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.

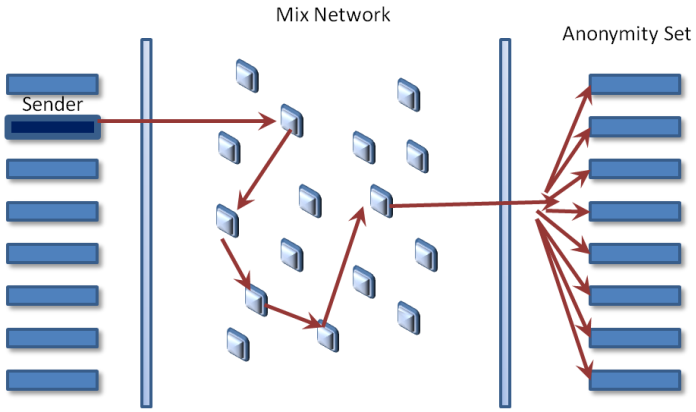

Fig.1. Basic model of a mix network

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*∪*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 × *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* ×*t*matrix: first estimating user profiles when SDA and then de-anonymize mixing round with $P'(s,r) \coloneqq \tilde{P}_{sen(s),SDA}\big(rec(r)\big), s \in S_i, r \in R_i$; 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*∪*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.