstract model of an anonymity system considering that users send messages to his contacts, and takes into account some messages sent by a particular user are replies. This attack assumes a more realistic scenario regarding the user behaviour on an email system; its aim is to estimate the distribution of contacts of Alice, and to deduce the receivers of all the messages sent by her.

The model considers N as the number of users in the system that send and receive messages. Each user n has a probability distribution D_n of sending a message to other users. For example, the target user Alice has a distribution D_A of sending messages to a subset of her k contacts. At first the target of the attack, Alice, is the only user that will be model as replying to messages with a probability r. The reply delay is the time between a message being received and sent again. The probability of a reply r and the reply delay rate are assumed to be known for the attacker, just as N and the probability that Alice initiates messages. Based on this information the attacker estimates: (i) the expected number of replies for a unit of time; (ii) The expected volume of discussion initiations for each unit of time; (iii) The expected volume of replies of a particular message.

Finally authors show a comparative performance of the Statistical Disclosure Attack (SDA) and the Two Sided Disclosure Attack (TS-SDA). It shows that TS-SDA obtains better results than SDA. The main advantage of the TS-SDA is its ability to uncover the recipient of replies. on reveal discussion initiations. Inconvenient details for application on real data is the assumption that all users have the same number of friends to which they send messages with uniform probability.

12 Perfect Matching Disclosure Attack (PMDA)

The PMDA [8] is based on graph theory, it considers all users in a round at once, instead of one particular user iteratively. No assumption on the users behaviour is required to reveal relationships between them. Comparing with previous attacks where Alice sends exactly one message per round, this model permits users to send or receive more than one message in each round. Bipartite graphs are employed to model a threshold mix, and through this, they show how weighted bipartite graphs can be used to disclosure users communication. A bipartite graph $G = (S \bigcup R, E)$ considers nodes divided in two distinct sets S (senders) and R (receivers) so that every edge E links one member in S and one member in R. It is required that every node is incident to exactly one edge. In order to build a threshold mix, it is thought that t messages sent during one round of the mix form the set S, and each node $s \in S$ is labeled with the sender's identity $\sin(s)$. Equally, the t messages received during one round form the set R where each node r is labeled with the receiver's identity $rec(r)$. A perfect matching M on G links all t sent and received messages. Additionally P' is $t \times t$ matrix containing weights w_s, r, representing probabilities for all possible edges in G.

The procedure for one round is: (i) sent messages are nodded in S, and marked with their senders identities; (ii) received messages are nodes in R, and marked with their receivers identities; (iii) derive the $t \times t$ matrix: first estimating user profiles when SDA and then de-anonymize mixing round with $P'(s,r) = \widetilde{P}_{\sin(S),SDA}(rec(r)), s \in S_i, r$; iv) replace each element of the matrix $P'(s,r)$ with $log_{10}(P'(s,r))$; v) having each edge associated with a log-probability, a maximum weighted bipartite matching on the graph $G = (S \bigcup R, E)$ outputs the most likely sender-receiver combination. This work shows that it is not enough to take the perspective of just one user of the system.

Results of experimentation show that this attack does not consider the possibility that users send messages with different frequencies. An extension proposal considers a Normalized SDA. Another related work concerning perfect matchings is perfect matching preclusion [43, 44] where Hamiltonian cycles on the hypercube are used.

13 Vida: How to Use Bayesian Inference to De-anonymize Persistent Communications

A generalization of the disclosure attack model of an anonymity system applying Bayesian techniques is introduced by Danezis et al. [45]. Authors build a model to represent long term attacks against anonymity systems, which are represented as N_{user} users that send N_{msg} messages to each other. Assume each user has a sending profile, sampled when a message is to be sent to determine the most likely receiver. The main contributions are two models: (1) Vida Black-box model represents long term attacks against any anonymity systems; (2) Vida Red-Blue allows an adversary to performance inference on selected target through traffic analysis.

Vida Black Box model describes how messages are generated and sent in the anonymity system. In order to perform inference on the unknown entities they use Bayesian methods. The anonymity system is represented by a bipartite graph linking input messages i_x with its correspondent output messages o_y without taking into account their identities. The edges are labelled with its weight that is the probability of the input message being sent out. Senders are associated with multinomial profiles, which are used to choose their correspondent receivers. Through Dirichlet distribution these profiles are sampled. Applying the proposed algorithm will derive a set of samples that will be used for attackers to estimate the marginal distributions linking senders with their respective receivers.

Vida Red-Blue model tries to respond to the needs of a real-world adversary, considering that he is interested in particular target senders and receivers. The adversary chooses Bob as a target receiver, it will be called "Red" and all other receivers will be tagged as "Blue". The bipartite graph is divided into two sub-graphs: one containing all edges ending on the Red target and one containing all edges ending on a Blue receiver. Techniques Bayesian are used to select the candidate sender of each Red message: the sender with the highest a-posterior probability is chosen as the best candidate.

The evaluation includes a very specific scenario which considers: (i) messages sent by up to 1000 senders to up to 1000 receivers; (ii) each sender is assigned 5 contacts randomly; (iii) everyone sends messages with the same probability; (iv) messages are anonymized using a threshold mix with a batch of 100 messages.

14 SDA with Two Heads (SDA-2H)

One of the most used strategies to attempt against SDA is sending cover traffic which consists of fake or dummy messages mixed with real ones that can hide Alice's true sending behaviour. SDA-2H [46] is an extension of SDA [39] and takes its predecessor as a baseline to improve it as it considers background traffic volumes in order to estimate the amount of dummy traffic that Alice sends. Dummy traffic serves as a useful tool to increase anonymity and they are classified based on their origin: (i) user cover, generated by the user Alice; (ii) background cover, generated by senders

other than Alice in the system; (iii) receiver-bound cover, generated by the mix. This work is centered on background cover which is created when users generated false messages along with their real ones. The objective for the attacker is to estimate how much of Alice's traffic is false based on the observations between the volume of incoming and outgoing traffic. Authors make several simulations and find that for a specific number of total recipients, the increase in the background messages makes it harder for the attacker to succeed having total recipients and Alice's recipients unchanged. They also find that when Alice's recipients stay and the number of total recipients increases, the attacker would need few rounds of observations to find Alice's recipients. A comparative between SDA and SDA-2H shows that SDA-2H may not be better than SDA in all cases, but SDA-2H takes into account the effect of background cover to achieve a successful attack.

15 Conclusions

In spite of widespread interest in datacenter networks, little has been published that reveals the nature of their traffic, or the problems that arise in practice. This chapter first shows how traffic analysis can be used to disclosure information, even considering patterns such as which servers talk to each other, when and for what purpose; or characteristics as duration streams or statistics. Although modern technologies have enhanced the way we conduct everyday business—these same technologies create new risks as they are deployed into the modern IT environment. The digital environment is changing and the focus must be on attackers, more work should be done to provide a useful guide for datacenter network designers. The real problem: Not only have attacks against the entire data center infrastructure increased, they've also become much more sophisticated. The influx of advanced attacks has become a serious issue for any data center provider looking to host modern technologies. As privacy research advances, we observe that some of our assumptions about the capabilities of privacy solutions also change. Risk reduction to acceptable levels should be taken into account to develop measures against internal and external threats.

Acknowledgment Part of the computations of this work were performed in EOLO, the HPC of Climate Change of the International Campus of Excellence of Moncloa, funded by MECD and MICINN.

References

1. Krishnamurthy, B.: Privacy and Online Social Networks: Can Colorless Green Ideas Sleep Furiously? IEEE Security Privacy **11**(3) (May 2013) 14–20
2. Dey, R., Jelveh, Z., Ross, K.: Facebook Users Have Become Much More Private: A Large-Scale Study. In: IEEE International Conference on Pervasive Computing and Communications Workshops. (19–23 March 2012) 346–352

3. Christofides, E., Desmarais, A.M.S.: Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes? CyberPsychology & Behavior **12**(3) (June 2013) 341–345
4. Gross, R., Acquisti, A.: Information Revelation and Privacy in Online Social Networks. In: 2005 ACM Workshop on Privacy in the Electronic Society, ACM (2005) 71–80
5. Goldberg, I., Wagner, D., Brewer, E.: Privacy-enhancing technologies for the Internet. In: IEEE Compcon'97. (February 23–26 1997) 103–109
6. Goldberg, I.: Privacy-Enhancing Technologies for the Internet, II: Five Years Later. In: Second International Workshop on Privacy Enhancing Technologies. (April 14–15 2003) 1–12
7. Goldberg, I.: Privacy Enhancing Technologies for the Internet III: Ten Years Later. In: Digital Privacy: Theory, Technologies and Practices, Auerbach Publications (December 2007) 3–18
8. Westin, A.F.: Privacy and Freedom. The Bodley Head Ltd (1997)
9. R. Gellman, P.D.: Online Privacy: A Reference Handbook. ABC-CLIO (2011)
10. Berendt, B., Günther, O., Spiekermann, S.: Privacy in e-Commerce: Stated Preferences vs. Actual Behavior. Communications of the ACM **48**(4) (April 2005) 101–106
11. Narayanan, A., Shmatikov, V.: De-Anonymizing Social Networks. In: IEEE Symposium on Security and Privacy, Washington, DC, USA, IEEE Computer Society (2009) 173–187
12. Commission, E.: Privacy Enhancing Technologies (PETs): The Existing Legal Framework (May 2007)
13. Fritsch, L.: State of the Art of Privacy-Enhancing Technology (PET). Technical report, Norsk Regnesentral, Norwegian Computing Center (2007)
14. Group, M.: Privacy Enhancing Technologies". Technical report, Ministry of Science, Technology and Innovation (March 2005)
15. Adams, C.: A Classification for Privacy Techniques. University of Ottawa Law & Technology Journal **3**(1) (July 2006) 35–52
16. Pfitzmann, A., Hansen, M.: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management: A Consolidated Proposal for Terminology. http://dud.inf.tu-dresden.de/Anon_Terminology.shtml (February 2008) v0.31.
17. Solove, D.J.: A Classification for Privacy Techniques. University of Pennsylvania Law Review **154**(3) (January 2006) 477–560
18. Diaz, C., Gürses, S.: Understanding the Landscape of Privacy Technologies. In: The 13th International Conference on Information Security (Information Security Summit). (2012) 1–6
19. Chaum, D.L.: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of ACM **24**(2) (February 1981) 84–90
20. Reiter, M.K., Rubin, A.D.: Crowds: Anonymity for Web Transactions. ACM Transactions on Information and System Security **1**(1) (November 1998) 66–92
21. Levine, B.N., Shields, C.: Hordes: A Multicast Based Protocol for Anonymity. Journal of Computer Security **10**(3) (September 2002) 213–240
22. Gulcu, C., Tsudik, G.: Mixing Email with Babel. In: Symposium on Network and Distributed System Security, Washington, DC, USA, IEEE Computer Society (1996) 1–15
23. Berthold, O., Federrath, H., Kopsell, S.: Web MIXes: A System for Anonymous and Unobservable Internet Access. In: International Workshop On Designing Privacy Enhancing Technologies: Design Issues In Anonymity And Unobservability, Springer-Verlag New York, Inc. (2001) 115–129
24. Goldschlag, D.M., Reed, M.G., Syverson, P.F.: Hiding Routing Information. In: First International Workshop on Information Hiding, London, UK, UK, Springer-Verlag (May 30 - June 1 1996) 137–150
25. Back, A., Goldberg, I., Shostack, A.: Freedom Systems 2.1 Security Issues and Analysis (May 2001)
26. Back, A., Goldberg, I., Shostack, A.: I2P (2003)
27. Dingledine, R., Mathewson, N., Syverson, P.: Tor: The Second-generation Onion Router. In: 13th Conference on USENIX Security Symposium - Volume 13, Berkeley, CA, USA, USENIX Association (2004) 21–21

28. Loesing, K.: Privacy-Enhancing Technologies for Private Services. PhD thesis, University of Bamberg (2009)
29. Edman, M., Yener, B.: On Anonymity in an Electronic Society: A Survey of Anonymous Communication Systems. ACM Computing Surveys **42**(1) (December 2009) 1–35
30. Benson, T., Anand, A., Akella, A., Zhang, M.: Understanding Data Center Traffic Characteristics. ACM SIGCOMM Computer Communication Review **40**(1) (January 2010) 92–99
31. Pang, R., Allman, M., Bennett, M., Lee, J., Paxson, V., Tierney, B.: A First Look at Modern Enterprise Traffic. In: 5th ACM SIGCOMM Conference on Internet Measurement, Berkeley, CA, USA, USENIX Association (October 19-21 2005) 2–2
32. Guha, S., Chandrashekar, J., Taft, N., Papagiannaki, K.: How Healthy Are Today's Enterprise Networks? In: 8th ACM SIGCOMM Conference on Internet Measurement, New York, NY, USA, ACM (October 20-22 2008) 145–150
33. Kandula, S., Sengupta, S., Greenberg, A., Patel, P., Chaiken, R.: The Nature of Data Center Traffic: Measurements & Analysis. In: 9th ACM SIGCOMM Conference on Internet Measurement Conference, New York, NY, USA, ACM (November 4-6 2009) 202–208
34. Greenberg, A., Maltz, D.A.: What Goes Into a Data Center? (2009)
35. Balakrishnan, M.: Reliable Communication for Datacenters. PhD thesis, Cornell University (September 2008)
36. Raymond, J.F.: Traffic Analysis: Protocols, Attacks, Design Issues and Open Problems. In: International Workshop On Design Issues In Anonymity And Unobservability, Springer-Verlag New York, Inc. (July 25-26 2000) 10–29
37. Kedogan, D., Agrawal, D., Penz, S.: Limits of Anonymity in Open Environments. In: 5th International Workshop on Information Hiding, London, UK, UK, Springer-Verlag (October 7–9 2002) 53–69
38. Agrawal, D., Kesdogan, D.: Measuring Anonymity: The Disclosure Attack. IEEE Security Privacy **1**(6) (2003) 27–34
39. Danezis, G.: Statistical Disclosure Attacks: Traffic Confirmation in Open Environments. In: IFIP Advances in Information and Communication Technology, Kluwer (2003) 421–426
40. Danezis, G., Serjantov, A.: Statistical Disclosure or Intersection Attacks on Anonymity Systems. In: 6th Information Hiding Workshop. (May 23–25 2004) 293–308
41. Mathewson, N., Dingledine, R.: Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. In: 4th International Conference on Privacy Enhancing Technologies. (May 23-25 2004) 17–34
42. Danezis, G., Diaz, C., Troncoso, C.: Two-sided Statistical Disclosure Attack. In: 7th International Conference on Privacy Enhancing Technologies, Berlin, Heidelberg, Springer-Verlag (June 20–22 2007) 30–44
43. Brigham, R., Harary, F., Violin, E., Yellen, J.: Perfect-Matching Preclusion. Congressus Numerantium **174** (2005) 185–192
44. Park, J.H., Son, S.H.: Conditional Matching Preclusion for Hypercube-like Interconnection Networks. Theoretical Computer Science **410**(27–29) (June 2009) 2632–2640
45. Danezis, G., Troncoso, C.: Vida: How to use Bayesian Inference to De-anonymize Persistent Communications. In: 9th International Symposium of Privacy Enhancing Technologies, Springer Berlin Heidelberg (August 5-7 2009) 56–72
46. Al-Ameen, M., Gatz, C., Wright, M.: SDA-2H: Understanding the Value of Background Cover Against Statistical Disclosure. In: 14th International Conference on Computer and Information Technology. (December 22-24 2011) 196–201

Index

S. U. Khan, A. Y. Zomaya (eds.), Handbook on Data Centers,
DOI 10.1007/978-1-4939-2092-1

Sensors **2015**, *15*, 4052-4071; doi:10.3390/s150204052

OPEN ACCESS
sensors
ISSN 1424-8220
www.mdpi.com/journal/sensors

Article

Extracting Association Patterns in Network Communications

Javier Portela [1], Luis Javier García Villalba [1], Alejandra Guadalupe Silva Trujillo [1,2], Ana Lucila Sandoval Orozco [1] and Tai-hoon Kim [3,*]

[1] Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), Faculty of Information Technology and Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, Madrid 28040, Spain; E-Mails: jportela@estad.ucm.es (J.P.); javiergv@fdi.ucm.es (L.J.G.V.); asilva@fdi.ucm.es (A.G.S.T); asandoval@fdi.ucm.es (A.L.S.O.)

[2] Facultad de Ingeniería, Universidad Autónoma de San Luis Potosí (UASLP), Zona Universitaria Poniente, San Luis Potosí 78290, Mexico

[3] Department of Convergence Security, Sungshin Women's University, 249-1 Dongseon-dong 3-ga, Seoul 136-742, Korea

* Author to whom correspondence should be addressed; E-Mail: taihoonn@daum.net; Tel.: +82-10-8592-4900.

Academic Editor: Neal N. Xiong

Received: 12 November 2014 / Accepted: 29 January 2015 / Published: 11 February 2015

Abstract: In network communications, mixes provide protection against observers hiding the appearance of messages, patterns, length and links between senders and receivers. Statistical disclosure attacks aim to reveal the identity of senders and receivers in a communication network setting when it is protected by standard techniques based on mixes. This work aims to develop a global statistical disclosure attack to detect relationships between users. The only information used by the attacker is the number of messages sent and received by each user for each round, the batch of messages grouped by the anonymity system. A new modeling framework based on contingency tables is used. The assumptions are more flexible than those used in the literature, allowing to apply the method to multiple situations automatically, such as email data or social networks data. A classification scheme based on combinatoric solutions of the space of rounds retrieved is developed. Solutions about relationships between users are provided for all pairs of users simultaneously, since the dependence of the data retrieved needs to be addressed in a global sense.

Keywords: anonymity; mixes; network communications; statistical disclosure attack

1. Introduction

When information is transmitted through the Internet, it is typically encrypted in order to prevent others from being able to view it. The encryption can be successful, meaning that the keys cannot be easily guessed within a very long period of time. Even if the data themselves are hidden, other types of information may be vulnerable. In the e-mail framework, anonymity concerns the senders "identity, receivers" identity, the links between senders and receivers, the protocols used, the size of data sent, timings, *etc*. Since [1] presented the basic ideas of the anonymous communications systems, researchers have developed many mix-based and other anonymity systems for different applications, and attacks on these systems have also been developed. Our work aims to develop a global statistical attack to disclose relationships between users in a network based on a single mix anonymity system.

2. Introducing Anonymous Communications

The infrastructure of the Internet was initially planned and developed to be an anonymous channel, but nowadays, it is well known that anybody can spy on it with different non-robust tools, like, for example, using sniffers and spoofing techniques. Since the Internet's proliferation and the use of some services associated with it, such as web searchers, social networks, webmail and others, privacy has become a very important research area, not just for security IT experts or enterprises. Connectivity and the enormous flow of information available on the Internet are a very powerful tool to provide knowledge and to implement security measures to protect systems.

Anonymity is a legitimate means in many applications, such as web browsing, e-vote, e-bank, e-commerce and others. Popular anonymity systems are used by hundreds of thousands people, such as journalists, whistle blowers, dissidents and others. It is well known that encryption does not guarantee the anonymity required for all participants. Attackers can identify traffic patterns to deduce who, when and how often users are in communication. The communication layer is exposed to traffic analysis, so it is necessary to anonymize it, as well as the application layer that supports anonymous cash, anonymous credentials and elections.

Anonymity systems provide mechanisms to enhance user privacy and to protect computer systems. Research in this area focuses on developing, analyzing and executing anonymous communication networks attacks.

Two categories for anonymous communication systems are commented on below: high latency systems and low latency systems. Both systems are based on Chaum's proposal [1] that introduced the concept of mixing.

- High latency anonymity systems aim to provide a strong level of anonymity and are oriented to limited activity systems that do not demand quick responses, such as email systems. These systems are message-oriented systems.

- Low latency anonymity systems can be used for interactive traffic, for example web applications, instant messaging and others. These systems are connection-based systems and are used to defend from a partial attacker who can compromise or observe just a part of the system. According to its nature, these systems are more susceptible to timing attacks and traffic analysis attacks. The majority of these systems depend on onion routing [2] for anonymous communication.

In low latency communication systems, an attacker only needs to observe the flow of the data stream to link sender and receptor users. Traditionally, in order to prevent this attack, dummy packets are added and delays incorporated into stream data to make the traffic between users uniform. The previously mentioned scenario can be useful for passive attackers that do not insert timing partners into the traffic to compromise anonymity. An active attacker can control routers of the network. Timing attacks are one of the main challenges in low latency anonymous communication systems. These attacks are closely related to traffic analysis in mix networks.

Traffic analysis techniques belong to the family of methods to infer information from the patterns in a communication system. Even when communication content has been ciphered, information routing needs to be sent clearly for routers to know the next package's destination in the network. Every data packet traveling on the Internet contains the node addresses of sending and recipient nodes. Therefore, it is well understood that, actually, no packet can be anonymous at this level.

2.1. Mixes and the Mix Network Model

Mixes are considered the base for building high latency anonymous communication systems. In network communications, mixes provide protection against observers hiding the appearance of messages, patterns, length and links between senders and receivers. Chaum [1] introduced also the concept of anonymous email. Their model suggested hiding the correspondence between senders and receivers encrypting messages and reordering them through a path of mixes before relaying them to their destinations. The set of the most likely receivers is calculated for each message in the sequence, and intersection of sets will make it possible to know who the receiver of the stream is.

A mix networks aims to hide the correspondences between the items in its input and those in its output, changing the incoming packets appearance through cryptographic operations (see Figure 1). The anonymity set is the set of all possible entities who might execute an action. The initial process in order for Alice send a message to Bob using a mix system is to prepare the message. The first phase is to choose the message transmission path; it has a specific order for iteratively sending messages before arriving at its final destination. It is recommended to use more than one mix in every path for improving system security. The next phase is to utilize the public keys of the chosen mixes for encrypting the message in the inverse order that they were chosen. Therefore, the public key of the last mix initially encrypts the message, then the next one before the last one, and finally, the public key of the first mix will be used. Every time a message is encrypted, a layer is built, and the next node address is included. This way, when the first mix gets a message prepared, this will be decrypted with its correspondent private key and will get the next node address.

An observer or an active attacker should not be able to find the link between the bit pattern of encoded messages arriving at the mix and decoded messages departing from it. Appending a block of random bits at the end of the message has the purpose of making messages uniform in size.

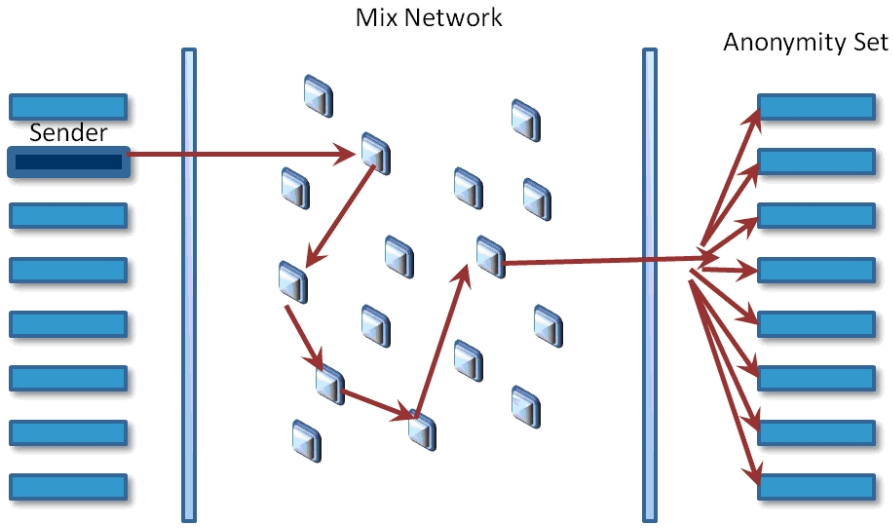


Figure 1. Mix network model.

2.2. ISDN-Mixes

The first proposal for the practical application of mixes [3] showed the way a mix-net could be used with ISDN lines to anonymize a telephone user's real location. The origin of this method took into account the fact that mixes in their original form imply a significant data expansion and significant delays, and therefore, it was often considered infeasible to apply them to services with higher bandwidth and real-time requirements. The protocol tries to defeat these problems.

2.3. Remailers

The first Internet anonymous remailer was developed in Finland and was very simple to use. A user added an extra header to the e-mail pointing out its final destination: an email address or a Usenet newsgroup. A server receives messages with embedded instructions about where to send them next without revealing their origin. All standard-based email messages include the source and transmitting entities at the headers. The full headers are usually eliminated. The application replaces the original email's source address with the remailer's address.

Babel [4], Mixmaster [5] and Mixminion [6] are some others anonymous communication designs. The differences between systems will not be addressed in our work. We centered only on senders and receivers active in a period of time, and we do not take into account message reordering, because this does not affect our attack. Onion routing [2] is another design used to provide low latency connection for web browsing and other interactive services. It is important to specify that our method does not address this kind of design; they can be treated by short-term timing or packet counting attacks [7].

3. The Family of Mix Systems Attacks

The attacks against mix systems are intersection attacks and aim to reduce the anonymity by linking senders with the messages that they send, receivers with the messages that they receive or linking senders with receivers. Attackers can derive relations of frequency through observation of the network, compromising mixes or keys, delaying or altering messages. They can deduce the messages' most probable destinations through the use of false messages sent to the network and using this technique to isolate target messages and to derive their properties. Traffic analysis belongs to a family of techniques used to deduce pattern information in a communication system. It has been proven that cipher by itself does not guarantee anonymity. See [8] for a review of traffic analysis attacks.

3.1. The Disclosure Attack

In [9], Agrawal and Kesdogan presented the disclosure attack, an attack centered on a single batch mix, aiming to retrieve information from a particular sender, called Alice. The attack is global, in the sense that it retrieves information about the number of messages sent by Alice and received by other users, and passive, in the sense that attackers cannot alter the network, for example, by sending false messages or delaying existent messages.

It is assumed that Alice has exactly m recipients and that Alice sends messages with some probability distribution to each of her recipients; also that she sends exactly one message in each batch of b messages. The attack is modeled considering a bipartite graph G. Through numerical algorithms, disjoint sets of recipients would be identified to reach, through intersection, the identification of Alice recipients. The authors use several strategies in order to estimate the average number of observations for achieving the disclosure attack. The assumptions are: (i) Alice participates in all batches; and (ii) only one of Alice's peer partners is in the recipient set of all batches. This attack is computationally expensive, because it takes an exponential time analyzing the number of messages to identify a mutually disjoint set of recipients. The main bottleneck for the attacker derives from an NP-complete problem when it is applied to big networks. The authors claim the method performs well on very small networks.

3.2. Statistical Disclosure Attacks

In [10], Danezis presents the statistical disclosure attack, maintaining some of the assumptions made in [9]. In the statistical disclosure attack, recipients are ordered in terms of probability. Alice must demonstrate consistent behavior patterns in the long term to obtain good results. The Statistical Disclosure Attack (SDA) requires less computational effort by the attacker and gets the same results. The method tries to reveal the most likely set of Alice's friends using statistical operations and approximations.

Statistical disclosure attacks when threshold mixing or pool mixing are used are treated also in [11], maintaining the assumptions of precedent articles, that is, focusing on one user, Alice, and supposing that the number of recipients of Alice is known. Besides, the threshold parameter B is also supposed to be known. One of the main characteristics of intersection attacks counts on a fairly consistent sending pattern or a specific behavior of anonymous network users.

Mathewson and Dingledine in [12] make an extension of the original SDA. One of the more significant differences is that they regard real social networks to have scale-free network behavior and also consider that such behavior changes slowly over time. The results show that increasing message variability makes the attack slow by increasing the number of output messages; assuming all senders choose with the same probability all mixes as entry and exit points and the attacker is a partial observer of the mixes.

Two-sided statistical disclosure attacks [13] use the possibilities of replies between users to make the attack stronger. This attack assumes a more realistic scenario, taking into account the user behavior on an email system. Its aim is to estimate the distribution of contacts of Alice and to deduce the receivers of all of the messages sent by her. The model considers N as the number of users in the system that send and receive messages. Each user n has a probability distribution Dn of sending a message to other users. At first, the target, Alice, is the only user that will be modeled as replying to messages with a probability r. An inconvenient detail for applications on real data is the assumption that all users have the same number of friends and send messages with uniform probability.

Perfect matching disclosure attacks [14] try to use simultaneous information about all users to obtain better results related to the disclosing of the Alice set of recipients. This attack is based on graph theory, and it does not consider the possibility that users send messages with different frequencies. An extension proposal considers a normalized SDA.

Danezis and Troncoso [15] present a new modeling approach, called Vida, for anonymous communication systems. These are modeled probabilistically, and Bayesian inference is applied to extract patterns of communications and user profiles. The authors developed a model to represent long-term attacks against anonymity systems. Assume each user has a sending profile, sampled when a message is to be sent to determine the most likely receiver. Their proposal includes: (1) the Vida black-box model representing long-term attacks against any anonymity systems. Bayesian techniques are used to select the candidate sender of each message: the sender with the highest *a posteriori* probability is chosen as the best candidate. The evaluation includes a very specific scenario considering the same number of senders and receivers. Each sender is assigned to five contacts randomly, and everyone sends messages with the same probability.

In [16], a new method to improve the statistical disclosure attack, called the hitting set attack, is introduced. Frequency analysis is used to enhance the applicability of the attack, and duality checking algorithms are also used to resolve the problem of improving the space of solutions. Mallesh and Wright [17] introduces the reverse statistical disclosure attack. This attack uses observations of all users sending patterns to estimate both the targeted user's sending pattern and her receiving pattern. The estimated patterns are combined to find a set of the targeted user's most likely contacts.

In [18], an extension to the statistical disclosure attack, called SDA-2H, is presented, considering the situation where cover traffic, in the form of fake or dummy messages, is employed as a defense.

Perez-Gonzalez *et al.* [19] presents a least squares approximation to the SDA, to recover users' profiles in the context of pool mixes. The attack estimates the communication user partners in a mix network. The aim is to estimate the probability of Alice sending a message to Bob; this will derive sender and receiver profiles applicable for all users. The assumptions are: the probability of sending a message from a user to a specific receiver is independent of previous messages; the behavior of all users are independent from one other; any incoming message in the mix is considered *a priori* sent by any

user with a uniform probability; and the parameters used to model the statistical behavior do not change over time.

In [20], a timed binomial pool mix is used, and two privacy criteria to develop dummy traffic strategies are taken into account: (i) increasing the estimation error for all relationships by a constant factor; and (ii) guaranteeing a minimum estimation error for any relationship. The model consists of a set of N senders exchanging messages with a set of M receivers. To simulate the system, consider the same number of senders and receivers and assume users send messages with the same probability. Other work also based on dummy or cover traffic is presented in [21]. This assumes users are not permanently online so, so they cannot send cover traffic uniformly. They introduce a method to reveal Alice's contacts with high probability, addressing two techniques: sending dummy traffic and increasing random delays for messages in the system.

Each one of the previous works has assumed very specific scenarios, but none of them solves the problems that are presented by real-world data. In order to develop an effective attack, the special properties of network human communications must be taken into account. Researchers have hypothesized that some of these attacks can be extremely effective in many real-world contexts. Nevertheless, it is still an open problem under which circumstances and for how long of an observation these attacks would be successful.

4. Framework and Assumptions

This work addresses the problem of retrieving information about relationships or communications between users in a network system, where partial information is obtained. The information used is the number of messages sent and received by each user. This information is obtained in rounds that can be determined by equally-sized batches of messages, in the context of a threshold mix, or alternatively by equal length intervals of time, in the case that the mix method consists of keeping all of the messages retrieved at each time interval and then relaying them to their receivers, randomly reordered.

The basic framework and assumptions needed to develop our method are the following:

- The attacker knows the number of messages sent and received by each user in each round.
- The round can be determined by the system (batches) in a threshold mix context or can be based on regular intervals of time, where the attacker gets the aggregated information about messages sent and received, in the case of a timed mix, where all messages are reordered and sent each period of time.
- The method is restricted, at this moment, to threshold mixing with a fixed batch size or, alternatively, to a timed mix, where all messages received in a fixed time period are relayed randomly, reordered with respect to their receivers.
- No restriction is made from before about the number of friends any user has nor about the distribution of messages sent. Both are considered unknown.
- The attacker controls all users in the system. In our real data application, we aim at all email users of a domain sent and received within this domain.

The method introduced in this work allows one to address these general settings in order to derive conclusions about the relationships between users. Contrary to other methods in the literature, there are

no restrictions about user relationships (number of friends, distribution of messages), and therefore, it can be used in a wider context. Furthermore, our proposition is new in the methodological sense: this is a novel approach to the problem, by means of contingency table setting and the extraction of solutions by sampling.

In an email context, this attack can be used if the attacker has access, at regular time intervals, to the information represented by the number of messages received and the number of messages sent for each user, in a closed domain or intranet, where all users are controlled. This situation can also be extended to mobile communications or social networks and could be used, for example, in the framework of police communication investigations.

5. Marginal Information and Feasible Tables

The attacker obtains, in each round, information about how many messages each user sends and receives. Usually, the sender and receiver set is not the same, even if some users are senders and also receivers in some rounds. Furthermore, the total number of users of the system N is not present in each round, since only a fraction of them are sending or receiving messages. Figure 2 represents a round with only six users.

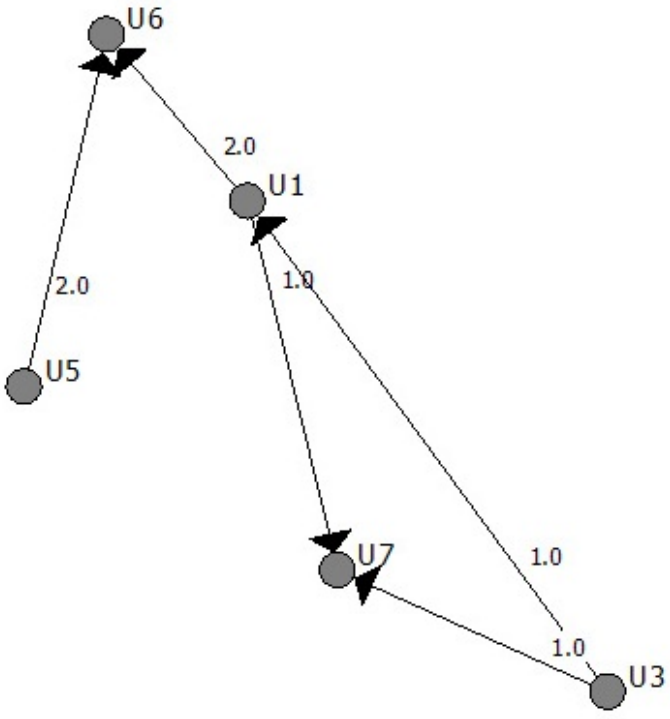


Figure 2. Graphical representation of one round.

The information of this round can be represented in a contingency table (see Table 1), where the element (i, j) represents the number of messages sent from user i to user j:

Table 1. Example of a contingency table.

Senders\Receivers	**U1**	**U6**	**U7**	**Total Sent**
U1	0	2	1	3
U3	1	0	1	2
U5	0	2	0	2
Total received	1	4	2	7

The attacker only sees the information present in the aggregated marginals, which means, in rows, the number of messages sent by each user, and in columns, the number of messages received by each user. In our example, only the sending pairs of vectors (U1 U3 U5) (3 2 2) and receiver pairs of vectors (U1 U6 U7) (1 4 2) are known.

There are many possible tables that can lead to the table with the given marginals that the attacker is seeing, making it impossible, in most cases, to derive direct conclusions about relationships. The feasible space of the tables' solution of the integer programming problem can be very large. In the example, there are only 16 possible different solutions and only one true solution.

Solutions (feasible tables) can be obtained via algorithms, such as the branch and bound algorithm or other integer programming algorithms. In general they do not guarantee covering evenly all possible tables/solutions, since they are primarily designed to converge to one solution. The simulation framework presented in this article allows us to obtain a large quantity of feasible tables (in the most problematic rounds, it takes approximately three minutes to obtain one million feasible tables). In many of the rounds with a moderate batch size, all feasible tables are obtained.

An algorithm that takes into account the information contained over all of the rounds retrieved is developed in the next section.

6. Statistical Disclosure Attack Based on Partial Information

The main objective of the algorithm we propose is to derive relevant information about the relationship (or not) between each pair of users. The information obtained by the attacker is the marginal sums, by rows and columns, of each of the rounds $1, ..., T$, where T is the total number of rounds. Note that in each round, the dimension of the table is different, since we do not take into account users that are not senders (row marginal = 0), nor users that are not receivers (column marginal = 0). We say element (i, j) is "present" at one round if the i and j corresponding marginals are not zero. That means that user i is present in this round as the sender and user j is present as the receiver.

A final aggregated matrix A can be built, summing up all of the rounds and obtaining a table with all messages sent and received from each user for the whole time interval considered for the attack. Each element (i, j) of this final table would represent the number of messages sent by i to j in total. Although the information obtained in each round is more precise and relevant (because of the lower dimension and combinatoric possibilities), an accurate estimate of the final table is the principal objective, because a zero in elements (i, j) and (j, i) would mean no relationship between these users (no messages sent from i to j nor from j to i). A positive number in an element of the estimated final table would mean

that some message is sent in some round, while a zero would mean no messages are sent in any round, that is, no relationship.

We consider all rounds as independent events. The first step is to obtain the higher number of feasible tables that is possible for each round, taking into account time restrictions. This will be the basis of our attack. In order to obtain feasible tables we use Algorithm 1, based on [22]. It consists of filling the table column by column and computing the new bounds for each element before it is generated.

Algorithm 1

1. Begin with column one, row one:
   Generate n_{11} from an integer uniform distribution in the bounds according to Equation (1), where $i = 1, j = 1$.
   Let r be the number of rows.
2. For each row element n_{k1} in this column, if row elements until $k-1$ have been obtained, new bounds for n_{k1} are according to Equation (1):

$$\max(0, (n_{+1} - \sum_{i=1}^{k-1} n_{i1}) - \sum_{i=k+1}^{r} n_{i+}) \leq n_{k1} \leq \min(n_{k+}, n_{+1} - \sum_{i=1}^{k-1} n_{i1}) \quad (1)$$

   The element n_{k1} is then generated by an integer uniform in the fixed bounds.
3. The last row element is automatically filled, since the lower and upper bounds coincide, letting $n_{(k+1)+} = 0$ by convenience.
4. Once this first column is filled, the row margins n_{i+} and total count n are actualized by subtraction of the already fixed elements, and the rest of the table is treated as a new table with one less column.

The algorithm fixes column by column until the whole table is filled.

The time employed depends on the complexity of the problem (number of elements, mean number of messages). In our email data, even for a large number of elements, this has not been a problem. For large table sizes in our applications, it takes approximately 3 min to obtain one million feasible tables in rounds with 100 cells and 10 on a PC with Intel processor 2.3 GHz and 2 GB RAM.

Repeating the algorithm as it is written for each generated table does not lead to uniform solutions, that is some tables are more probable than others due to the order used when filling columns and rows. Since we must consider *a priori* all solutions for a determined round equally possible, two further modifications are made: (i) random reordering of rows and columns before a table is generated; and (ii) once all tables are generated, only distinct tables are kept to make inferences. These two modifications have resulted in an important improvement of the performance of our attack, lowering the mean misclassification rate to about a 20% in our simulation framework.

Deciding the number of tables to be generated poses an interesting problem. Computing the number of distinct feasible tables for a contingency table with fixed marginals is still an open problem that has been addressed via algebraic methods [23] and by asymptotic approximations [24], but in our our case, the margin totals are small and depend on the batch size; therefore, it is not guaranteed that asymptotic approximations hold. The best approximation so far to count the feasible tables is to use the generated tables.

Chen *et al.* [22] show that an estimate of the number of tables can be obtained by averaging over all of the generated tables the value $\frac{1}{q(T)}$ according to the Algorithm 2.

Algorithm 2

① $q(T)$ is the probability of obtaining the table T and is computed iteratively, imitating the simulation process according to Equation (2).

② $q(t_1)$ is the probability of the actual values obtained for Column 1, obtained by multiplying the uniform probability for each row element in its bounds. $q(t_2 \mid t_1)$ and subsequent terms are obtained in the same way, within the new bounds restricted to the precedent columns fixed values:

$$q(T) = q(t_1)q(t_2 \mid t_1)q(t_3 \mid t_1,t_2)...q(t_c \mid t_1,t_2,...,t_{c-1}) \tag{2}$$

The number of feasible tables goes from moderate values, such as 100,000, that can be easily addressed, getting all possible tables via simulation, to very high numbers, such as 10^{13}. Generating all possible tables for this last example would take, with the computer we are using, a Windows 7 PC with 2.3 GHz and 4 GB RAM, at least 51 days. The quantity of feasible tables is the main reason why it is difficult for any deterministic intersection-type attack to work, even with low or moderate user dimensions. Statistical attacks need to consider the relationships between all users to be efficient, because the space of solutions for any individual user is dependent on all other users' marginals. Exact trivial solutions can be, however, found at some time in the long run, if a large number of rounds are obtained.

In our setting, we try to obtain the largest number of tables that we can, given our time restrictions, obtaining a previous estimate of the number of feasible tables and fixing the highest number of tables that can be obtained for the most problematic rounds. However, an important issue is that once a somewhat large number of tables is obtained, good solutions depend more on the number of rounds treated (time horizon or total number of batches considered) than on generating more tables. In our simulations, there is generally a performance plateau in the curve that represents the misclassification rate *versus* the number of tables generated, since a sufficiently high number of tables is reached. This minimum number of tables to be generated depends on the complexity of the application framework.

The final information obtained consists of a fixed number of generated feasible tables for each round. In order to obtain relevant information about relationships, there is a need to fix the most probable zero elements. For each element, the sample likelihood function at zero $\widehat{f}(X \mid p_{ij} = 0)$ is estimated. This is done by computing the percent of tables with that element being zero in each round that the element is present and multiplying the estimated likelihood obtained in all of these rounds (the element will be zero for the final table if it is zero for all rounds).

If we are estimating the likelihood for the element (i, j) and are generating M tables per round, we use the following expressions:

$n_t^{(i,j)}$ = the number of tables with element $(i, j) = 0$ in round t.

$N_{present}$ = the number of rounds with element (i, j) present.

X = the sample data, given by marginal counts for each round.

$$log(\widehat{f}(X \mid p_{ij} = 0)) = -N_{present} \log(M) + \sum_{t=1,(i,j)present}^{T} \log(n_t^{(i,j)}) \tag{3}$$

Final table elements are then ordered by the estimated likelihood at zero, with the exception of elements that were already trivial zeros (elements that represent pair of users that have never been present at any round).

Elements with the lowest likelihood are then considered candidates to insert as a "relationship". The main objective of the method is to detect accurately:

1 cells that are zero with a high likelihood (no relationship $i \to j$);
2 cells that are positive with high likelihood (relationship $i \to j$).

In our settings the likelihood values at $p_{ij} = 0$ are bounded in the interval $[0, 1]$. Once these elements are ordered by most likely to be zero to less, a classification method can be derived based on this measure. A theoretical justification of the consistency of the ordering method is given below.

Proposition 1. *Let us consider, a priori, that for any given round k, all feasible tables, given the marginals, are equiprobable.*

Let p_{ij} be the probability of element (i, j) being zero at the final matrix A, which is the aggregated matrix of sent and received messages over all rounds. Then, the product of the proportion of feasible tables with $x_{ij} = 0$ at each round, Q^{ij} leads to an ordering between elements, such that if $Q^{ij} > Q^{i'j'}$; then, the likelihood of data for $p_{ij} = 0$ is bigger than the likelihood of data for $p_{i'j'} = 0$.

Proof. If all feasible tables for round k are equiprobable, the probability of any feasible table is $p_k = \frac{1}{\#[X]_k}$, where $\#[X]_k$ is the total number of feasible tables in round k.

For elements with $p_{ij} = 0$, it is necessary that $x_{ij} = 0$ for any feasible table. The likelihood for $p_{ij} = 0$ is then:

$$P([X]_k \mid p_{ij} = 0) = \frac{\#[X \mid x_{ij} = 0]_k}{\#[X]_k}$$

where $\#[X \mid x_{ij} = 0]_k$ denotes the number of feasible tables with the element $x_{ij} = 0$.

Let $k = 1, ..., t$ independent rounds. The likelihood at $p_{ij} = 0$, considering all rounds, is:

$$Q^{ij} = \prod_{k=1}^{t} P([X]_k \mid p_{ij} = 0) = \prod_{k=1}^{t} \frac{\#[X \mid x_{ij} = 0]_k}{\#[X]_k}$$

and the log likelihood:

$$\log(Q^{ij}) = \sum_{k=1}^{t} \log(\#[X \mid x_{ij} = 0]_k) - \sum_{k=1}^{t} \log(\#[X]_k)$$

Then, the proportion of elements with $x_{ij} = 0$ at each round leads to an ordering between elements, such that if $Q^{ij} > Q^{i'j'}$, then the likelihood of data for $p_{ij} = 0$ is bigger than the likelihood of data for $p_{i'j'} = 0$. □

Our method is not based on all of the table solutions, but on a consistent estimator of Q^{ij}. For simplicity, let us consider a fixed number of M sampled tables at every round.

Proposition 2. *Let* $[X]_k^1$,...,$[X]_k^M$ *be a random sample of size* M *of the total* $\#[X]_k$ *of feasible tables for round* k*. Let* $w_k^{(i,j)} = \frac{\#[X|x_{ij}=0]_k^M}{M}$ *be the sample proportion of feasible tables with* $x_{ij} = 0$ *at round* k*. Then, the statistic* $q^{ij} = \prod_{k=1}^{t} \frac{\#[X|x_{ij}=0]_k^M}{M}$ *is such that, for any pair of elements* (i, j) and (i', j') , $q^{ij} > q^{i'j'}$ *implies, in convergence, a higher likelihood for* $p_{ij} = 0$ *than for* $p_{i'j'} = 0$*.*

Proof. (1) Let $\#[X]_k$ be the number of feasible tables at round k. Let $[X]_k^1$,...,$[X]_k^M$ be a random sample of size M of the total $\#[X]_k$. Random reordering of columns and rows in Algorithm 1, together with the elimination of equal tables, assures that it is a random sample. Let $\#[X \mid x_{ij} = 0]_k^M$ be the number of sample tables with element $x_{ij} = 0$. Then, the proportion $w_k^{(i,j)} = \frac{\#[X|x_{ij}=0]_k^M}{M}$ is a consistent and unbiased estimator of the true proportion $W_k^{(i,j)} = \frac{\#[X|x_{ij}=0]_k}{\#[X]_k}$. This is a known result from finite population sampling. As $M \to \#[X]_k$, $w_k^{(i,j)} \to W_k^{(i,j)}$.

(2) Let $k = 1, ..., t$ independent rounds. Then, given a sample of proportion estimators $w_1^{(i,j)}, ..., w_t^{(i,j)}$ of $W_1^{(i,j)}, ..., W_t^{(i,j)}$, consider the function

$$f(w_1^{(i,j)}, ..., w_t^{(i,j)}) = \sum_{k=1}^{t} \log(w_k^{(i,j)}) \text{ and } f(W_1^{(i,j)}, ..., W_t^{(i,j)}) = \sum_{k=1}^{t} \log(W_k^{(i,j)}).$$

Given the almost sure convergence of each $w_k^{(i,j)}$ to each $W_k^{(i,j)}$ and the continuity of the logarithm and sum functions, the continuous mapping theorem assures convergence in probability, $f(w_1^{(i,j)}, ..., w_t^{(i,j)}) \xrightarrow{P} f(W_1^{(i,j)}, ..., W_t^{(i,j)})$. Then, $log(q^{ij}) = f(w_1^{(i,j)}, ..., w_t^{(i,j)})$ converges to $log(Q^{ij}) = f(W_1^{(i,j)}, ..., W_t^{(i,j)})$. Since the exponential function is continuous and monotonically increasing, applying the exponential function to both sides leads to the convergence of q^{ij} to Q^{ij}, so that $q^{ij} > q^{i'j'}$ implies, in convergence, $Q^{ij} > Q^{i'j'}$ and, then, higher likelihood for $p_{ij} = 0$ than for $p_{i'j'} = 0$. □

Given all pairs of senders and receivers (i, j) ordered by the statistic q^{ij}, it is necessary to select a cut point in order to complete the classification scheme and to decide whether a pair communicates ($p_{ij} > 0$) or not ($p_{ij} = 0$). That is, it is needed to establish a value c, such that $q^{ij} > c$ implies $p_{ij} = 0$ and $q^{ij} \leq c$ implies $p_{ij} > 0$.The defined statistic q^{ij} is bounded in $[0, 1]$, but this is not strictly a probability, so fixing *a priori* a cut-point, such as 0.5, is not an issue. Instead, there are some approaches that can be used:

1. In some contexts (email, social networks), the proportion of pairs of users that communicate is approximately known . This information can be used to select the cut point from the ordering. That is, if about 20% of pairs of users are known to communicate, the classifier would give a value "0" (no communication) to the upper 80% elements (i, j), ordered by the statistic q^{ij}, and a value "1" (communication) to the lower 20% of elements.
2. If the proportion of zeros is unknown, it can be estimated, using the algorithm for obtaining feasible tables over the known marginals of the matrix A and estimating the proportion of zeros by the mean proportion of zeros over all of the simulated feasible tables.

7. Performance of the Attack

In this section, simulations are used to study the performance of the attack.

Each element (i, j) of the matrix A can be zero (no communication) or strictly positive. The percentage of zeroes in this matrix is a parameter, set *a priori* to observe its influence. In a closed-center

email communications, this number can be between 70% and 99% . However, intervals from 0.1 (high communication density) to 0.9 (low communication density) are used here for different practical situations. Once this percentage is set, a randomly chosen percent of elements are set to zero and then are zero for all of the rounds.

The mean number of messages per round for each positive element (i, j) is also set *a priori*. This number is related, in practice, to the batch size that the attacker can obtain. As the batch size (or time length interval of the attack) decreases, the mean number of messages per round decreases, making the attack more efficient.

Once the mean number of messages per round is determined for each positive element (λ_{ij}), a Poisson distribution with mean λ_{ij}, $P(\lambda_{ij})$, is used to generate the number of messages for each element, for each of the rounds.

External factors, given by the context (email, social networks, *etc*.) that have an effect on the performance of the method are monitored to observe their influence:

1. The number of users: In a network communication context with N users, there exist N potential senders and N receivers in total, so that the maximum dimension of the aggregated matrix A is N^2. As the number of users increases, the complexity of round tables and the number of feasible tables increases, so that it could negatively affect the performance of the attack.
2. The percent of zero elements in the matrix A: These zero elements represent no communication between users. As will be seen, this influences the performance of the method.
3. The mean frequency of messages per round for positive elements: This is directly related to the batch size, and when it increases, the performance is supposed to be affected negatively.
4. The number of rounds: As the number of rounds increases, this is supposed to improve the performance of the attack, since more information is available. One factor related to the settings of the attack method is also studied.
5. The number of feasible tables generated by round: This affects computing time, and it is necessary to study to what extent it is useful to obtain too many tables. This number can be variable, depending on the estimated number of feasible tables for each round .

The algorithm results in a binary classification, where zero in an element (i, j) means no relationship of sender-receiver from i to j and one means a positive relationship of sender-receiver.

Characteristic measures for binary classification tests include the sensitivity, specificity, positive predictive value and negative predictive value. Letting TP be true positives, FP false positives, TN true negatives and FN false negatives:

Sensitivity=$\frac{TN}{TN+FP}$ measures the capacity of the test to recognize true negatives.
Specificity=$\frac{TP}{TP+FN}$ measures the capacity of the test to recognize true positives.
Positive predictive value=$\frac{TP}{TP+FP}$ measures the precision of the test to predict positive values.
Negative predictive value =$\frac{TN}{TN+FN}$ measures the precision of the test to predict positive values.
Classification rate=$\frac{TN+TP}{TN+TP+FN+FP}$ measures the percent of elements well classified.

Figures 3 and 4 show the simulation results. When it is not declared, values of $p_0 = 0.7$, $\lambda = 2$, $N = 50$ users and the number of rounds = 100 are used as base values.

Figure 3 shows that as the number of cells (N^2, where N is the number of users) increases and the percent of cells that are zero decreases, the number of feasible tables per round increases. For a moderate number of users, such as 50, the number of feasible tables is already very high, greater than 10^{20}. This does not have a strong effect on the main results, except for lower values. As can be seen in Figure 4, once a sufficiently high number of tables per round is generated, increasing this number does not lead to significant improvement of the correct classification rate.

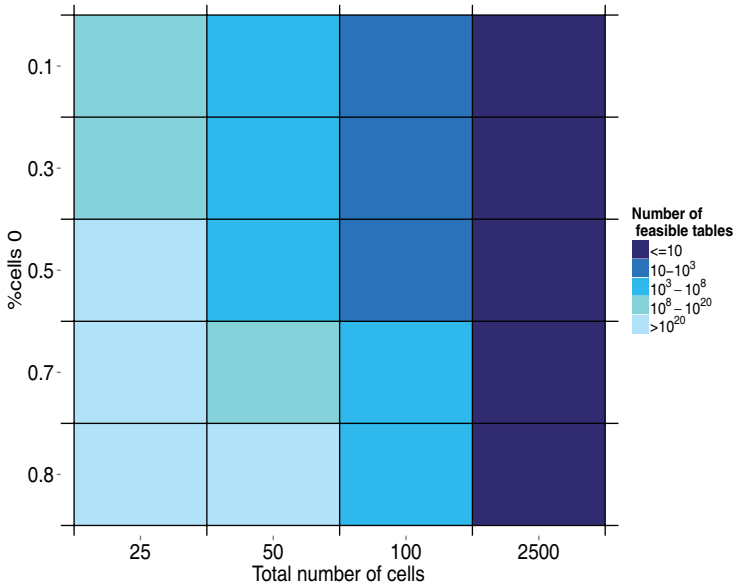


Figure 3. Number of feasible tables per round, depending on % of cells of zero and the total number of cells.

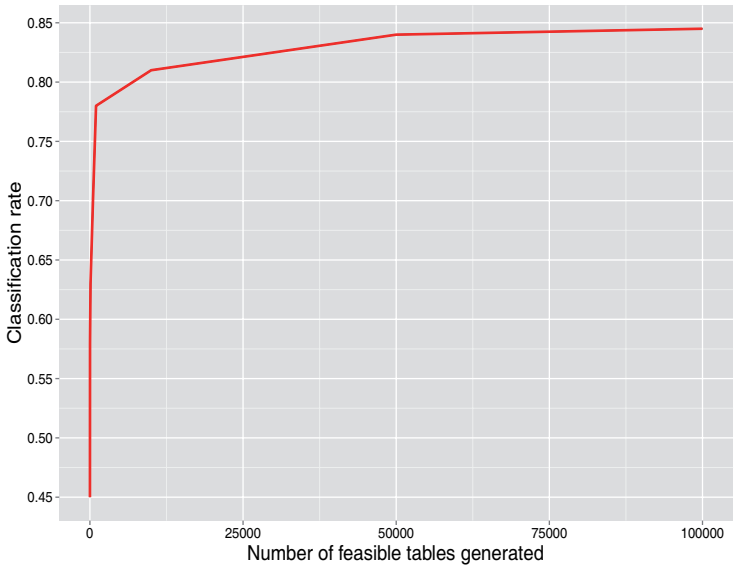


Figure 4. Classification rate as function of the number of feasible tables generated per round.

Figure 5 shows that the minimum classification rate is attained at a percent of cells of zero (users that do not communicate) near 0.5. As this percent increases, the true positive rate decreases, and the true negative rate increases.

As the attacker gets more information, that is more rounds are retrieved, the classification rate gets better. Once a high number of rounds is obtained, there is no further significant improvement, as is shown in Figure 6.

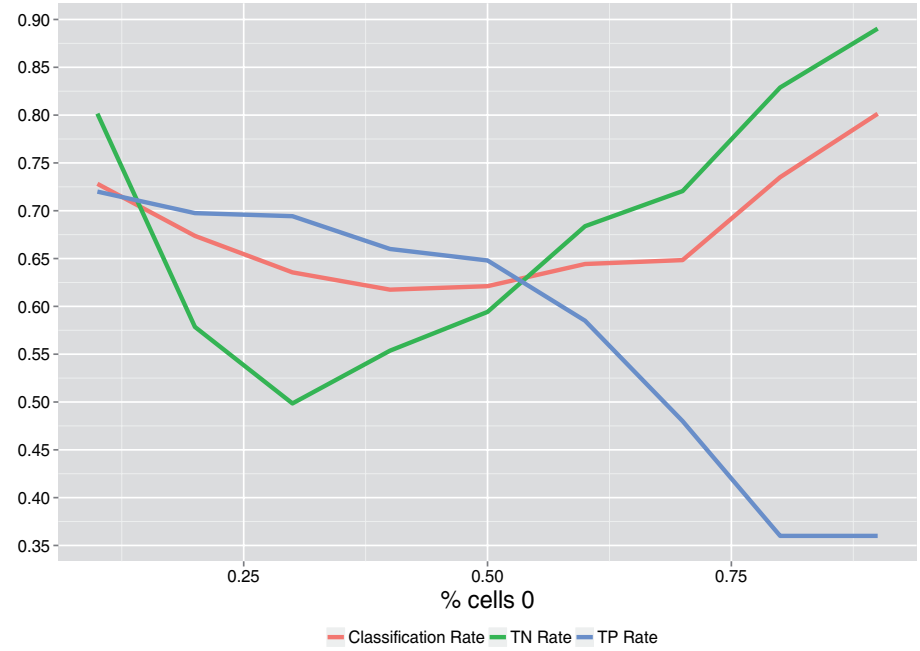


Figure 5. Classification rate, true positive rate and true negative rate *vs.* the percent of cells of zero.

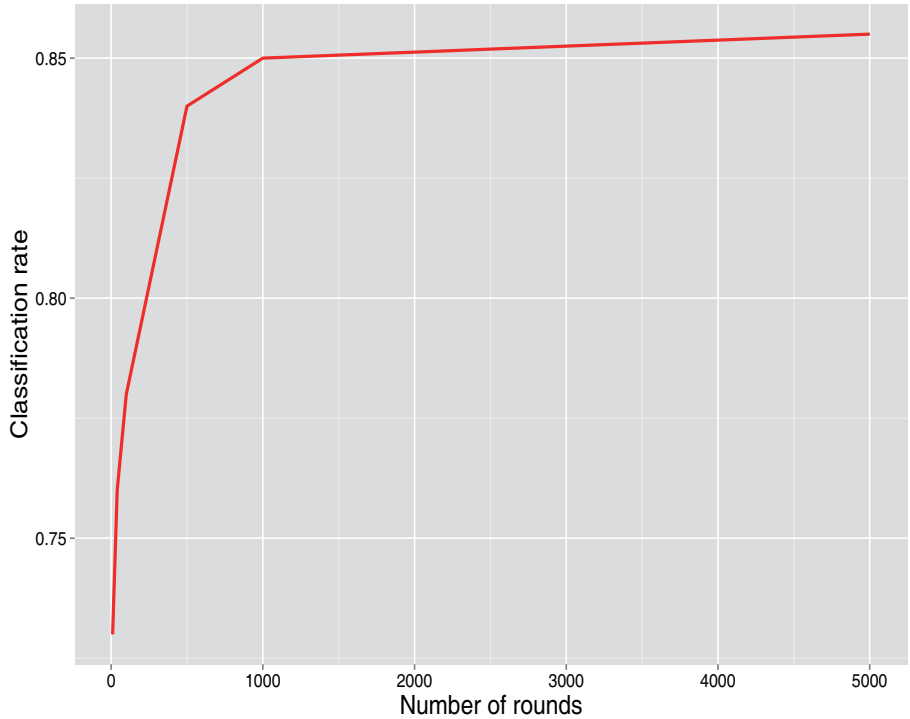


Figure 6. Classification rate *vs.* the number of rounds obtained.

In Figure 7, it is shown that as the number of messages per round (λ) for users that communicate increases, the classification rates decrease. This is a consequence of the complexity of the tables involved

(more feasible tables). This number is directly related to the batch size, so it is convenient for the attacker to obtain data in small batch sizes and for the defender to group data in large batch sizes, leading to lower latency.

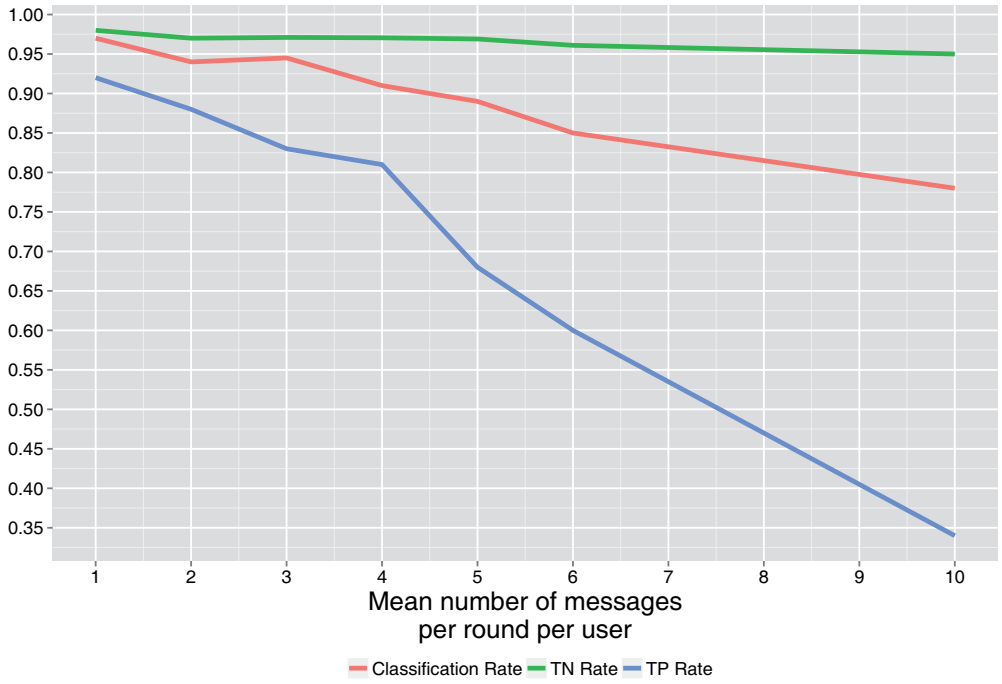


Figure 7. Classification rates *vs.* the mean number of messages per round.

The complexity of the problem is also related to the number of users, as can be seen in Figure 8, where the classification rate decreases as the number of users increases.

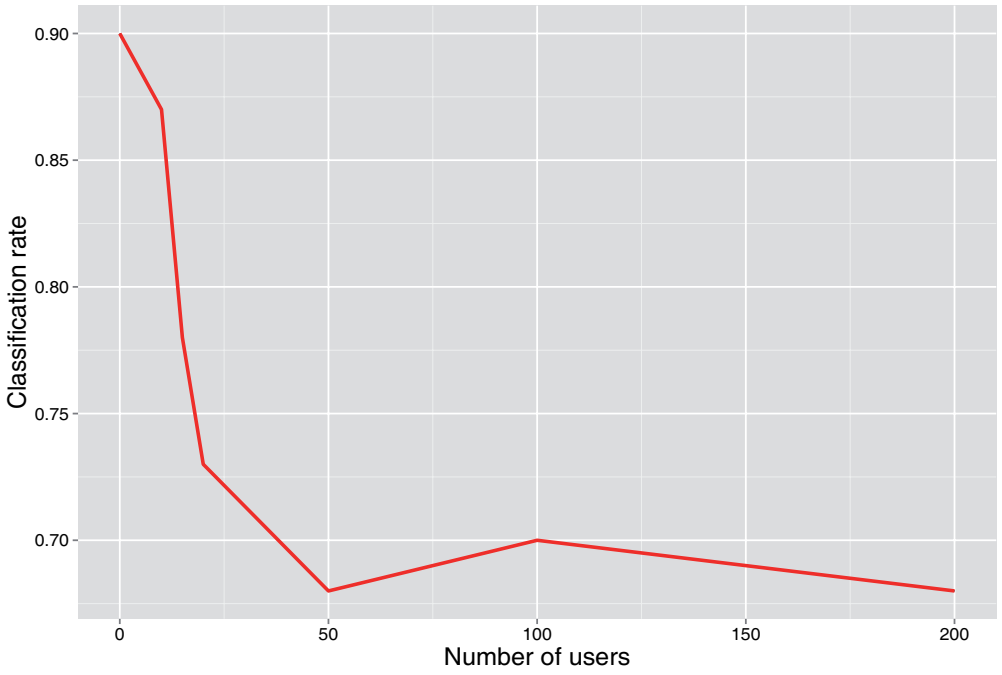


Figure 8. Classification rate *vs.* the number of users.

8. Conclusions

This work presents a method to detect relationships (or non-existent relationships) between users in a communication framework, when the retrieved information is incomplete. The method can be extended to other settings, such as pool mixes, or situations where additional information can be used. Parallel computing has also been successfully used in order to obtain faster results. The method can also be used for other communication frameworks, such as social networks or peer-to-peer protocols, and for real de-anonymization problems not belonging to the communications domain, such as disclosing public statistical tables or forensic research. More research has to be done involving the selection of optimal cut points, the optimal number of generated tables or further refinements of the final solution, which may be through the iterative filling of cells and cycling the algorithm.

Acknowledgments

Part of the computations of this work were performed in EOLO, the HPC of Climate Change of the International Campus of Excellence of Moncloa, funded by MECD and MICINN. This is a contribution to CEI Moncloa. The authors would like to acknowledge funding support provided by Red Garden Technologies Mexico. Also, the authors thank the comments of MSc. Facundo Armenta Armenta for his valuable feedback to carry out this project.

Author Contributions

J. Portela, L. J. García Villalba and A. G. Silva Trujillo are the authors who mainly contributed to this research, performing experiments, analysis of the data and writing the manuscript. A. L. Sandoval Orozco and T.-H. Kim analyzed the data and interpreted the results. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Chaum, D.L. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM* **1981**, *24*, 84–88.
2. Dingledine, R.; Mathewson, N.; Syverson, P. Tor: The second generation onion router. In Proceedings of the 13th USENIX Security Syposium, 9–13 August 2004; pp. 303–320.
3. Pfitzmann, A.; Pfitzmann, B.; Waidner, M. ISDN-Mixes : Untraceable communication with small bandwidth overhead. In *GI/ITG Conference: Communication in Distributed Systems*; Springer-Verlag: Berlin, Germany, 1991; pp. 451–463.
4. Gulcu, C.; Tsudik, G. Mixing E-mail with Babel. In Proceedings of the Symposium on Network and Distributed System Security, San Diego, CA, USA, 22–23 February 1996; pp. 2–16.

5. Moller, U.; Cottrell, L.; Palfrader, P.; Sassaman, L. Mixmaster Protocol Version 2. Internet Draft Draft-Sassaman-Mixmaster-03, Internet Engineering Task Force, 2005. Available online: http://tools.ietf.org/html/draft-sassaman-mixmaster-03 (accessed on 9 February 2015).
6. Danezis, G.; Dingledine, R.; Mathewson, N. Mixminion: Design of a type III anonymous remailer protocol. In Proceedings of the 2003 Symposium on Security and Privacy, Oakland, CA, USA, 11–14 May 2003; pp. 2–5.
7. Serjantov, A.; Sewell, P. Passive Attack Analysis for Connection-Based Anonymity Systems. In Proceedings of European Symposium on Research in Computer Security, Gjovik, Norway, 13–15 October 2003; pp. 116–131.
8. Raymond, J.F. Traffic analysis: Protocols, attacks, design issues, and open problems. In Proceedings of the International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability, Berkeley, CA, USA, 25–26 July 2000; pp. 10–29.
9. Agrawal, D.; Kesdogan, D. Measuring anonymity: The disclosure attack. *IEEE Secur. Priv.* **2003**, *1*, 27–34.
10. Danezis, G. Statistical Disclosure Attacks: Traffic Confirmation in Open Environments. In *Proceedings of Security and Privacy in the Age of Uncertainty*; De Capitani di Vimercati, S., Samarati, P., Katsikas, S., Eds.; IFIP TC11, Kluwer: Athens, Greece, 2003; pp. 421–426.
11. Danezis, G.; Serjantov, A. Statistical disclosure or intersection attacks on anonymity systems. In Proceedings of the 6th International Conference on Information Hiding, Toronto, ON, Canada, 23–25 May 2004; pp. 293–308.
12. Mathewson, N.; Dingledine, R. Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. In Proceedings of Privacy Enhancing Technologies Workshop, Toronto, ON, Canada, 26–28 May 2004; pp. 17–34.
13. Danezis, G.; Diaz, C.; Troncoso, C. Two-sided statistical disclosure attack. In Proceedings of the 7th International Conference on Privacy Enhancing Technologies, Ottawa, ON, Canada, 20–22 June 2007; pp. 30–44.
14. Troncoso, C.; Gierlichs, B.; Preneel, B.; Verbauwhede, I. Perfect Matching Disclosure Attacks. In Proceedings of the 8th International Symposium on Privacy Enhancing Technologies, Leuven, Belgium, 23–25 July 2008; pp. 2–23.
15. Danezis, G.; Troncoso, C. Vida: How to Use Bayesian Inference to De-anonymize Persistent Communications. In Proceedings of the 9th International Symposium on Privacy Enhancing Technologies, Seattle, WA, USA, 5–7 August 2009; pp. 56–72.
16. Kesdogan, D.; Pimenidis, L. The hitting set attack on anonymity protocols. In Proceedings of the 6th International Conference on Information Hiding, Toronto, ON, Canada, 23–25 May 2004; pp. 326–339.
17. Mallesh, N.; Wright, M. The reverse statistical disclosure attack. In Proceedings of the 12th International Conference on Information Hiding, Calgary, AB, Canada, 28–30 June 2010; pp. 221–234.
18. Bagai, R.; Lu, H.; Tang, B. On the Sender Cover Traffic Countermeasure against an Improved Statistical Disclosure Attack. In Proceedings of the IEEE/IFIP 8th International Conference on Embedded and Ubiquitous Computing, Hong Kong, China, 11–13 December 2010; pp. 555–560.

19. Perez-Gonzalez, F.; Troncoso, C.; Oya, S. A Least Squares Approach to the Static Traffic Analysis of High-Latency Anonymous Communication Systems. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1341–1355.
20. Oya, S.; Troncoso, C.; Pérez-González, F. Do Dummies Pay Off? Limits of Dummy Traffic Protection in Anonymous Communications. Available online: http://link.springer.com/chapter/10.1007/978-3-319-08506-7_11 (accessed on 3 February 2015).
21. Mallesh, N.; Wright, M. An analysis of the statistical disclosure attack and receiver-bound. *Comput. Secur.* **2011**, *30*, 597–612.
22. Chen, Y.; Diaconis, P.; Holmes, S.P.; Liu, J.S. Sequential Monte Carlo methods for statistical analysis of tables. *J. Am. Stat. Assoc.* **2005**, *100*, 109–120.
23. Rapallo, F. Algebraic Markov Bases and MCMC for Two-Way Contingency Tables. *Scand. J. Stat.* **2003**, *30*, 385–397.
24. Greenhilla, C.; McKayb, B.D. Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums. *Adv. Appl. Math.* **2008**, *41*, 459–481.

URSI2015

LIBRO DE ACTAS

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

PAMPLONA

XXX Simposium Nacional de la Unión Científica Internacional de Radio
2, 3 y 4 de septiembre

International Workshop on THz Engineering
1 de septiembre

041 - ACIO: Access Control in Organizations, a Sensing Enterprise Approach (12:40-13:00h.)

Iván Prada Gamallo[1], Alicia González Cabestreros[1], Oscar Lázaro de Barrio[1], Mikel Uriarte Itsazelaia[2], Oscar López Pérez[2], Jordi Blasi Uribarri[2], Eneko Olivares Gorriti[3], Carlos E. Palau Salvador[3]

[1]Asociación Innovalia; [2]Nextel S.A.; [3]Universitat Politécnica de Valencia

The paper analyses access control in organization from a use case approach, from the results achieved in the ACIO project. Delivering enhanced physical/logical access control models and solutions based in the sensing enterprise paradigm. The paper views the access control services from a usability and security perspective, providing comparative information on different mechanisms that helps to improve the fluency, monitoring and efficiency of work and assures the real-time authorization of physical and logical users to specific tasks and areas within an organization while maintaining the required level of information security. ACIO has developed services and mechanisms for access control that make use of internal organisation data as well as location and other attribute information. The application domain used to validate the proposal and present performance evaluation results is port transportation, and explicitly the Port of Valencia.

042 - Sistema para la Detección de Comunicaciones entre usuarios de Correo Electrónico (13:00-13:20h.)

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba

Universidad Complutense de Madrid, España

The disclosure attacks methods found in literature perform simulations in very specific scenarios, any of them has applied real data to show its performance. We consider this an important issue because users behavior in a communication system has certain peculiarities such as: the average number of email messages sent by user, the different number of friends for each user, among others. The statistical disclosure attacks must consider the relationship between users in order to be efficient, because the solution space for a single user depends on the marginal values of other users. The aim of our work is to show the results after carrying out our statistical disclosure attack performed with data of an anonymous communication system email from an university community.

Sistema para la Detección de Comunicaciones entre Usuarios de Correo Electrónico

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba
asilva@fdi.ucm.es, jportela@estad.ucm.es, javiergv@fdi.ucm.es
Grupo de Análisis, Seguridad y Sistemas (GASS)
Departamento de Ingeniería de Software e Inteligencia Artificial (DISIA)
Facultad de Informática, Universidad Complutense de Madrid (UCM)
Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040, Madrid

***Abstract*— The disclosure attacks methods found in literature perform simulations in very specific scenarios, any of them has applied real data to show its performance. We consider this an important issue because users behavior in a communication system has certain peculiarities such as: the average number of email messages sent by user, the different number of friends for each user, among others. The statistical disclosure attacks must consider the relationship between users in order to be efficient, because the solution space for a single user depends on the marginal values of other users. The aim of our work is to show the results after carrying out our statistical disclosure attack performed with data of an anonymous communication system email from an university community.**

I. Introducción

A dos años de las revelaciones de un técnico de la CIA respecto a los programas de registro y espionaje cibernético del gobierno de EUA, el tema sigue siendo polémico. Estudios demuestran luego de tales revelaciones, ha aumentado el intercambio de opiniones, debates, artículos, que han motivado a algunos a cambiar sus hábitos al navegar por Internet, la forma en que usan su correo electrónico, los buscadores, las redes sociales o sus dispositivos móviles [1].

En los últimos años se han desarrollado sistemas para proteger la privacidad de las personas, herramientas que permiten la navegación anónima en Internet [2], el uso de correos electrónicos también en un entorno anónimo, entre otros [3] [4] [5]. A la par de estos sistemas, también han surgido técnicas para contrarrestar las funciones anónimas como aquellas llamadas de revelación de identidad, que permiten conocer quién se comunica con quién en un entorno anónimo.

Los métodos encontrados en la literatura de ataques de revelación de identidades realizan simulaciones en escenarios muy específicos, y en nuestro conocimiento, no existe alguna que haya llevado a cabo pruebas con datos reales que son un factor importante dado que el comportamiento de los usuarios en un sistema de comunicación tiene ciertas particularidades como: el promedio de mensajes enviados por cada usuario, la variabilidad en el número de receptores o amigos, entre otros. En los ataques estadísticos de revelación de identidades se debe considerar las relaciones entre cada uno de los usuarios para ser eficientes, debido a que el espacio de soluciones para un usuario depende de los valores marginales del resto de los usuarios.

El objetivo de nuestro trabajo es mostrar los resultados luego de llevar a cabo un ataque de revelación de identidades realizado con datos reales, datos de un sistema de comunicación de correo electrónico anónimo de una comunidad universitaria. Organizamos nuestro trabajo de la siguiente manera. En la sección II explicamos brevemente nuestro algoritmo. Abordamos los resultados obtenidos en la sección III y finalmente, en la sección IV mostramos nuestras conclusiones y trabajos futuros.

II. Algoritmo

El marco base y los supuestos necesarios para llevar a cabo nuestro ataque son:

- El atacante conoce el número de mensajes enviados y recibidos por cada usuario por ronda.
- Una ronda puede definirse por intervalos de tiempo en los que el atacante observa la red o bien determinados por el sistema (lotes).
- En nuestra aplicación hemos utilizado datos de diferente lote, asi como también hemos variado el horizonte temporal de muestreo.
- Se considera cada una de las rondas como eventos independientes.
- El algoritmo está considerado para un sistema mix simple.
- No existen limitantes respecto al número de receptores o amigos de cada usuario, ni tampoco de la distribución de mensajes enviados. Ambos se consideran desconocidos.
- El atacante controla todos los usarios de la red.

La información de una ronda puede definirse a través de una tabla de contingencia como se muestra en la Tabla I, donde los renglones son los emisores y las columnas son los receptores. De esta manera, la celda (i, j) contiene el número de mensajes

TABLE I
Ejemplo de una tabla de contingencia

Emisores \ Receptores	U1	U6	U7	Total de enviados
U1	0	2	1	3
U3	1	0	1	2
U5	0	2	0	2
Total de recibidos	1	4	2	7

enviados por el usuario i al usuario j.

Sin embargo, el atacante solo ve las marginales de la tabla que corresponden al número total de mensajes enviados y recibidos.

El algoritmo da como resultado información relevante de la existencia de relación o no relación entre cada par de usuarios. Los pasos a seguir son los siguientes:

1. El atacante obtiene información de n rondas a través de la observación de la red tal como se describe en [6].
2. Se generan las tablas factibles para cada ronda de acuerdo al algoritmo utilizado en [7].
3. Se calculan los elementos que tienen mayor probabilidad de ser cero. Lo anterior, según el porcentaje de tablas factibles donde el elemento es cero en las rondas en donde esta presente. El elemento será cero en la tabla final si es cero en todas las rondas.
4. Se lleva a cabo la clasificación de cada celda (i, ij) donde 1 indica que existe relación entre el usuario i y el usuario j; y 0 indica que no existe relación entre ellos.

Bajo el contexto de una red de comunicación con N usuarios, existen en total N emisores y receptores potenciales. La tabla final A de dimensión N^2 se construye resumiendo todas las rondas y obteniendo una tabla con todos los mensajes enviados y recibidos por cada usuario en el intervalo de tiempo del ataque.

En [7] se encuentran los detalles del algoritmo, en el presente trabajo discutiremos los resultados obtenidos luego de su aplicación en datos reales de un sistema de comunicación de correo electrónico anónimo. También mostraremos algunas consideraciones derivadas de las distintas simulaciones que realizamos, mismas que serán la base para trabajos futuros a llevar a cabo.

III. Aplicación del Algoritmo

Se aplicó el algoritmo en datos de correo electrónico de 32 subdominios. Dichos datos corresponden a correos intra-facultades de la Universidad Complutense de Madrid del año 2010 previamente anonimizados y fueron proporcionados por el Centro de Cálculo de la UCM. De esta manera cada subdominio corresponde a una facultad o entidad administrativa. Cabe mencionar que cada uno de los subdominios cuenta con un número de usuarios diferente.

En subdominios con poco número de usuarios, se tiene un flujo de datos reducido, en tanto aquellos subdominios con mayor número de receptores/emisores potenciales maneja un flujo considerable. En la Figura 1 presentamos el número de emisores y receptores por subdomio.

Para empezar se obtienen las tablas factibles para cada una de las rondas y se calculan los elementos que tienen más probabilidad de ser cero y uno. Por los resultados del algoritmo consideramos un test de clasificación binaria, en donde utilizamos métricas que evaluan la sensitividad, la especificidad, el valor predictivo positivo y el valor predictivo negativo. Consideramos TP a los verdaderos positivos, FP a los falsos positivos, TN a los verdaderos negativos y FN a los falsos negativos:

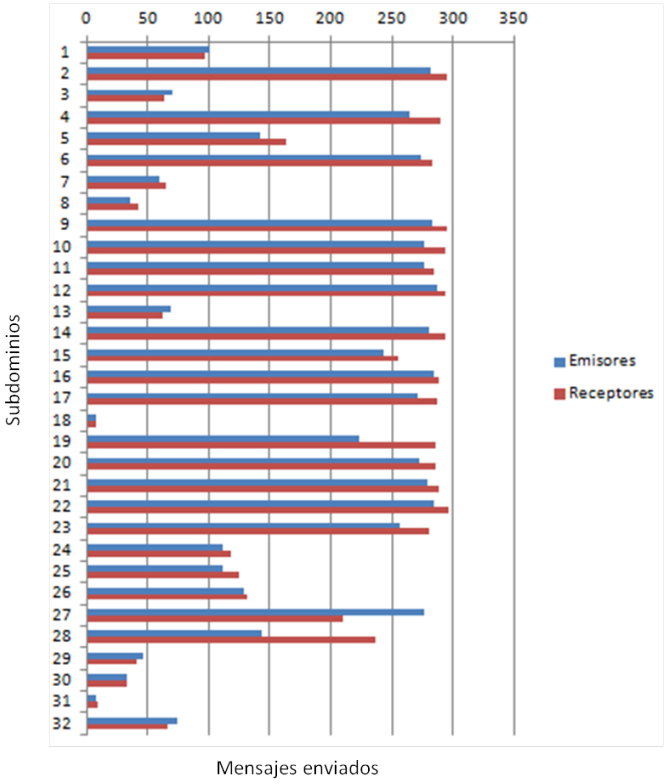


Fig. 1
Número de emisores y receptores por subdominio.

- Sensitividad = TN / TN + FP mide la capacidad del test para reconocer valores negativos verdaderos.
- Especificidad = TP / TP + FN mide la capacidad del test para reconocer valores positivos verdaderos.
- Valor predictivo positivo (VPP) = TP / TP + FP mide la precisión del test para predecir valores positivos.
- Valor predictivo negativo (VPN) = TN / TN + FN mide la precisión del test para predecir valores negativos.

Una tasa de clasificación cercana a 1 representa una predicción perfecta. La Figura 2 muestra la tasa de clasificación para cada subdominio.

Los resultados de sensitividad, que se le conoce tambien como la fracción de verdaderos positivos se muestran en la Figura 3 para un horizonte temporal de 12 meses.

Otra de las variantes que realizamos para ver con más detalle el comportamiento de nuestro algoritmo fue tomar los muestreos para diferentes períodos de tiempo 1, 2, 3 y 12 meses. A efectos de comparar el comportamiento de nuestro algoritmo en subdominios con pocos y muchos usuarios hemos elegido las facultades 32 y 22 respectivamente. En la Figura 4 vemos la tasa de clasificación obtenida con nuestro algoritmo para diferentes horizontes de tiempo el subdominio 32 que tiene poco número de usuarios. En tanto la Figura 5 muestra los resultados obtenidos para el subdominio 22 que cuenta con un gran número de usuarios.

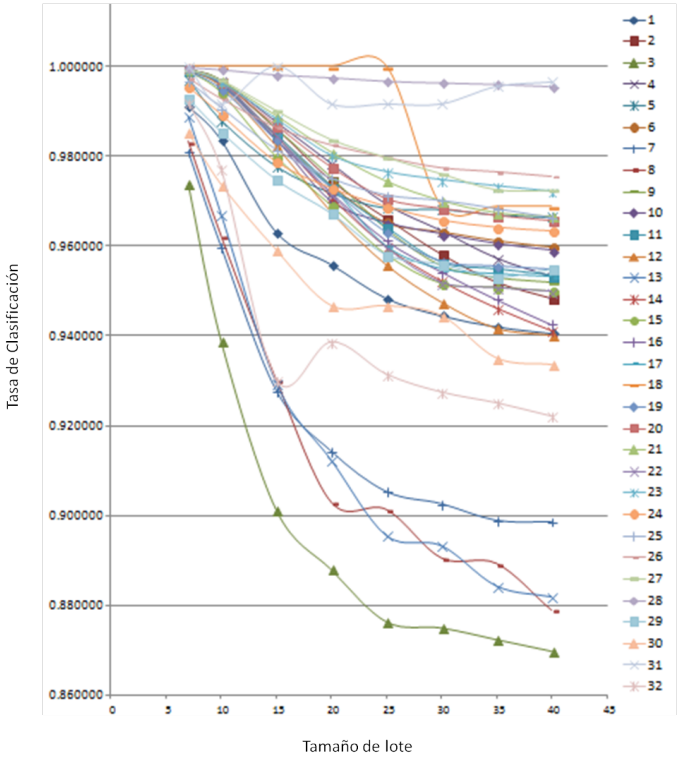


Fig. 2
TASA DE CLASIFICACIÓN POR SUBDOMINIO EN UN LOTE DE 12 MESES.

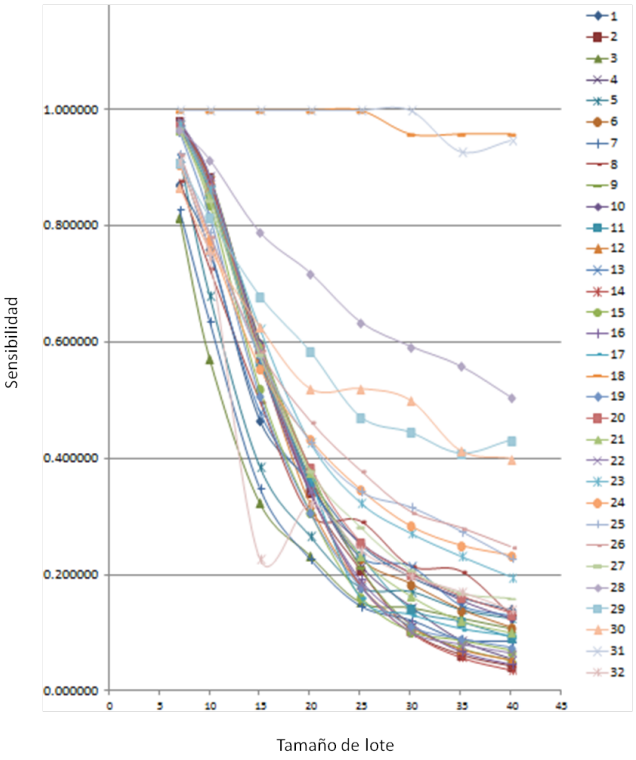


Fig. 3
SENSITIVIDAD POR SUBDOMINIO EN UN LOTE DE 12 MESES.

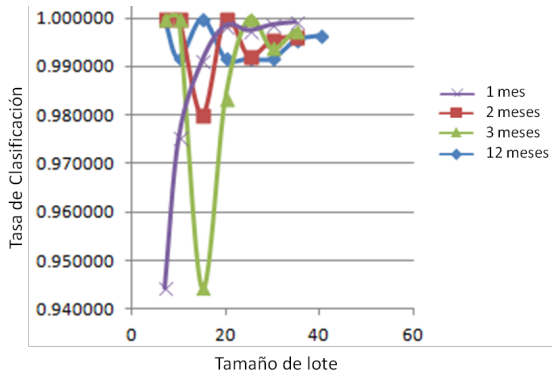


Fig. 4
TASA DE CLASIFICACIÓN POR MESES DEL SUBDOMINIO 32.

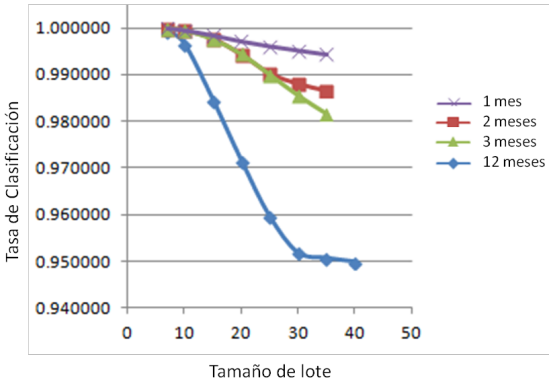


Fig. 5
TASA DE CLASIFICACIÓN POR MESES DEL SUBDOMINIO 22.

Para subdominios con menor número de emisores y receptores potenciales N se obtienen mejores resultados de clasificación. Respecto al tamaño de lote se puede observar que para pocos usuarios, entre más grande es el tamaño de lote, la tasa de clasificacion mejora; en cambio en subdominios donde hay mayor número de usuarios, entre más grande es el tamaño de lote la tasa de clasificación obtenida se va degradando.

Si el atacante obtiene mayor información, esto es, entre más rondas recolecte, la tasa de clasificación es mejor. También es importante señalar que en algunos casos a pesar de variar el tamaño del lote no conduce a mejoras significativas. Podemos decir que una celda es cero fijo cuando se detecta que el emisor y receptor no han coincidido o no han estado presentes en una misma ronda.

Las Figuras 6 y 7 muestran los resultados del número de ceros fijos que se generan en un entorno con pocos o muchos usuarios del subdominio 32 y 22 tomados como base. Ambos casos se comportan de la misma forma al variar el tamaño de lote.

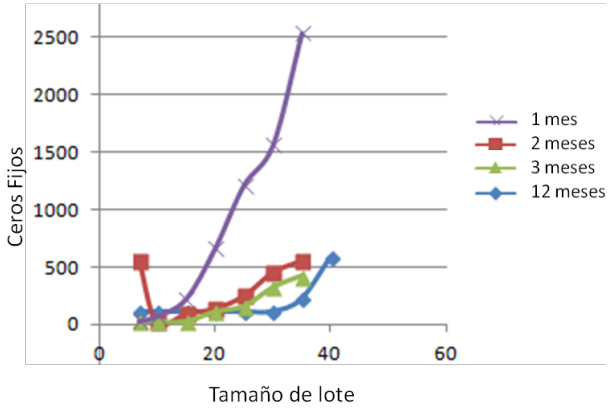


Fig. 6
NÚMERO DE CEROS FIJOS POR MESES DE SUBDOMINIO32.

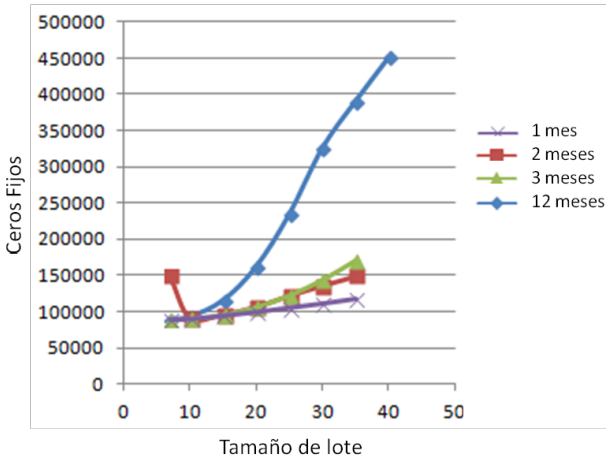


Fig. 7
NÚMERO DE CEROS FIJOS POR MESES DE SUBDOMINIO22.

Para obtener una mejor clasificación hemos utilizado puntos de corte, lo que ha arrojado valores altos de predicción positiva (VPP) y bajos de predicción negativa (VPN). Esto significa que cuando el algoritmo clasifica una celda (i, j) como elementos que sí se comunican esto es bastante preciso. Sin embargo, cuando la celda es clasificada como 0 (no hay comunicación) suele ser menos preciso. La Tabla II presenta los verdaderos y falsos positivos y negativos, el valor predictivo positivo y negativo de algunas facultades con tamaño de lote 20. Se observa que el algoritmo no detecta un alto porcentaje de comunicaciones verdaderos lo que se traduce en una baja sensitividad. Si el objetivo del atacante es simplemente capturar con la mejor precisión el mayor numero de pares de elementos que se comunican, el algoritmo propuesto es el adecuado.

TABLE II
TASAS OBTENIDAS PARA DIFERENTES FACULTADES

Facultad	TP	FP	TN	FN	VPP	VPN	Sensitividad
1	264	0	6818	143	1	0.979	0.648
2	1259	1	88831	510	0.999	0.994	0.711
3	231	1	4088	304	0.995	0.924	0.43
4	973	0	89177	451	1	0.994	0.68
5	415	0	28322	504	1	0.98	0.45

IV. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo presenta los resultados obtenidos de la aplicación de nuestro algoritmo que detecta las relaciones existentes o no existentes entre usuarios de un sistema de correo electrónico anónimo de una universidad. De acuerdo a los resultados obtenidos podemos observar que existen variantes en los mismos, al aplicarlo a sistemas con pocos o muchos usuarios. Asimismo pudimos notar que entre mayor sea la información obtenida por el atacante, dará mejores resultados. De los valores altos de predicción positiva (VPP) y bajos de predicción negativa (VPN) podemos decir que cuando el algoritmo detecta que un par de usuarios se comunican es altamente preciso, sin embargo el caso contrario suele ser menos preciso. Dentro de los trabajos futuros podemos aplicar nuestro algoritmo teniendo en consideración los resultados obtenidos.

AGRADECIMIENTOS

Los autores agradecen el apoyo brindado por el "Programa de Financiación de Grupos de Investigación UCM validados de la Universidad Complutense de Madrid - Banco Santander". Los autores también agradecen el apoyo proporcionado por Red Garden Technologies (México). Asimismo, los autores agradecen los valiosos comentarios y sugerencias de MSc. Facundo Armenta Armenta que han contribuido a la mejora de este trabajo.

REFERENCES

[1] L. Rainie and M. Madden, "Americans' Privacy Strategies Post-Snowden", disponible en www.pewinternet.org/2015/03/16/americans-privacy-strategies-post-snowden/, 2015.
[2] R. Dingledine, N. Mathewson and P. Syverson, "Tor: The Second Generation Onion Router", in *Proceedings of the 13th USENIX Security Symposium*, Sand Diego, CA, USA, August 9-13, 2004, pp. 302-320.
[3] C. Gulcu and G. Tsudik, "Mixing E-mail with Babel", in *Proceedings of the Symposium on Network and Distributed System Security*, San Diego, CA, USA, February 22-23, 1996, pp. 2-16.
[4] U. Moller, L. Cottrell, P. Palfrader and L. Sassaman, "Mixmaster Protocol Version 2. RFC 3668, *Internet Engineering Task Force*, June 2005.
[5] G. Danezis, R. Dingledine and R. Mathewson, "Mixminion: Design of a Type III Anonymous Remailer Protocol", in *Proceedings of the Symposium on Security and Privacy*, Washington, DC, USA, May 11-14, 2003, pp. 2-15.
[6] G.Danezis, "Statistical Disclosure Attacks: Traffic Confirmation in Open Enviroments", in *Proceedings of Security and Privacy in the Age of Uncertainty*, 2003, pp. 421-426.
[7] J. Portela, L. J. García Villalba, A. Silva, A. L. Sandoval Orozco and T.-H. Kim, "Extracting Association Patterns in Network Communications", *Sensors*, vol. 15, no. 2, February 2015.

J Supercomput
DOI 10.1007/s11227-015-1524-7

Disclosing user relationships in email networks

Javier Portela[1] · **Luis Javier García Villalba**[1] · **Alejandra Silva Trujillo**[1] · **Ana Lucila Sandoval Orozco**[1] · **Tai-Hoon Kim**[2]

Abstract To reveal patterns of communications of users in a network, an attacker may repeatedly obtain partial information on behavior and finally derive relationships between pairs of users through the modeling of this statistical information. This work is an enhancement of a previously presented statistical disclosure attack. The improvement of the attack is based on the use of the EM algorithm to improve the estimation of messages sent by users and to derive what pairs of users really communicate. Two methods are presented using the EM algorithm and the best method is used over real email data over 32 different network domains. Results are encouraging with high classification and positive predictive value rates.

✉ Tai-Hoon Kim
taihoonn@empal.com

Javier Portela
jportela@estad.ucm.es

Luis Javier García Villalba
javiergv@fdi.ucm.es

Alejandra Silva Trujillo
asilva@fdi.ucm.es

Ana Lucila Sandoval Orozco
asandoval@fdi.ucm.es

1 Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), Faculty of Information Technology and Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, Spain

2 School of Information and Computing Science, University of Tasmania, Centenary Building, Room 350, Private Bag 87, Hobart, TAS 7001, Australia

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

Keywords Anonymity · Mixes · Network communications · Statistical disclosure attack

1 Introduction

Nowadays, privacy is one of the most important topics for security communications. The extensive use of mobile devices, Internet and online applications is a growing phenomenon that has led to the development of new techniques and technologies to provide users safe environments. Several research groups have been given the task of developing applications that support this purpose. It is well known that the government and organizations want to take advantage of every piece of data we leave on the Internet, purchasing habits, health records, and many other actions for legal (or non-legal) purposes. Such practices are covered under the argument that monitoring online activities is necessary to detect potential threats that could undermine national security.

Privacy enhancing technologies is a term employed to define the use of technology which helps accomplish personal information protection and in business context this technology can protect corporate confidential information and the integrity of data. One of the basic characteristic of these technologies is the way to transmit messages with high or low latency. Low latency systems are focused on fast transmission of messages (for example, web browsing and other interactive services), and high-latency systems provide their users with an exchange of messages in an anonymous or pseudonymous way.

Even when data are encrypted, there is still an open target to identify significant information such as the sender and receiver network address. An adversary can see all the messages over the network and execute traffic analysis. Many people who are aware of the surveillance for industries and governments want to maintain their privacy protecting their communications. Some of them avoid being profiled or stigmatized. In government and militaries scenarios, it is important to safeguard privacy in communications, because a change in pattern communication can lead to dangerous consequences. Anonymity area is a legitimate for protect privacy; there are applications such as web browsing, e-vote, e-bank, e-commerce and others. Dissidents living under an authoritarian regime, journalists who want to maintain anonymous research, whistle blowers and others use the available tools to perform activities without being identified. Anonymity systems provide mechanisms to enhance user privacy and to protect computer systems.

1.1 Mix networks and mix systems attacks

For high-latency anonymous communication systems, mixed networks are considered as the base. Mixes provide protection against observers hiding the appearance of messages, patterns, length, and links between senders and receivers. In 1981, privacy preserving communications research was initiated by Chaum [1]. He described a model to hide the correspondence between senders and receivers through encrypting messages. Assume the scenario where Alice wants to send a message to Bob using a mix network; the initial process is to prepare the message, and then choose the message

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

transmission path. Every step in the transmission path is a layer built on the message because the public keys of the chosen mixes are used to encrypt the message in the inverse order that they were chosen.

The mixes idea is analogous to send a message with multiple envelopes. The sender sets an envelope with the first mix's address, and so on until it uses the last mix's address on the innermost envelope. Every envelope is implemented through encryption using public keys of the mixes.

The attacks against mix systems are intersection attacks whose aim is to reduce the anonymity by linking senders with the messages they send, receivers with the messages they receive, or linking senders with receivers. Attackers can derive relations of frequency through observation of the network, compromising mixes or keys, delaying or altering messages. They can deduce messages from the most probable destinations through the use of false messages sent to the network, and using this technique to isolate target messages and derive their properties. Traffic analysis belongs to a family of techniques used to infer patterns of information in a communication system [8]. In this family, the statistical disclosure attacks are included.

1.2 Statistical disclosure attacks

Based on Graph Theory, the disclosure attack considers a bipartite graph G = (A U B, E) [9]. The set of edges E represents the relationship between senders and recipients A and B. This attack is very expensive in computational terms because it takes an exponential time taking into account the number of messages to be analyzed trying to identify mutually disjointed sets of recipients. The main bottleneck for the attacker derives to a NP-complete problem. Based on the previous attack, the Statistical Disclosure Attack [10] requires less computational effort by the attacker and gets the same results. The method aims to reveal the most likely set of Alice's friends using statistical properties on the observations and recognize potential recipients. To achieve good results, Alice must follow regular communication patterns in the long term.

To our knowledge, there is not another variant of Statistical Disclosure Attacks that performs with real-world data, each assuming very specific scenarios to succeed. The Two-Sided Statistical Disclosure Attack [13] involved a more realistic scenario in assuming regular user behavior on email systems. This attack develops an abstract anonymity system where users send messages to their contacts, and some of these messages are replies. The goal is to estimate the contacts distribution of Alice, and to deduce the receivers of all its messages. An inconvenient detail for application on real data is the assumption all users have an equal number of friends and send messages with uniform probability.

One of the main characteristics in intersection Attacks is that to perform well it has to be a consistent sending pattern or specific behavior for users participating in an anonymity network. In [12], the sender behavior assumes Alice sends n messages with a probability $Pm(n)$. For each round in which Alice sent a message, the attacker observes the number of messages m_i sent by Alice and calculates the arithmetic mean. The results presented are based on simulations on pool mixes, each mix retains the messages in its pool with the same probability at every round. The attacker

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

has great opportunities to succeed if it takes long-term observations even when it partially observes the network. In [14], there are no assumptions on user's behavior to reveal patterns of communication. However, this work shows that it is not enough to just consider the perspective of one user in the system, because all are correlated. Experimentation does not consider user send messages with different frequencies. Another generalization of the disclosure attack model is [15] where Bayesian techniques are applied. Authors develop a model to represent long-term attacks against anonymity systems.

One of the most used techniques to protect against Statistical Disclosure Attack is sending cover traffic which consists of using fake or dummy messages along with real ones to hide Alice's true sending behaviour. In [19], consider background traffic volumes to estimate the amount of dummy traffic that Alice generates. The attacker goal is to estimate how much of Alice's traffic is false based on observations of the incoming and outgoing traffic. Simulations show that for a specific number of recipients, if the background messages increase, it makes it more difficult to succeed considering that the total recipients and Alice's recipients are unchanged.

In [22], a novel statistical disclosure attack that takes into account possible interactions between users under very general conditions is presented. In this article, this method is reviewed under an important modification that improves its performance.

2 Modeling approach

This work addresses the problem of retrieving information about relationships or communications between users in a network system, where partial information is obtained. The information used is the number of messages sent and received by each user. This information is obtained in rounds that can be determined by equally sized batches of messages, in the context of a threshold mix, or alternatively by equal length intervals of time, in the case that the mix method consists of keeping all of the messages retrieved at each time interval and then relaying them to their receivers, randomly reordered.

The attacker can only retrieve the number of messages sent and received for each user in each round as represented in Figs. 1 and 2. In each round, not all the users must be present. A final adjacency matrix A that represents aggregated information from all the rounds is also built by the attacker.

The main objective of the attacker is to derive, for each pair of users ij, if there has been positive communication or not during the study period. Considering a final true adjacency matrix A' where all messages from the original A matrix for all rounds are

Fig. 1 Round example. The attacker only sees the unshaded information

	Receivers			
Senders	u2	u3	u5	
u1	2	1	0	3
u4	1	0	2	3
u7	1	0	1	2
	4	1	3	**8**

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

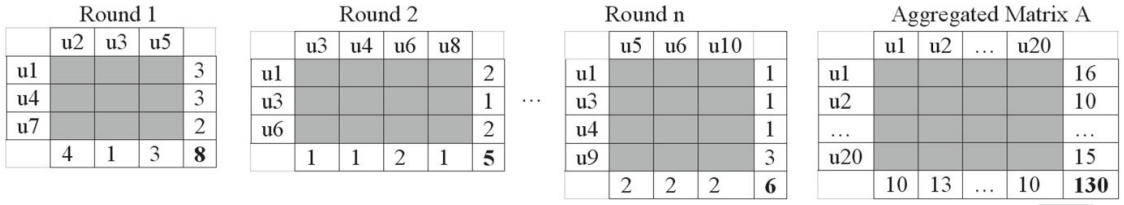


Fig. 2 Information retrieved by the attacker in rounds

summed up, and marking as 1 matrix elements that are strictly positive (there has been communication in at least one round) and allowing that 0 elements that are already zero, the objective of the attacker is to develop a classifier that predicts each cell into 1 (communication) or 0 (not communication). This classifier would lead to an estimate matrix $\widehat{A'}$ and diagnostic measures could be computed based on the true matrix A' and its estimate $\widehat{A'}$. This is an unsupervised problem, since generally the attacker does not know a priori any communication pattern between users.

Let us consider the following general settings for the attack:

- The attacker knows the number of messages sent and received by each user in each round.
- The round can be determined by the system (batches) in a threshold mix context or can be based on regular intervals of time, where the attacker gets the aggregated information about messages sent and received, in the case of a timed mix, where all messages are reordered and sent each period of time.
- No restriction is made from before about the number of friends any user has nor about the distribution of messages sent. Both are considered unknown.
- The attacker controls all users in the system. In our real- data application, we aim at all email users of a domain sent and received within this domain.

In [22], a method to build a classifier is developed. It is based on the random generation of feasible tables for each round, that is, possible tables that match the row and column marginal values (the only information retrieved by the attacker). These tables are used to set an ordering between pairs of users, based on the likelihood of communication, and to classify these pairs as 0 or 1. Some advantages of the algorithm are presented below:

1. It automatically detects many pairs of users that surely did never communicate (since they never coincide in any round).
2. It automatically detects many pairs of users that surely did communicate in some round (when there exist rounds where logical constraints set the cell to be strictly positive).
3. For the rest of the pairs of users, the method establishes an ordering from higher likelihood of communication to lower.
4. A cut point for the ordered list above is used to build a classifier and each pair of users is classified as 1 = did communicate or 0 = did not communicate.

The method was applied to simulated data with good results, and the refinement presented here is applied on real email data. The performance of the method is affected by these features:

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

1. The number of users. As the number of users increases, the complexity of round tables and the number of feasible tables increases, so that it negatively affects the performance of the attack.
2. The percentage of zero elements in the matrix A: These zero elements represent no communication between users.
3. The mean frequency of messages per round for positive elements: This is directly related to the batch size, and when it increases, the performance is negatively affected.
4. The number of rounds: As the number of rounds increases, this improves the performance of the attack, since more information is available.
5. The number of feasible tables generated by round: This affects computing time, and it is necessary to study to what extent it is useful to obtain too many tables. This number can be variable.

A further modification of the method is presented in the next section.

3 Application of the EM algorithm to detect communication patterns

In [22], the obtaining of feasible tables was addressed through a generalized extraction, attempting to obtain feasible tables over all combinatorial regions, giving equal weight to every table. Three features of the algorithm were used to achieve this global representation: uniform generation of table cell values, successive random rearrangement of rows and columns before table generation, and deletion of equal feasible tables once a number of tables were obtained.

In spite of the interesting results obtained in the previous research using this algorithm, feasible tables that match table marginal values usually have different probabilities of being true. A further refinement of the algorithm taking into account this fact is developed in this section.

In a first setting of the following refinement of the algorithm, the number of messages sent per round by user i to the user j is modeled by a Poisson distribution with parameter λ_{ij}. This is a simplification of the underlying non-homogeneous Poisson process (this rate may change over time). This simplification is motivated by the fact that the rounds, defined by the attacker, may be constructed by batches of messages or alternatively, by time periods. Also, approximating a non-homogeneous Poisson process by a homogeneous Poisson process is a frequent decision when information is limited, as is the case in the problem treated here.

Within this modeling approach, the number of messages sent by round by user i will follow a Poisson distribution with parameter $\lambda_i = \sum_{j=1}^{\text{receivers}} \lambda_{ij}$ and the number of messages received by round by user j will follow a Poisson distribution with parameter $\lambda_j = \sum_{i=1}^{\text{senders}} \lambda_{ij}$. Pairs of users that do not communicate will have a degenerated distribution with fixed rate $\lambda_{ij} = 0$.

Each round is an independent realization of a batch of messages sent and received. In each round, the attacker observes the number of messages sent by each user i, x_i^r, and the number of messages received by each user j, y_j^r. It should be noted that an unbiased estimator $\widehat{\lambda}_i$ of the rate λ_i is the average number of messages sent per round by the user i, $\overline{x_i} = \frac{1}{n}\sum_{r=1}^{n} x_i^r$. In the same way, $\overline{y_i} = \frac{1}{n}\sum_{r=1}^{n} y_i^r$ is an unbiased

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

estimator of λ_j. An initial estimator of λ_{ij} can be obtained through the independence assumption in the final aggregated table A obtained aggregating all the round marginals. In this case, using the well-known statistical results in contingency tables under the independence hypothesis, $\left(\sum_{r=1}^{n} x_i^r \sum_{r=1}^{n} y_i^r\right)/N$ is an estimator of the total number of messages sent from user i to j for all the rounds and λ_{ij} can be estimated by $\widehat{\lambda}_{ij} = \left(\sum_{r=1}^{n} x_i^r \sum_{r=1}^{n} y_i^r\right)/Nn$ where N is the total number of messages sent in all rounds. Obviously, the independence hypothesis does not apply, since senders have different preferences over the space of receivers, but it is a good departure point given the limited information available.

To refine the estimation of λ_{ij}, the EM algorithm (Dempster, Laird, Rubin, 1977) will be used. This algorithm allows us to estimate parameters by means of maximum likelihood approach, in situations where it is too difficult to obtain direct solutions from the maximum likelihood optimization equations. Generally, this algorithm is used when a probabilistic model exists where X is the observed data, θ is a vector of parameters, and Z is the latent, non-observed data.

The likelihood function is $L(\theta; X, Z) = p(X, Z \mid \theta)$. Since Z is unknown, likely function of θ is set as $L(\theta, X) = \sum_z P(X, Z \mid \theta)$. This function is not easy to maximize due to the complexity of sum up in Z (frequently multidimensional). The EM algorithm (Expectation–Maximization) allows us to approach the problem in two steps iteratively, after the assignment of an initial value $\theta^{(1)}$. In each step t, the next two operations are made:

1. *Expectation step (E-step)*: The expectation of $L(\theta, X, Z)$ under the distribution of Z conditional to the values of X and $\theta^{(t)}$ is derived: $Q(\theta \mid \theta^{(t)}) = E_{Z|X,\theta^{(t)}}[L(\theta, X, Z)]$.
2. *Maximization step (M-step)*: $Q(\theta \mid \theta^{(t)})$ is maximized in θ, obtaining a new value $\theta^{(t+1)}$ for θ.

This process is realized iteratively until convergence.

In the present problem, X^r is the information observed by the attacker and represents the marginal sums in each round r. Z^r are the unknown values of the cells of the table in the round r. The parameter vector is denoted by λ.

For each round, Z^r cell values are a priori pairwise independent, and rounds are generated independently. Also, Z^r values that do not match the round marginals X^r have 0 probability. Then,

$$P(Z^r \mid X^r, \lambda) \propto \prod_{i,j} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

for all Z^r compatible with X^r marginal values. Proportionality is fixed with respect to the sum over all feasible tables in round r: $\sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!}$, where T^r represents the set of all feasible tables with marginals X^r and z_{ijt}^r is referred to the cell values for each table t from the set T^r.

$P(Z^r \mid X^r, \lambda) = 0$ for all Z^r incompatible with X^r marginal values.

Calling X and Z the information for all rounds:

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

$$P(Z \mid X, \lambda) = \prod_{r=1}^{n} \prod_{i,j} \left(\sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!} \right)^{-1} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

for all Z^r compatible with X^r marginal values.

Since $P(X = x \mid \lambda)$ is the probability of all feasible tables leading to x,

$$P(X \mid \lambda) = \prod_{r=1}^{n} \sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!}.$$

Then, the likelihood is

$$L(\lambda; X, Z) = P(X, Z \mid \lambda) = P(Z \mid X, \lambda) \cdot P(X \mid \lambda) = \prod_{r=1}^{n} \prod_{i,j} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

In this expression, z_{ij}^r (the cell values in each round) are latent values, not observed. The EM algorithm is applied in this context to estimate the λ_{ij} values by maximum likelihood. The initial value for λ_{ij} will be set as the independence hypothesis estimate $\widehat{\lambda}_{ij} = \frac{\overline{x_i y_j}}{n}$.

1. *E-step*: In this step, it is necessary to approach the expectation of $L(\lambda, X, Z)$ under the distribution $P(Z \mid X, \lambda)$. The Monte Carlo method is used to approach $E_{Z|X,\lambda}[L(\lambda, X, Z)]$ by $\frac{1}{m} \sum_{k=1}^{m} L(\lambda, X, Z_k)$, where Z values are obtained by k generations from the conditional distribution $P(Z \mid X, \lambda)$ for each round. Since working with the logarithm, the likelihood leads to the same optimization process the following approximation is applied:

$$\widehat{E}_{Z|X,\lambda}[log(L(\lambda, X, Z))] = \frac{1}{m} \sum_{k=1}^{m} \sum_{r=1}^{n} \sum_{i,j} \log \left(\frac{\lambda_{ij}^{z_{ijk}^r} e^{-\lambda_{ij}}}{(z_{ijk}^r)!} \right).$$

   To obtain samples from $P(Z \mid X, \lambda)$ for each round, the algorithm presented in [22] is applied, but in this case the feasible tables are generated in each cell generation, instead of the uniform distribution, a Poisson distribution with rate $\widehat{\lambda}_{ij}$ truncated by X^r marginal limitations.
2. *M-step*: to maximize the expression $\widehat{E}_{Z|X,\lambda}[L(\lambda, X, Z)]$ with respect to λ_{ij}, the maximization process is developed as is usual in the Poisson distribution parameter estimation. This results in $\widehat{\lambda}_{ij} = \overline{\overline{z}}_{ij}$ where the mean is taken over the sample feasible tables and all the rounds.This estimated value $\widehat{\lambda}_{ij}$ will be used subsequently in the Monte Carlo table generation referred in step 1.

Steps 1 and 2 are repeated iteratively until convergence.

This application of the EM algorithm leads to the final estimates $\widehat{\lambda}_{ij}$. To obtain an estimate of the adjacency matrix $\widehat{A'}$, the ordering of cells is then fixed based on

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

probability of zero for each cell, that is, under the Poisson modeling, $P(z_{ij} = 0) = e^{-\lambda_{ij}}$. A cut point is then selected to apply to ordering list. It can be based on external information, or based on estimation through extracting feasible tables from the A matrix, restricted to sure zero and positive cells already detected by the EM algorithm. The chosen cut point is used to classify cells ij into 0 or 1 obtaining the estimate $\widehat{A}'$ of the true adjacency matrix A'.

The later approach uses the Poisson distribution to model the number of messages sent per round, as is usual in applications. Next, another approach is applied.

Let us model the distribution of the number of messages sent per round by user i to the user j as a discrete tabulated distribution with parameters $(p_{ij0}, p_{ij1}, p_{ij2}, \ldots)$ where p_{ijt} represents the probability the sender i sends t messages to the user j in a round.

To develop a new version of the EM algorithm above, denoting p by the matrix of parameters, it results in

$$P(Z \mid X, p) = \prod_{r=1}^{n} \prod_{i,j} \left(\frac{1}{\sum_{T^r} \prod_{i,j} p_{ijz_{ij}^r}} \right)^{-1} p_{ijz_{ij}^r}$$

for all Z compatible with the marginals X, and the E-Step gives

$$\widehat{E}_{Z|X,p}[log(L(p, X, Z))] = \frac{1}{m} \sum_{k=1}^{m} \sum_{r=1}^{n} \sum_{i,j} \log \left(p_{ijz_{ij}^r} \right).$$

Simple maximization in each p_{ijt} leads to estimate p_{ijt} through the sample proportion the cell ij takes the value t:

$$\widehat{p}_{ijt} = \frac{1}{nm} \sum_{k=1}^{m} \sum_{r=1}^{n} I(z_{ij}^r = t)$$

Since the range of values for each cell is a priori unknown, the estimators $\widehat{p}_{ijk}$ are finally adjusted to sum up to one for each cell ij. For the Monte Carlo approach in the E-Step, each value z_{ij} is generated in each round through the algorithm applied in [22], using in each cell generation the discrete distribution with parameters $(\widehat{p}_{ij0}, \widehat{p}_{ij1}, \widehat{p}_{ij2}, \ldots)$ truncated by marginal limitations.

The initialization of $(\widehat{p}_{ij0}, \widehat{p}_{ij1}, \widehat{p}_{ij2}, \ldots)$ in the EM algorithm in this version is set as in the base algorithm (uniform distribution).

4 Application to email data

Data obtained from the Computation Center of the Universidad Complutense de Madrid are used as a basis to study the performance of the method. Time of sending, sender to receiver (both anonymized) for each message are obtained for 12 months,

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

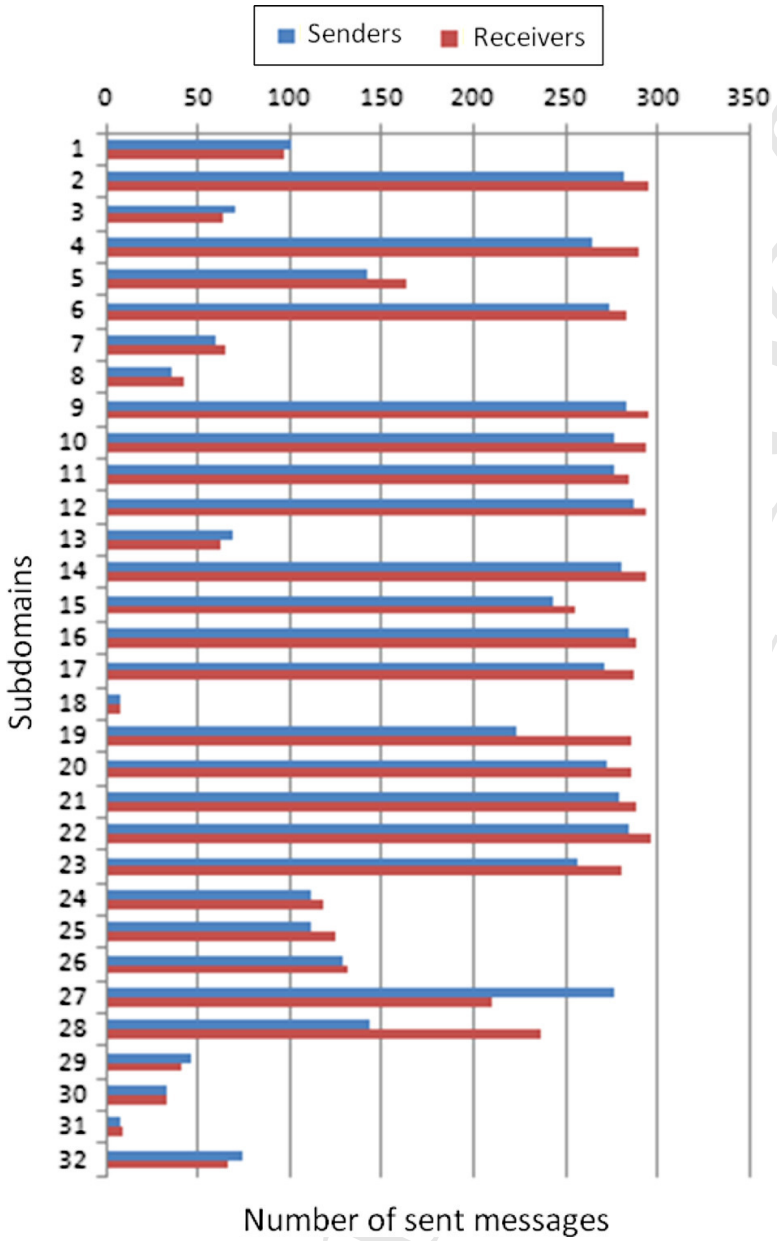


Fig. 3 Number of senders and receivers in different faculty subdomains

in 32 Faculty subdomains. Messages that evidently are sent to lists, institutional messages and messages that come from out of the subdomain or that are sent out of the subdomain are deleted. E-mail data patterns are very specific. This is a very sparse data, and true *A* adjacency matrix for each faculty ranges between 90 and 96 % zero cells (not communication between pairs). User's activity has high variance, ranking from about ten messages to 2500. The number of different receivers for each user is also disperse, from 1 to 40. These numbers affect the detection of communications since active users are more likely to be detected. Figure 3 shows the variability between faculty subdomains in terms of senders and receivers.

The classification algorithm is initially applied to ten faculties to study its performance under the three forms presented:

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

Table 1 Classification rate after five iterations for the three forms of the algorithm and different batch size, for four faculties

Faculty	Batch size	Basic method (uniform)	EM Poisson	EM discrete
1	7	0.997	0.989	0.997
1	15	0.984	0.983	0.986
1	20	0.976	0.977	0.980
2	7	0.985	0.984	0.99
2	15	0.976	0.977	0.981
2	20	0.965	0.966	0.974
3	7	0.975	0.976	0.98
3	15	0.902	0.91	0.92
3	20	0.89	0.88	0.91
4	7	0.991	0.991	0.991
4	15	0.985	0.986	0.988
4	20	0.972	0.974	0.977

(a) The original form in [22], that is, obtaining feasible tables in an uniform setting, trying to explore the space of feasible tables giving equal weight to all the tables.
(b) The application of the EM algorithm under the Poisson model approach.
(c) The application of the EM algorithm under the discrete tabulated distribution model approach.

Given that the method is computationally demanding, the EM algorithm is realized only for five iterations since it has been observed there is no further improvement. It has also been developed with different batch sizes. As can be seen in Table 1, results show that the simple discrete tabulated distribution outperforms the base algorithm and the Poisson modeling approach. Classification rate is the percent of cells ij of the aggregated matrix A that are correctly classified as 0 (not communication) or 1 (communication).

Batch size and complexity of data in terms of percent of zero cells determine the performance of the attack. For the low batch sizes presented in Table 1, classification rate is high, since many trivial solutions are detected besides the use of the algorithm to detect communications. In Fig. 4, the algorithm is applied in the EM-discrete tabulated form to all faculties for different batch sizes over the 12 months horizon. As batch size increases, performance rapidly decreases. For batch sizes over 100, classification rate is often lower than 80 % (not shown in the figure).

For the method presented here, conservative cut points for the classification based on the cells ordering are used. This leads to results with high positive predictive value, and a somewhat lower negative predictive value (many false negatives). That is, when the algorithm classifies a pair ij as "communicating pair", it is very accurate. When the cell is classified as "not communicating pair", it is less accurate. Table 2 presents the True and False positives and negatives, positive predictive value (TP/(TP+FP)) and negative predictive value TN/(TN+FN)) for some faculties and batch 20. The drawback of the algorithm is that it does not detect a high percentage of the true

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

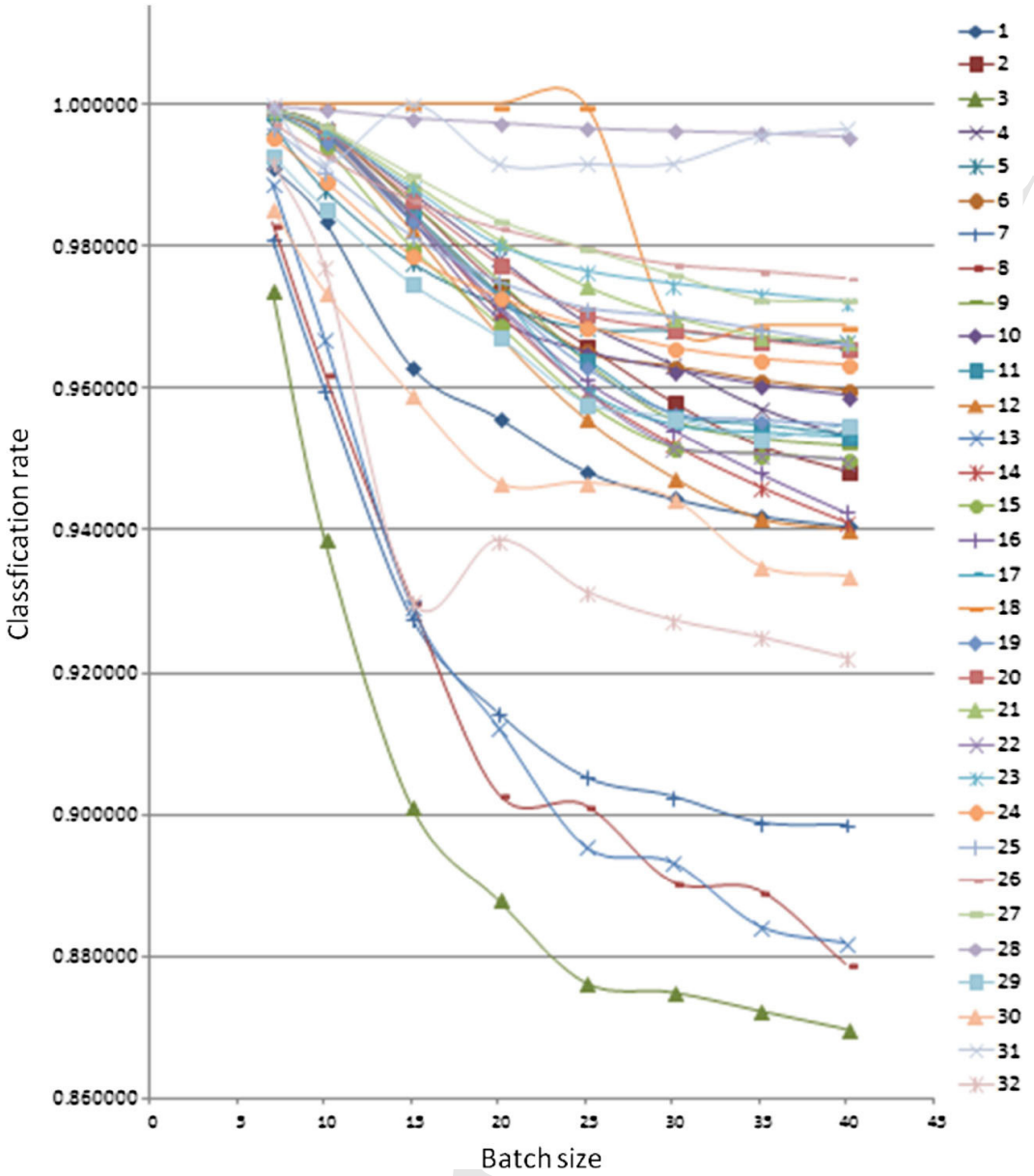


Fig. 4 Classification rate for all the faculty domains, and different batch sizes

Table 2 Rates for different faculties after five iterations of the EM algorithm with discrete distribution, batch 20

Faculty	TP	FP	TN	FN	PPV	NPV	Sensitivity
1	264	0	6818	143	1	0.979	0.648
2	1259	1	88,831	510	0.999	0.994	0.711
3	231	1	4088	304	0.995	0.924	0.43
4	973	0	89,177	451	1	0.994	0.68
5	415	0	28,322	504	1	0.98	0.45

communications (low sensitivity). If the aim of the attacker is simply to capture as many communicating pairs as possible with high reliability, the algorithm presented here is very appropriate.

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

5 Conclusions

A modified version of a disclosure attack through the use of the EM algorithm has been presented. The improvement of the algorithm is significant, increasing the classification rate and other measures such as the Positive Predictive Value. When applying the EM algorithm, it has been shown that the Poisson modeling version does not significantly change the performance of the base algorithm, but its counterpart, the discrete distribution version, improves it significantly. The application of the attack over real data is successful for low batch sizes. The batch size depends on the capacity of the attacker. Often it will be difficult for the attacker to retrieve information in small batches, and then to attain high standards of classification. Using conservative cutpoints for the cell ordering (for example, classifying as 1 only the highest 2 % of cells) is an issue to test in these cases. In general, the positive predictive value of the method is very high: when a pair of users is classified as positive communication, it is very reliable. Complexity of the data increases with the number of users and decreases with the percentage of cells 0. It significantly affects the algorithm. Generation of feasible tables does not need to be heavy. In general, less than 10,000 tables per round are sufficient.EM algorithm seems to converge quickly in 3–4 iterations in our experience. For some faculties, after a few iterations, the results of the EM algorithm deteriorate. Further research is to be done using time structure, and combining the results of cells ij with cells ji, if they exist, to improve the classification, and to study the sensibility and performance of the method over other email data.

Acknowledgments Part of the computations of this work was performed in EOLO, the HPC of Climate Change of the International Campus of Excellence of Moncloa, funded by MECD and MICINN. This work was supported by the "Programa de Financiación de Grupos de Investigación UCM validados de la Universidad Complutense de Madrid - Banco Santander".

References

1. Chaum DL (1981) Untraceable electronic mail, return addresses, and digital pseudonyms. Commun ACM 24:84–88
2. Dingledine R, Mathewson N, Syverson P (2004) Tor: the second generation onion router. In: Proceedings of the 13th USENIX security symposium, pp 303–320. San Diego, 9–13 Aug 2004
3. Pfitzmann A, Pfitzmann B, Waidner M (1991) ISDN-Mixes: untraceable communication with small bandwidth overhead. In: GI/ITG conference: communication in distributed systems, pp 451–463. Springer, Berlin
4. Gulcu C, Tsudik G (1996) Mixing e-mail with Babel. In: Proceedings of the symposium on network and distributed system security, San Diego, 22–23 Feb 1996, pp 2–16
5. Moller U, Cottrell L, Palfrader P, Sassaman L (2015) Mixmaster protocol version 2. Internet draft draft-sassaman-mixmaster-03, internet engineering task force. http://tools.ietf.org/html/draft-sassaman-mixmaster-03. Accessed 9 Feb 2015
6. Danezis G, Dingledine R, Mathewson N (2003) Mixminion: design of a type III anonymous remailer protocol. In: Proceedings of the 2003 symposium on security and privacy, pp 2–5. Oakland, 11–14 May 2003
7. Serjantov A, Sewell P (2003) Passive attack analysis for connection-based anonymity systems. In: Proceedings of European symposium on research in computer security, pp 116–131. Gjovik, 13–15 October 2003

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

8. Raymond JF (2000) Traffic analysis: protocols, attacks, design issues, and open problems. In: Proceedings of the international workshop on designing privacy enhancing technologies: design issues in anonymity and unobservability, pp 10–29. Berkeley, 25–26 July 2000
9. Agrawal D, Kesdogan D (2003) Measuring anonymity: the disclosure attack. IEEE Secur Priv 1:27–34
10. Danezis G (2003) Statistical disclosure attacks: traffic confirmation in open environments. In: Proceedings of security and privacy in the age of uncertainty, IFIP TC11, pp 421–426. Kluwer, Athens
11. Danezis G, Serjantov A (2004) Statistical disclosure or intersection attacks on anonymity systems. In: Proceedings of the 6th international conference on information hiding, pp 293–308. Toronto, 23–25 May 2004
12. Mathewson N, Dingledine R (2004) Practical traffic analysis: extending and resisting statistical disclosure. In: Proceedings of privacy enhancing technologies workshop, pp 17–34. Toronto, 26–28 May 2004
13. Danezis G, Diaz C, Troncoso C (2007) Two-sided statistical disclosure attack. In: Proceedings of the 7th international conference on privacy enhancing technologies, pp 30–44. Ottawa, 20–22 June 2007
14. Troncoso C, Gierlichs B, Preneel B, Verbauwhede I (2008) Perfect matching disclosure attacks. In: Proceedings of the 8th international symposium on privacy enhancing technologies, pp 2–23. Leuven, 23–25 July
15. Danezis G, Troncoso C (2009) Vida: how to use bayesian inference to de-anonymize persistent communications. In: Proceedings of the 9th international symposium on privacy enhancing technologies, pp 56–72. Seattle, 5–7 August 2009
16. Kesdogan D, Pimenidis L (2004) The hitting set attack on anonymity protocols. In: Proceedings of the 6th international conference on information hiding, pp 326–339. Toronto, 23–25 May 2004
17. Bagai R, Lu H, Tang B (2010) On the sender cover traffic countermeasure against an improved statistical disclosure attack. In: Proceedings of the IEEE/IFIP 8th international conference on embedded and ubiquitous computing, pp 555–560. Hong Kong, 11–13 December 2010
18. Perez-Gonzalez F, Troncoso C, Oya S (2014) A least squares approach to the static traffic analysis of high-latency anonymous communication systems. IEEE Trans Inf Forensics Secur 9:1341–1355
19. Oya S, Troncoso C, Pérez-González F (2015) Do dummies pay off? limits of dummy traffic protection in anonymous communications. http://link.springer.com/chapter/10.1007/978-3-319-08506-7_11. Accessed 3 Feb 2015
20. Mallesh N, Wright M (2011) An analysis of the statistical disclosure attack and receiver-bound. Comput Secur 30:597–612
21. Chen Y, Diaconis P, Holmes SP, Liu JS (2005) Sequential Monte Carlo methods for statistical analysis of tables. J Am Stat Assoc 100:109–120
22. Portela J, García Villalba LJ, Silva A, Sandoval AL, Kim T (2015) Extracting association patterns in network communications. Sensors 15:4052–4071

Journal: 11227 Article No.: 1524 TYPESET DISK LE CP Disp.:2015/10/10 Pages: 14 Layout: Small-X

Journal: 11227
Article: 1524

Springer
the language of science

Author Proof

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details required	Author's response
1.	Please check Figs. 4 and 6 has processed as Tables 1 and 2 and Fig. 5 has processed as Fig. 4. Accordingly citations were renumbered.	

Ataque y Estimación de la Tasa de Envíos de Correo Electrónico mediante el Algoritmo EM

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel, Luis Javier García Villalba

***Abstract*—Monitoring a communication channel is an easy task, with appropriate tools anyone can gain information, an attacker can eavesdrop the communication such that the communicators cannot detect the eavesdropping. An attacker is capable of observing the network and deduces users' communication patterns, even when data is incomplete, communication patterns can be used to infer information about a specific subject. The attacker is able to know who communicates whom, what time, frequency, among others. Traffic analysis is a powerful tool because it is difficult to safeguard against. The purpose of this work is to develop an attack and estimate the sending rate of an email system using the EM algorithm.**

***Index Terms*—Anonymity, EM Algorithm, Privacy, Statistical Disclosure Attack.**

I. Introducción

Por definición, la privacidad en términos simples es la protección de nuestros datos ante terceras partes [1]. Se puede garantizar la protección del contenido de un mensaje a través de mecanismos de cifrado. Sin embargo en relación al envío de paquetes de datos se tiene poco control, dado que un adversario al observar la red, puede deducir los patrones de comportamiento de comunicación de los usuarios que la integran, saber quién se comunica con quien, con qué frecuencia, cuándo se comunican, entre otros más detalles; todo ello a partir de un análisis de tráfico y aún cuando tales patrones estén incompletos. En este sentido, el algoritmo EM es un método para encontrar la máxima probabilidad estimada de los parámetros de modelos probabilísticos con datos incompletos, y ha sido utilizado en análisis de flujo de tráficos [2], detección de botnets [3], desarrollo de algoritmos para protección de intimidad en minería de datos [4] y eliminación de spammers [5].

En el presente trabajo nos enfocamos en llevar a cabo un ataque y calcular la estimación de tasa de envíos entre los usuarios de correo electrónico de una universidad a través del algoritmo EM. En la sección II damos una breve introducción describiendo la base de dichos ataques. La sección III describirá el modelo de nuestro ataque. En la sección IV presentamos los resultados de la aplicación de nuestro atraque y finalmente, en la sección V desarrollamos las conclusiones.

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel y Luis Javier García Villalba, Grupo de Análisis, Seguridad y Sistemas (GASS, http://gass.ucm.es), Departamento de Ingeniería del Software e Inteligencia Artificial (DISIA), Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España. E-mail: asilva@fdi.ucm.es, jportela@estad.ucm.es, javiergv@fdi.ucm.es.

II. Ataques Probabilísticos de Revelación de Identidades

A partir del concepto de una red *mix* [6] se han desarrollado múltiples sistemas y contramedidas o ataques. Una red *mix* se conforma por una serie de servidores llamados mixes que se encargan de recolectar mensajes de diversos emisores o usuarios, luego los mensajes son reordenados y finalmente enviados de forma aleatoria a sus respectivos remitentes.

El objetivo de una red mix es prevenir que terceras personas puedan deducir el patrón de comunicaciones de la red. A los usuarios del sistema anónimo se les conoce también como el conjunto anónimo. En la Figura 1 se muestra gráficamente el modelo.

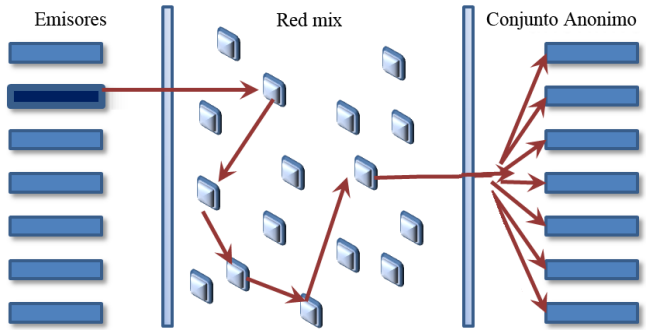


Figura 1. Modelo de red mix.

Suponiendo que Alicia desea enviar un mensaje a Bob a través de un sistema de mixes, se realiza lo siguiente:

i. Preparar la ruta de transmisión del mensaje, tal ruta es la que se utiliza iterativamente antes que el mensaje llegue a su destino.
ii. Cifrar el mensaje a través de las llaves públicas de cada uno de los mixes elegidos como ruta en el orden inverso, como se muestra en la Figura 2.

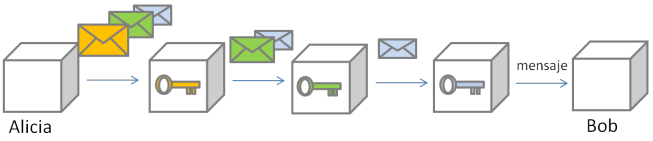


Figura 2. Funcionamiento de una red mix.

El receptor descifra el mensaje a través de la llave privada correspondiente y toma la dirección del siguiente mix.

Los ataques probabilísticos de revelación también son conocidos como ataques de intersección, su base radica en un

ataque a una red mix simple de lotes, cuyo objetivo es obtener información de un emisor particular mediante la vinculación de los emisores con los mensajes que envían, los receptores con los mensajes que reciben, o enlazar a emisores con receptores [7]. Un atacante puede derivar dichas relaciones a través de la observación de la red, retrasando o modificando los mensajes para comprometer los sistemas mix.

El análisis de tráfico pertenece a una familia de técnicas utilizadas para deducir patrones de información en un sistema de comunicación. Se ha demostrado que el cifrado de datos por sí solo no garantiza el anonimato [8].

La base de los ataques de intersección asume que Alicia cuenta con m amigos (o receptores), a quienes envía mensajes con la misma probabilidad a cada uno de ellos. Además también asume enviar un mensaje en cada lote de b mensajes. El modelo usa algoritmos numéricos de conjuntos disjuntos para identificar los receptores de Alicia. En nuestro ataque, no existen tales restricciones.

En el ataque probabilístico propuesto en [8], los receptores se ordenan en términos de probabilidad. Y para que el algoritmo propuesto derive a buenos resultados, Alicia debe mantener patrones de envío consistentes en un largo plazo.

III. Modelo de Ataque

A. *Términos y Definiciones*

Se dice que un emisor o un receptor están "activos" cuando reciben o envían un mensaje en un período de tiempo determinado.

Nuestro ataque no se centra en un usuario en concreto dada la interdependencia de los datos, por lo que nos centramos en obtener la máxima información de todos los usuarios.

La información utilizada es el número de mensajes enviados y recibidos por cada uno de los usuarios. Dicha información se puede establecer por intervalos de tiempo de una longitud determinada o por lotes de mensajes del mismo tamaño, e incluso por ambos para conformar rondas.

Para formar las rondas, el atacante puede construir un conjunto de posibles emisores y/o receptores, tomando en cuenta a aquellos usuarios que están "activos"; tal conjunto de usuarios es el conjunto anónimo.

En la Figura 3 mostramos el ejemplo de la representación gráfica de una ronda y cuya tabla de contigencia es la Tabla I. La Tabla II representa las marginales para la ronda de ejemplo, donde para fines prácticos hemos incluido a pocos usuarios. Se denomina tablas factibles a aquellas tablas donde las marginales coinciden con las de la tabla de contingencia.

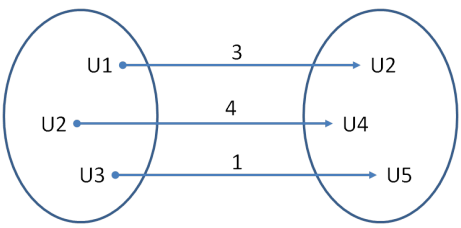


Figura 3. Relación entre emisores y receptores.

Tabla I
Ejemplo de tabla de contingencia

Emisores	Receptores			Total Enviados
	U2	U4	U5	
U1	3	0	0	3
U2	0	4	0	4
U3	0	0	1	1
Total recibidos	3	4	1	8

Tabla II
Ejemplo de tabla de contingencia con marginales

Emisores	Receptores			Total Enviados
	U2	U4	U5	
U1	-	-	-	3
U2	-	-	-	4
U3	-	-	-	1
Total recibidos	3	4	1	8

Para el desarrollo de nuestro ataque hemos considerado lo siguiente:

- El atacante conoce el número de mensajes enviados y recibidos por cada usuario para cada ronda.
- Cada ronda se toma como un evento independiente.
- El atacante controla a todos los usuarios del sistema.
- Las rondas pueden establecerse a través de lotes o por intervalos de tiempo.
- Nuestro modelo es aplicable a un sistema mix simple.
- No existe restricción alguna respecto al número de usuarios del sistema, así como tampoco tenemos restricción del número de amigos o receptores de cada usuario.
- No existe restricción en el número de mensajes enviados.

En una red de comunicación con N usuarios, existen en total N emisores y receptores potenciales. El resultado de las simulaciones es una tabla de dimensión N^2 conformada por receptores y emisores. Cada uno de los elementos de la tabla i, j puede ser 0 ó 1. Donde 1 indica que existe comunicación entre el usuario i y el usuario j, y 0 en caso contrario.

B. *Algoritmo*

Cada usuario i envía mensajes en cada ronda al usuario j según una distribución de Poisson con tasa $\lambda_{i,j}$. El número de mensajes enviados por ronda por el usuario i seguirá una distribución Poisson con tasa $\lambda_i = \sum_{j=1}^{receptores} \lambda_{ij}$ y el número de mensajes recibidos por ronda por el usuario j seguirá una distribución de Poisson $\lambda_j = \sum_{i=1}^{emisores} \lambda_{ij}$. Hay que remarcar que usuarios que no se comunican entre sí tendrán una tasa fija $\lambda_{ij} = 0$, con lo que ese trata de una distribución generada que asigna una probabilidad uno al valor 0.

Cada ronda es una realización independiente de envíos de mensajes. En cada ronda r al atacante observa el número de mensajes enviados por cada usuario i, $\overline{x_i} = \frac{1}{n}\sum_{r=1}^{n} x_i^r$. Del mismo modo $\overline{y_i} = \frac{1}{n}\sum_{r=1}^{n} y_i^r$ es un estimador insesgado de λ_j. Un estimador inicial de e λ_{ij} puede obtenerse asumiendo la hipótesis de independencia entre emisores y receptores. En este caso, y utilizando los resultados estadísticos conocidos

en tratamiento de tablas de contingencia, el número promedio de mensajes enviados por el usuario i al usuario j podrá ser aproximado por $\widehat{\lambda}_{ij} = \frac{\overline{x_i y_j}}{n}$. Obviamente la hipótesis de independencia no se cumple, al tener cada emisor preferencias distintas sobre sus receptores, pero a falta de más información previa sirve como primer punto de partida objetivo.

Para refinar la estimación de los λ_{ij} se utilizará el algoritmo EM [9]. Este algoritmo permite estimar parámetros por máxima verosimilitud en condiciones en las cuales es complicado obtener soluciones directamente de las ecuaciones. En general, se dispone de un modelo probabilístico donde X son datos observados, θ es un vector de parámetros, y Z son datos latentes no observados.

Si se conociera Z, la función de verosimilitud sería $L(\theta; X, Z) = p(X, Z \mid \theta)$. Al no conocerla, la función de verosimilitud de θ se calcula como $L(\theta, X) = \sum_z P(X, Z \mid \theta)$ pero en general no se puede maximizar fácilmente debido a la complejidad de sumar en Z (que a menudo es multidimensional). El algoritmo EM (*Expectation-Maximization*) permite abordar el problema en fases iterativamente, tras asignar inicialmente un valor θ^1. En cada paso t se realizan las siguientes operaciones:

1. *Expectation Step (E-Step)*: Se calcula la esperanza de $L(\theta, X, Z)$ bajo la distribución de Z condicionada a los valores deX y $\theta^{(t} : Q(\theta \mid \theta^{(t)}) = E_{Z|X,\theta^{(t)}}[L(\theta, X, Z)]$.
2. *Maximization Step (M-Step)*: Se maximiza $Q(\theta \mid \theta^{(t)})$ en θ, obteniendo un vlor nuevo $\theta^{(t+1}$ para el parámetro θ.

El proceso se realiza iterativamente hasta su convergencia, monitorizada por diferentes criterios de parada.

En el problema planteado, X^r es la información observada por el atacante y representa los valores marginales de cada ronda r (denotadas anteriormente por x_i^r y y_i^r. Z^r son las realizaciones desconocidas (los valores de las celdas en cada ronda). El vector de parámetros es λ.

Los valores Z^r en una ronda son independientes entre sí, y las rondas se suponen generadas independientemente unas de otras. Además, aquellos valores de Z^r que no suman las marginales X^r tienen probabilidad 0. Se tiene que:

$$P(Z^r = z^r \mid X^r, \lambda) \propto \prod_{i,j} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

para todo Z^r compatible con los valores marginales X^r. La proporcionalidad está referida a la suma sobre todas las tablas factibles de la ronda r: $\sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!}$, donde T^r representa el conjunto de todas las tablas factibles con marginales X^r y z_{ijt}^r se refiere a los valores de las celdas para cada tabla t del conjunto T^r.

Llamando X y Z a la información de todas las rondas:

$$P(Z \mid X, \lambda) = \prod_{r=1}^{n} \prod_{i,j} \left(\sum_{t \in T^r} \prod_{i,j} \frac{\lambda_{ij}^{z_{ijt}^r} e^{-\lambda_{ij}}}{(z_{ijt}^r)!} \right)^{-1} \frac{\lambda_{ij}^{z_{ij}^r} e^{-\lambda_{ij}}}{(z_{ij}^r)!}$$

para todo Z^r compatible con los valores marginales X^r.

En la expresión anterior los z_{ij}^r son valores latentes, no observados. El algoritmo EM se aplica en este contexto para estimar por máxima verosimilitud los valores de los λ_{ij}. El valor inicial de λ_{ij} será el mencionado anteriormente, la estimación básica bajo hipótesis de independencia $\widehat{\lambda}_{ij} = \frac{\overline{x_i y_j}}{n}$.

1. **E-Step**: En este paso es necesario aproximar la esperanza de $L(\theta, X, Z)$ bajo la distribución $P(Z \mid X, \lambda)$. Para obtener una aproximación a esta esperanza se puede ver que es

   $$E_{Z|X,\theta^{(t)}}[L(\theta, X, Z)] = \sum_Z L(\theta, X, Z) P(Z \mid X, \lambda)$$

   Si es posible obtener muestras de Z bajo la distribución condicionada $P(Z \mid X, \lambda)$,, se puede utilizar el método de Monte Carlo para aproximar $E_{Z|X,\theta^{(t)}}[L(\theta, X, Z)]$ por $\frac{1}{m} \sum_{k=1}^{m} L(\theta, X, Z_k)$. Habitualmente se trabaja con el logaritmo de la verosimilitud, En este caso la aproximación quedará finalmente

   $$\widehat{E}_{Z|X,\theta^{(t)}}[L(\theta, X, Z)] = \frac{1}{m} \sum_{k=1}^{m} \sum_{r=1}^{n} \sum_{i,j} \log \left(\frac{\lambda_{ij}^{z_{ijk}^r} e^{-\lambda_{ij}}}{(z_{ijk}^r)!} \right)$$

   Para obtener las muestras de $P(Z \mid X, \lambda)$ en cada ronda se utiliza el algoritmo en [10], generando tablas factibles bajo las restricciones de las marginales pero en este caso alterando la probabilidad bajo la cual se obtiene cada celda, que en este caso se obtendrá a partir de una distribución de Poisson con parámetro λ_{ij} truncada por las restricciones de las marginales.
2. **M-Step**: Resta maximizar la verosimilitud dados los valores de Z muestreados. Desarrollando $\widehat{E}_{Z|X,\theta^{(t)}}[L(\theta, X, Z)]$ y maximizando la expresión en λ_{ij} se obtiene fácilmente que para cada celda, el máximo de λ_{ij} se alcanza en la media muestral obtenida de z_{ij} sobre todas las rondas. Es decir $\widehat{\lambda}_{ij} = \overline{z}_{ij}$.

Los pasos 1. y 2. se realizan iterativamente hasta la convergencia del algoritmo.

El funcionamiento del algoritmo se ve afectado por los siguientes factores:

i. El número de rondas obtenido por el atacante.
ii. El tamaño de las tablas de cada ronda, y el número de usuarios.
iii. El número de tablas factibles generadas en cada ronda.
iv. El número de λ_{ij} ceros en la tabla agregada final ($\lambda_{ij} = 0$ significa que el usuario i nunca envía mensajes a j).

C. Aplicación del algoritmo: Datos simulados

Para efectos pedagógicos establecimos una tabla de lambdas suponiendo un sistema de solamente 3 usuarios. En la Tabla III se muestran las λ verdaderas que hemos utilizado.

En este caso y solo para propósitos de demostración no hemos restringido el hecho de que un usuario pueda enviarse mensajes a sí mismo. Por ejemplo, de acuerdo a la Tabla III se considera que el usuario U1 tiene una tasa de envío de 5 a sí mismo, y de 3 mensajes al usuario U3. En tanto el usuario U2 tiene una tasa de 2, 1 y 2 para los usuarios U1, U2 y U3 respectivamente.

Tabla III
EJEMPLO DE LAMBDAS INICIALES

Emisores	Receptores		
	U1	U2	U3
U1	5	0	3
U2	2	1	2
U3	4	2	0

Considerando las probabilidades de envío y recepción de los usuarios, construimos rondas de diferentes tamaños y que genera también diferente número de tablas factibles. Dado un vector de valores λ estimados, y el vector de λ reales, la distancia entre ellos se denota por,

$$d(\widehat{\lambda}, \lambda) = \sum_{i,j} (\widehat{\lambda}_{i,j} - \lambda_{i,j})^2$$

En la Figura 4 hemos llevado a cabo simulaciones con diferente número de rondas para ver en qué casos se obtienen los mejores resultados. Lo que pudimos observar es que cuando se calcula un mayor número de tablas factibles, el algoritmo arroja mejores resultados pues se cuenta con más información para refinar las soluciones.

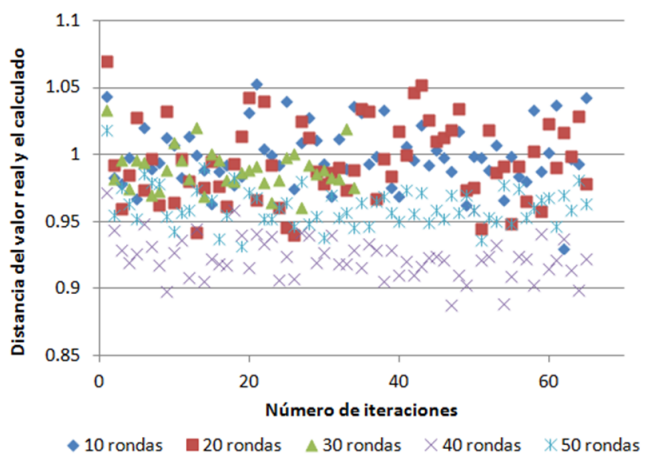


Figura 4. Aplicación del algoritmo en rondas simuladas.

D. Aplicación del Algoritmo: Datos de Email Reales

Para la aplicación del algoritmo hemos utilizado los datos de correo electrónico proporcionados por el Centro de Cálculo de la Universidad Complutense de Madrid del año 2010. Los datos se entregaron luego de ser anonimizados. Se categorizaron los correos por subdominio o facultad, siendo un total de 32. Se consideraron aquellos correos enviados entre los usuarios de la misma facultad. El número de usuarios de cada facultad es diferente. Hemos tomado un horizonte temporal de un mes para llevar a cabo las simulaciones.

Los datos email tienen características peculiares que hacen necesario introducir ciertas modificaciones en el proceso anteriormente expuesto. Concretamente, existen numerosas celdas cero (pares de usuarios que nunca se comunican entre sí) y además, los $\lambda_{i,j}$ que corresponden al número de mensajes enviados por ronda, pueden llegar a ser muy pequeños dependiendo del tamaño de la ronda o de la ventana temporal en la cual el atacante recoge datos, con lo cual es necesario introducir correcciones para evitar problemas de overflow.

Para refinar el resultado, en este caso se utilizan conjuntamente el algoritmo mencionado en [10] con el algoritmo EM presentado en este artículo.

Concretamente, se realizan los siguientes pasos:

1. Un proceso inicial utilizando el algoritmo mencionado en [10] para determinar celdas con ceros fijos (usuarios que nunca se comunican) o extremadamente probables. Estas celdas quedarán catalogadas como celdas con $\lambda_{i,j} = 0$.
2. A continuación se modifica el algoritmo EM utilizado simulando igualmente de distribuciones de Poisson truncadas en cada celda factible pero dejando como ceros los prefijados en el paso anterior. Se realiza un cierto número de iteraciones del algoritmo con esta modificación.
3. Para ilustrar el cambio en la estimación de λ, al no disponer de los λ reales, se presentan los gráficos que muestran cómo evoluciona la distancia a los λ estimados por el promedio del valor de las celdas calculado sobre todas las rondas. Esta información se presenta en la Figura 5, donde se observa que en cada una de las facultades conforme aumenta el número de iteraciones la distancia de lo estimado y lo real decrece.
4. Aparte de la estimación de los λ por celda, la utilización del algoritmo EM conjuntamente con el algoritmo presentado en [10] permite un refinamiento de la clasificación de las celdas en 0 y 1. Se calcula la tasa de error de clasificación basada en estos resultados.

Los factores que afectan a los resultados del algoritmo son los siguientes: El tamaño de las rondas, el número de usuarios, el horizonte temporal y el número de iteraciones del algoritmo.

En la Figura 6 mostramos la relación que existe del número de usuarios por cada subdominio o facultad. Por ejemplo la Facultad 18 y 30 tiene un número de usuarios bajo en comparación con las Facultades 11, 14, 17, 19 y 22.

Podemos notar que a diferencia de la aplicación del ataque con datos simulados, se obtienen mejores resultados al llevarlo a cabo con datos reales. Esto puede estar relacionado al hecho de que hay muchos ceros en los datos reales, es decir que en el período de tiempo de muestreo los usuarios se comunicaron poco.

Cabe señalar que un cero fijo se puede deducir si un usuario i no coincide en ninguna ronda con un usuario j. En la Figura 7 se muestra el número de ceros fijos de cada facultad, como se puede observar es bastante alto, lo que quiere confirma la poca comunicación entre usuarios.

En la Figura 8 mostramos el porcentaje de ceros que existe y que representa la no comunicación entre usuarios.

Finalmente la Figura 9 nos muestra la tasa de aciertos de clasificación obtenida, es decir la capacidad del algoritmo de detectar si existe comunicación entre i, j. Una métrica intuitiva sobre la calidad de un método de clasificación lo constituye la tasa de aciertos. Nuestro algoritmo obtuvo tasas de clasificación superiores a 0,95 para todas las facultades.

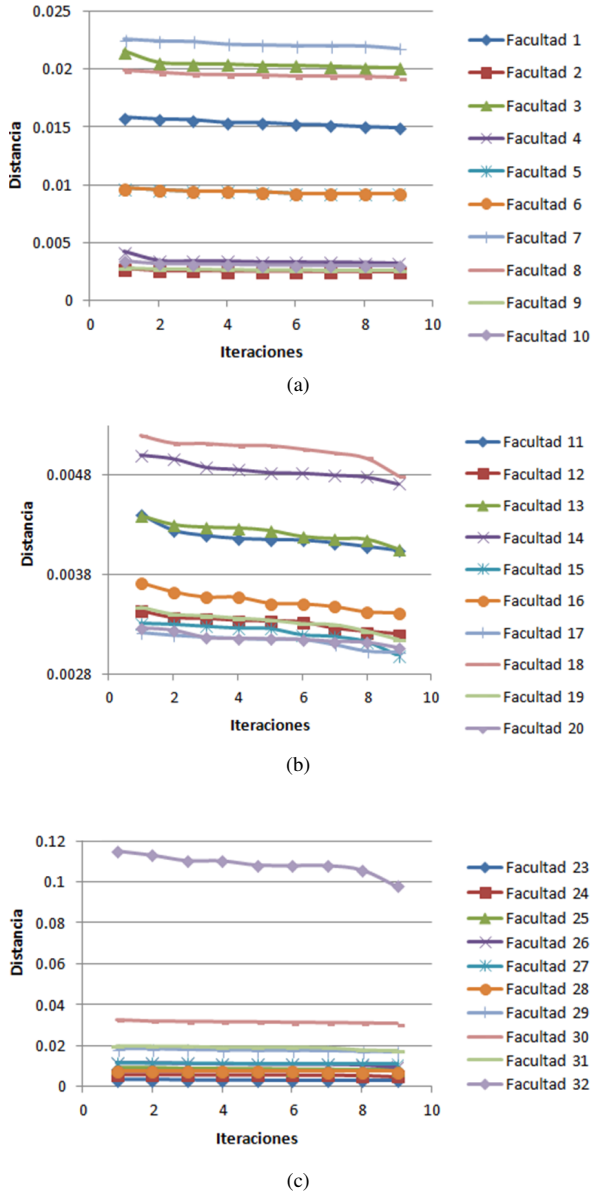


Figura 5. Aplicación del algoritmo en datos de correo electrónico

IV. Conclusiones

Con la aplicación del algoritmo EM pudimos realizar un ataque a un sistema anónimo de correo electrónico y estimar la tasa de envíos. Observamos que nuestro algoritmo arroja mejores estimaciones con datos reales debido a la características particulares de los datos email tales como que existen muchas celdas con cero. Por otro lado, factores como: el número de usuarios, el horizonte temporal, el tamaño de las rondas y el número de iteraciones afectan directamente los resultados. A pesar de estas variables, nuestro algoritmo obtuvo una tasa de clasificación de aciertos superior a 0,95 en todas las facultades. Dentro de los trabajos futuros consideramos llevar a cabo más simulaciones para observar el refinamiento de los resultados, dado que la tasa de clasificación mejora al aumentar el número de iteraciones y con ello se puede deducir mayor información del sistema de comunicación. Así como también habrá que considerar ampliar el muestreo a más meses.

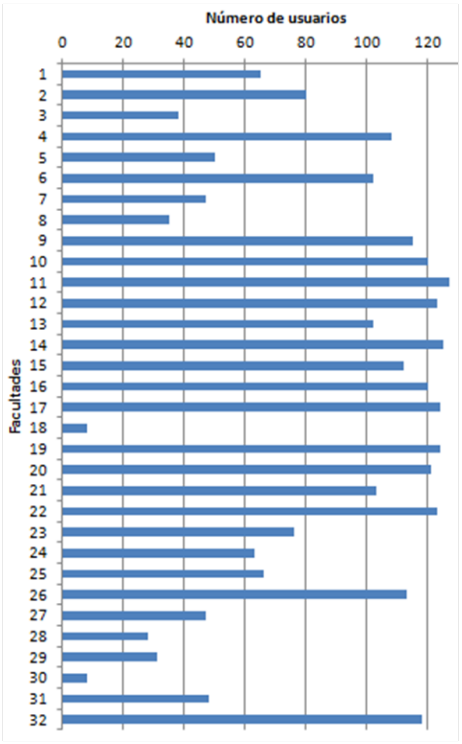


Figura 6. Número de usuarios por facultades.

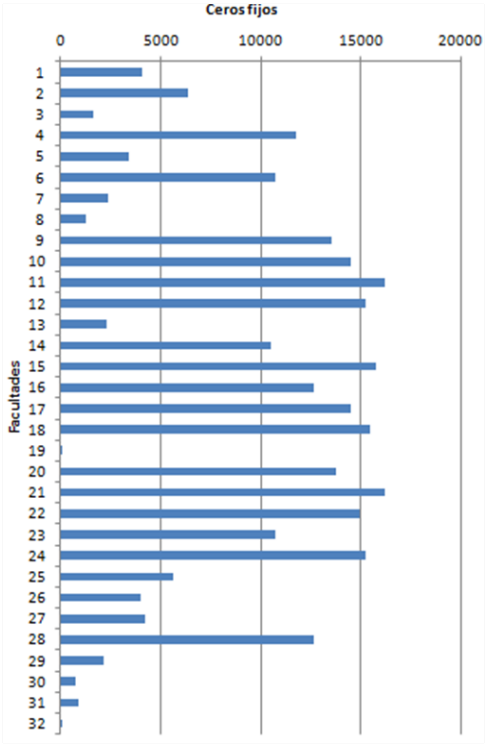


Figura 7. Ceros fijos por facultades.

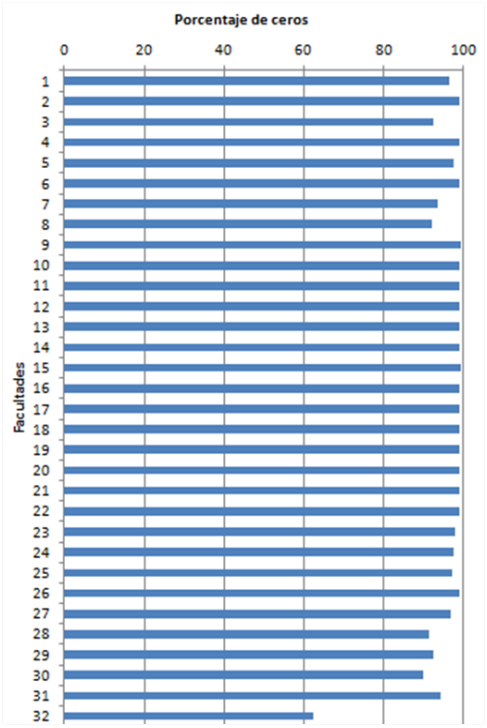


Figura 8. Porcentaje de Ceros por facultad.

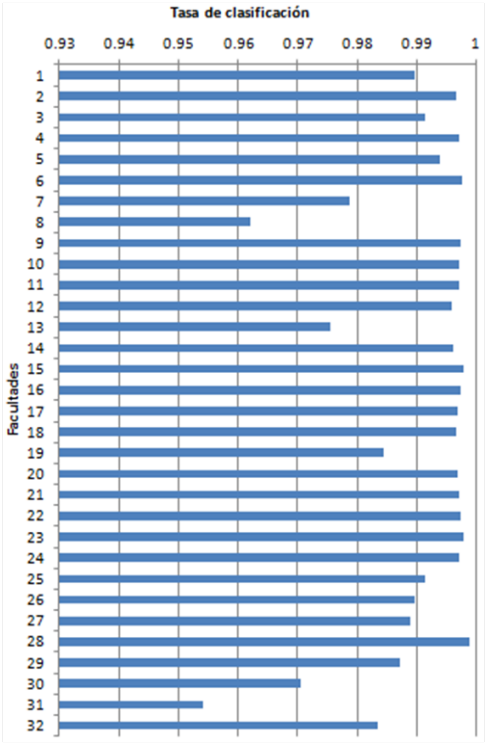


Figura 9. Tasa de clasificación por facultades.

REFERENCIAS

[1] A. F. Westin, "Privacy and Freedom," *Washington and Lee Law Review*, vol. 25, no. 1, p. 166, 1968.

[2] L. L. Chen, A. and J. Cao, "Tracking cardinality distribution in network traffic," in *Proceedings of IEEE 28th Conference on Computer Comunications*, pp. 819–827, April 2009.

[3] S. García, A. Zunino, and M. Campo, "Detecting botnet traffic from a single host," *Handbook of Research on Emerging Developments in Data Privacy*, pp. 426–446, aug 2015.

[4] D. Agrawal and C. C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," in *Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '01, (New York, NY, USA), pp. 247–255, ACM, 2001.

[5] V. C. Raykar and S. Yu, "Eliminating spammers and ranking annotators for crowdsourced labeling tasks," *J. Mach. Learn. Res.*, vol. 13, pp. 491–518, Feb. 2012.

[6] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, pp. 84–88, feb 1981.

[7] D. Agrawal and D. Kesdogan, "Measuring Anonymity: The Disclosure Attack," *IEEE Security & Privacy*, vol. 1, no. 6, pp. 27–34, 2003.

[8] G. Danezis, "Statistical Disclosure Attacks: Traffic Confirmation in Open Environments," in *Proceedings of Security and Privacy in the Age of Uncertainty* (Gritzalis, Vimercati, Samarati, and Katsikas, eds.), (Athens), pp. 421–426, IFIP TC11, Kluwer, May 2003.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[10] J. Portela, L. J. García Villalba, A. G. Silva Trujillo, A. L. Sandoval Orozco, and T.-h. Kim, "Extracting Association Patterns in Network Communications," *Sensors*, vol. 15, no. 2, pp. 4052–4071, 2015.

Alejandra Guadalupe Silva Trujillo received the Computer Science Engineering degree from Universidad Autónoma de San Luis Potosí (Mexico). She works as Security Engineer in the State Government. She is also a Lecturer in the Department of Computer Science of the Faculty of Engineering of the Universidad Autónoma de San Luis Potosí (UALSP). She is currently a Ph.D. student at Complutense Research Group GASS (http://gass.ucm.es). Her main research interests are privacy and anonymity.

Javier Portela García-Miguel received the Mathematics degree from the Universidad Complutense de Madrid (Spain). He holds a Ph.D. in Mathematics from the Universidad Complutense de Madrid. He is currently an Associate Professor in the Department of Statistics and Operational Research of the Faculty of Statistical Studies of the Universidad Complutense de Madrid and a Member Researcher at Complutense Research Group GASS (http://gass.ucm.es). His main research interests are privacy and anonymity.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a M.Sc. in Computer Networks (1996) and a Ph.D. in Computer Science (1999), both from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems, http://gass.ucm.es) which is located in the Faculty of Computer Science and Engineering at the UCM Campus. His professional experience includes projects with Hitachi, IBM, Nokia, Safelayer Secure Communications and H2020. His main research interests are cryptography, coding, information security and its applications.

Extracción de Características de Redes Sociales Anónimas a través de un Ataque Estadístico

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel and Luis Javier García Villalba

***Abstract*— Social network analysis (SNA) has received growing attention on different areas. SNA is based on examine relational data obtained from social systems to identify leaders, roles and communities in order to model profiles or predict a specific behavior in users' network. This information has a huge impact on research areas such as terrorism, financial crimes, analysis of fraud, and sociological studies because SNA helps to understand the dynamics of social networks. The aim of our work is develop a statistical disclosure attack and show the results and information obtained from a social network composed of a university community.**

***Index Terms*— Anonymity, Graph Theory, Privacy, Social Network Analysis, Statistical Disclosure Attack.**

I. INTRODUCCION

En los últimos años se han desarrollado tecnologías que permiten establecer comunidades sociales virtuales tal es el caso de Facebook, Twitter, Instagram, por mencionar algunas. Dichas tecnologías están transformando la manera en que se desarrollan las relaciones sociales y están generando gran impacto en nuestra sociedad.

Toda esta información también ha sido un foco de interés para campos de estudio como el análisis de fraude, el terrorismo, prevención de delitos financieros, donde se involucran estudios sociológicos que se pueden modelar como redes sociales. Existen diversas herramientas y técnicas que permiten entender la naturaleza de la información y encaminarse a una correcta toma de decisiones. Por ejemplo desde el punto de vista de la mercadotecnia, analizar una red social puede revelar quién es el sujeto de mayor influencia para etiquetarlo como un cliente potencial que puede a la vez generar más clientes. El estudio de las redes sociales también se puede abordar desde áreas como la epidemiología, la sociología, la criminalística, el terrorismo, la prevención de fraudes, entre otras. A través del uso de técnicas de Análisis en Redes Sociales se puede responder a preguntas como, ¿quién influye más dentro de una organización?, ¿quién controla el flujo de información?, ¿es posible desarticular la red?

Al evaluar las conexiones entre varios individuos pertenecientes a una red social nos proporciona la posibilidad de identificar los roles que juegan en ella, así como las dinámicas de las relaciones existentes. Por ejemplo, los sociólogos o historiadores desean conocer la interrelación entre los actores políticos o sociales de una determinada red social para identificar agentes de cambio [1]. Otras investigaciones se han enfocado en analizar los envíos de correos electrónicos con el objetivo de identificar comunidades y observar su comportamiento [2] [3] [4]. Para el análisis de blogs en línea, se emplean técnicas de inferencia colectiva que predicen el comportamiento de una entidad a través de sus conexiones. Mediante técnicas de aprendizaje automático o modelos de lenguaje natural se desea identificar al autor de un texto al llevar a cabo un análisis de su forma de escribir y el vocabulario empleado [5] [6].

En el presente trabajo se pretende estimar las características de una red social donde el atacante obtiene información parcial, considerando además las características propias de las redes sociales. Se muestra toda la información que se puede obtener a partir de nuestro ataque a un sistema anónimo de correo electrónico universitario. En la sección II presentamos las características y propiedades de las redes sociales, en la sección III abordamos algunos de los ataques a sistemas de correo electrónico anónimos. En la sección IV describimos el algoritmo para llevar a cabo el ataque de revelación de identidades. En la sección V presentamos los resultados obtenidos y la información que puede derivarse luego de llevar a cabo un análisis de la red social conformada por los usuarios del sistema de correo electrónico de una universidad. Finalmente en la sección VI presentamos las conclusiones y trabajos futuros.

II. REDES SOCIALES Y MÉTRICAS

En esta sección formalizamos la definición y el modelado de redes sociales, así como las métricas más importantes en el análisis de redes sociales.

A. Definición de una red social

Una red social es una estructura social compuesta de individuos, los cuales están conectados por uno o varios tipos de relaciones. Su representación puede hacerse a través de un grafo donde los vértices representan a las personas y las aristas son las relaciones entre ellas. Formalmente una red social se modela como un grafo $G = (V, E)$ donde:

- $V = (v_1, \ldots, v_n)$ es el conjunto de vértices o nodos que representan a entidades o individuos.
- E es el conjunto de relaciones sociales entre ellos (representadas como aristas en el grafo) donde $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$

El análisis estructural de una red social se fundamenta en desarrollar una matriz que representa las relaciones entre los

Alejandra Guadalupe Silva Trujillo, Javier Portela García-Miguel y Luis Javier García Villalba, Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial (DISIA), Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España. E-mail: asilva@fdi.ucm.es, jportela@estad.ucm.es, javiergv@fdi.ucm.es

usuarios y la construcción del grafo correspondiente. Imaginemos que deseamos analizar las relaciones de amistad entre un conjunto de 5 personas y que representamos con 1 la existencia de relación entre ellos y con 0 el caso contrario.

El modelo se puede ver en la Tabla I.

Tabla I
Ejemplo de representación de amistad

	1	**2**	**3**	**4**
1	0	1	0	1
2	1	0	1	1
3	0	1	0	0
4	0	1	0	0

Representamos estas mismas relaciones de amistad por medio de un grafo mostrado en la Figura 1.

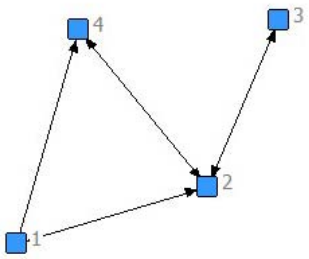


Figura 1. Ejemplo de grafo

Los dos modelos anteriores nos dicen lo mismo respecto a las relaciones de amistad de los participantes. También nos permiten determinar propiedades como la posición de cada amigo en la red, la intensidad en la relación de amistad, entre otros.

B. Propiedades de una red social

En el análisis de redes se utilizan diferentes métricas para clasificación y comparación estructural de las redes y las posiciones que hay en ellas. El análisis puede enfocarse desde el punto de vista de la centralidad, en los clúster fuertemente conectados, en las posiciones que son estructuralmente equivalentes, o en la existencia de posiciones únicas. Otras medidas permiten la comparación de toda la estructura de la red.

A continuación describimos las métricas más importantes y su interpretación.

- **Grado**: El grado de centralidad de un nodo es el número de usuarios o nodos que tienen relación directa con él. En nuestro caso, hacemos uso de grafos dirigidos. Existen dos tipos de grados: 1) grados de entrada, es la suma del número de aristas que terminan en él; 2) grados de salida, es la suma de aristas que se originan en él.
- **Densidad**: Es el porcentaje del número de relaciones existentes y el número de relaciones posibles.
- **Coeficiente de agrupamiento**: Es una métrica que calcula el nivel de interconexión de un nodo con sus vecinos.
- **Centralidad**: Es el número de nodos a los que un nodo está directamente unido.

Existen ciertas características de las redes sociales del mundo real, una de ellas es la llamada mundo pequeño, donde los valores de diámetro son pequeños, en relación al número de nodos [7]. Otra particularidad es que son comúnmente redes libres de escala, muestran un elevado coeficiente de agrupamiento lo que significa que los amigos de amigos son amigos. Por otro lado, al encontrar que el coeficiente de agrupamiento es significativamente mayor a la densidad de la red se puede decir que la red tiene un alto nivel de agrupación. Las redes sociales tienen una distribución de grados que sigue una ley de potencias, donde la mayoría de los nodos tienen pocas conexiones y hay pocos nodos que tienen muchas.

III. ATAQUES PROBABILÍSTICOS DE REVELACIÓN

Se ha demostrado que un atacante puede revelar las identidades de los usuarios de una red mix a través del análisis de tráfico, observando el flujo de los mensajes de entrada y salida. En la literatura existen trabajos en donde un atacante puede obtener información parcial para estudiar una red social anónima, tomando en cuenta las vulnerabilidades a ataques de captura de ruta [8] [9]. Tales ataques utilizan la vulnerabilidad del tráfico de la red para comprometer la identidad de los usuarios que componen la red social.

IV. ALGORITMO

A. Marco Base

El marco base y los supuestos necesarios para llevar a cabo nuestro ataque son:

- Una ronda está formada de grupos de emisores y receptores, el atacante obtiene cuántos mensajes envía y recibe cada usuario. Dicha ronda puede definirse por intervalos regulares de tiempo en los que el atacante observa la red o bien por el sistema (batches).
- Se considera cada una de las rondas como eventos independientes.
- El algoritmo está considerado para un sistema mix simple.
- No existen limitantes respecto al número de receptores o amigos de cada usuario, así como tampoco de la distribución de mensajes enviados. Ambas variables se consideran desconocidas.
- Se asume que el atacante controla a todos los usuarios de la red.

En la Figura 2 representamos una ronda conformada solamente por 6 usuarios para efectos didácticos.

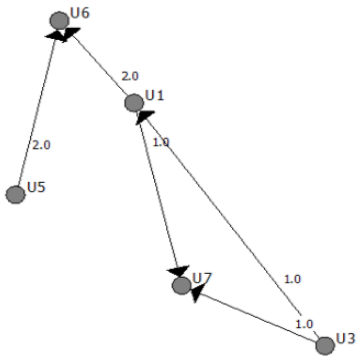


Figura 2. Ejemplo de ronda

La información de una ronda puede definirse a través de una tabla de contingencia como se muestra en la Tabla II, donde los renglones son los emisores y las columnas son los receptores. Cada celda (i, j) contiene el número de mensajes enviados del usuario i al usuario j. Por ejemplo, el usuario 1 envía 2 mensajes al usuario 6 y 1 al usuario 7. Sin embargo, el atacante solo ve las marginales de la tabla que corresponden al número de mensajes enviados o recibidos. El objetivo del atacante es deducir las celdas sombreadas de la Tabla II. Existen múltiples soluciones que corresponden a las marginales; nuestro algoritmo calcula el mayor número posible de tablas-soluciones con el objeto de deducir las relaciones de los usuarios que conforman la red. A cada tabla-solución le llamamos tabla factible.

El algoritmo arroja como resultado información relevante de la existencia de la relación (o no relación) entre cada par de usuarios. Los pasos a seguir son los siguientes:

Tabla II. Ejemplo de una ronda

Emisores/ Receptores	**U1**	**U6**	**U7**	**Total de mensajes enviados**
U1	0	2	1	3
U3	1	0	1	2
U5	0	2	0	2
Total de mensajes recibidos	1	4	2	7

1. El atacante obtiene la información de n rondas a través de la observación de la red tal como se describe en [10].
2. Se generan las tablas factibles para cada ronda a través del algoritmo utilizado en [11].
3. Se construye una tabla agregada A que corresponde a la suma de todas las rondas considerando todos los mensajes enviados y recibidos por cada usuario en el intervalo de tiempo del ataque. Cada celda (i, j) representa el número total de mensajes enviados de i a j.
4. Se calculan los elementos que tienen mayor probabilidad de ser cero, en base al porcentaje de tablas factibles donde el elemento es cero en las rondas en donde está presente. El elemento será cero en la tabla final A' si es cero en todas las rondas.
5. Se lleva a cabo la clasificación de cada celda (i, j) en una matriz estimada $\widehat{A}'$ donde 1 indica que existe relación entre el usuario i y el usuario j, y 0 indica que no existe relación entre ellos.

V. Representación de resultados

Aplicamos nuestro algoritmo a datos proporcionados por el Centro de Cálculo de la Universidad Complutense de Madrid que fueron previamente anonimizados. Tal información la dividimos en 32 subdominios o facultades que componen el sistema de correo electrónico. Para efectos de demostración hemos elegido solamente la Facultad A.

En la Tabla III presentamos los resultados obtenidos luego de aplicar nuestro algoritmo para la Facultad A con datos de 3 meses y en la Tabla IV para 12 meses. Hemos considerado tamaños de lote de 10, 30 y 50. Podemos observar que los valores estimados con lotes de mensajes menores se acercan más a los valores reales de la red.

Tabla III
Resultados de la Facultad A con datos de 3 meses

	Batch	**Nodos**	**Aristas**	**Media Grado**	**Densidad**	**Coeficiente de Agrupamiento**
Estimado	10	85	406	4.776	0.057	0.335
	30	85	406	4.776	0.057	0.335
	50	85	403	4.741	0.056	0.334
Real	-	85	406	4.776	0.057	0.335

Tabla IV
Resultados de la Facultad A con datos de 12 meses

	Batch	**Nodos**	**Aristas**	**Media Grado**	**Densidad**	**Coeficiente de Agrupamiento**
Estimado	10	116	929	8.009	0.070	0.482
	30	116	923	7.957	0.069	0.490
	50	116	924	7.966	0.069	0.479
Real	-	116	929	8.009	0.070	0.482

En la Figura 3 se muestra el grafo estimado y real de la Facultad A con 85 usuarios, con un horizonte temporal de tres meses. En la Figura 4 se muestran los resultados para un período de 12 meses. Las diferencias son poco perceptibles es por ello que hemos colocado los dos grafos superpuestos en donde las aristas de color verde corresponden a las relaciones que nuestro algoritmo no ha detectado tanto en 3 y 12 meses. También podemos notar que ambas redes exhiben características propias de mundo pequeño y escala libre.

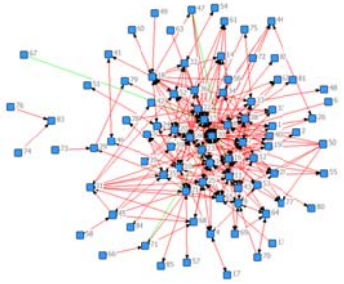


Figura 3. Grafo simulado y real de la Facultad A de un horizonte temporal de 3 meses.

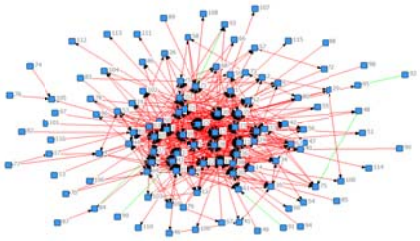


Figura 4. Grafo simulado y real de la Facultad A de un horizonte temporal de 12 meses.

Al hacer un análisis para conocer los nodos más importantes de cada uno de los grafos donde varía el tamaño del lote se obtienen casi los mismos resultados. Esto quiere decir que nuestro algoritmo es capaz de reconocer quiénes son los nodos más influyentes dentro una red a pesar de incrementar el número de nodos.

En la Tabla V presentamos los cinco grados más altos de centralidad calculado para cada red estimada con diferente lote (10, 15, 20 y 30), la última columna corresponde a los de la red real. Por otro lado, en la Tabla VI mostramos los cinco grados de centralidad más bajos.

Tabla V
Los cinco nodos con mayor grado de centralidad de la Facultad A

Batch 10	Batch 30	Batch 50	Real
0.286	0.286	0.286	0.286
0.214	0.214	0.214	0.214
0.190	0.190	0.190	0.190
0.167	0.167	0.167	0.167
0.167	0.167	0.167	0.167

Tabla VI
Los cinco grados de centralidad más bajos de la Facultad A

Batch 10	Batch 15	Batch 20	Real
0.012	0.012	0.012	0.012
0.012	0.012	0.012	0.012
0.012	0.012	0.012	0.012
0.012	0.012	0.012	0.012
0.012	0.012	0	0.012

En la Figura 5 presentamos la comparativa de los grados estimados y reales de la Facultad A para 3 y 12 meses; entre más cercano esté un punto a la diagonal mejor es la estimación. En caso contrario los puntos aparecen muy por encima o debajo de la diagonal.

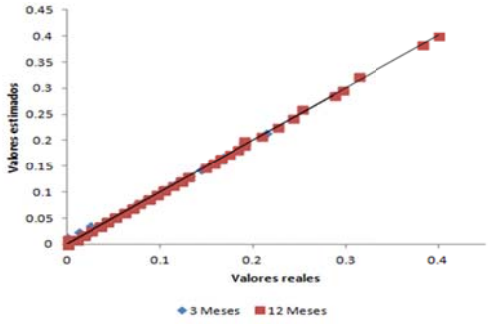


Figura 5. Comparativo de los grados estimados y reales.

IV. CONCLUSIONES

En este documento describimos un ataque probabilístico de revelación de identidades para un sistema de correo electrónico anónimo universitario, asimismo hemos representado tal sistema como una red social. Al llevar a cabo un análisis de redes sociales y obtener las métricas que permiten conocer la centralidad de los usuarios involucrados podemos detectar aquellos elementos más importantes, quiénes controlan el flujo de la información, quiénes son los que pueden influenciar al resto de la red.

De los resultados obtenidos pudimos observar que el ataque es mejor en lotes pequeños, así como también el grafo estimado es bastante similar al real.

Dentro de los trabajos futuros a realizar está el utilizar puntos de corte más amplios para la clasificación de celdas, para observar si se obtienen mejores resultados en lotes de mayor tamaño.

REFERENCIAS

[1] J. Imizcoz, "Introducción actores sociales y redes de relaciones: reflexiones para una historia global", Bilbao: Universidad del País Vasco, 2001, pp. 19-30.

[2] J. Tyler, D. Wilkinson, B. Huberman, "Email as sprectroscopy: automated discovery of community structure within organizations" in *Proceedings of Communities and technologies*, 2003, pp. 81-96.

[3] C. Bekkerman, and A. McCallum, "Extracting social networks and contact information from email and the web", in *Proceedings of CEAS-1*, 2004.

[4] M. Van.Alstyne, and J.Zhang, "EmailNet: autormatically mining social networks from organizational email communications", in *Proceedings of Annual Conference of the North American Association for Computational Social and Organizational Sciences (NAACSOS'03)*, Pittsburg, PA, 2003.

[5] A. Anderson, M. Corney, O.de Vel, and G. Mohay, "Identifying the Authors of Suspect E-mail", *Communications of the ACM*, 2001.

[6] A. McCallum, X. Wang, and A. Corrada-Emmanuel, "Topic and Role Discovery in Social Networks" *Journal of Artificial Intelligence Research 30*, 2007, pp. 249-272.

[7] F. Moradi, T. Olovsson and P. Tsigas, "Towards Modeling Legitimate and Unsolicited Email Traffic Using Social Network Properties", in *Proceedings of the Fifth Workshop on Social Network Systems (SNS '12)*. ACM, New York, 2012.

[8] S. Nagaraja, "Anonymity in the wild: mixes on unstructured networks", In *Proceedings of the 7th Workshop on Privacy Enhancing technologies (PET 2007)*, Ottawa, Canada, June 20-22, 2007.

[9] G, Danezis, C. Diaz, C. Troncoso, and B. Laurie, "Drac: an architecture for anonymous low-volume communications", in *Proceedings of the 10th Privacy Enhancing Technologies Symposium (PETS 2010)*.

[10] G. Danezis, "Statistical Disclosure Attacks: Traffic Confirmation in Open Enviroments", Security and Privacy in the Age of Uncertainty, IFIP Advances in Information and Communication Technology (Sabrina de Capitani di Vimercati, Pierangela Samarati, SokratisKatsikas, Eds.), pp. 421-426, April 2003.

[11] J. Portela García-Miguel, L. J. García Villalba, A. G. Silva Trujillo, A. L. Sandoval Orozco and T.-H. Kim, "Extracting Association Patterns in Network Communications" in Sensors Vol. 15, Issue 2, February 2015.

Alejandra Guadalupe Silva Trujillo received the Computer Science Engineering degree from Universidad Autónoma de San Luis Potosí (Mexico). She works as Security Engineer in the State Government. She is also a Lecturer in the Department of Computer Science of the Faculty of Engineering of the Universidad Autónoma de San Luis Potosí (UALSP). She is currently a Ph.D. student at the Universidad Complutense de Madrid (Spain) and a Research Assistant at Complutense Research Group GASS (http://gass.ucm.es). Her main research interests are privacy and anonymity.

Javier Portela García-Miguel received the Mathematics degree from the Universidad Complutense de Madrid (Spain). He holds a Ph.D. in Mathematics from the Universidad Complutense de Madrid. He is currently an Associate Professor in the Department of Statistics and Operational Research of the Faculty of Statistical Studies of the Universidad Complutense de Madrid and a Member Researcher at Complutense Research Group GASS (http://gass.ucm.es). His professional experience includes projects with Nokia, Safelayer Secure Communications and H2020. His main research interests are privacy and anonymity.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de M\'alaga (Spain) in 1993 and holds a M.Sc. in Computer Networks (1996) and a Ph.D. in Computer Science (1999), both from the Universidad Polit\'ecnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems, http://gass.ucm.es) which is located in the Faculty of Computer Science and Engineering at the UCM Campus. His professional experience includes projects with Hitachi, IBM, Nokia, Safelayer Secure Communications and H2020. His main research interests are cryptography, coding, information security and its applications.