トラストアンカーの無いネットワークにおける利用者アカウントの認証法

ヴィシェゴロデツェフ・マラット　　宮本大輔　　若原恭

東京大学
113-8658 東京都文京区弥生2-11-16
kamimachi@cnl.t.u-tokyo.ac.jp, {daisu-mi, wakahara}@nc.u-tokyo.ac.jp

あらまし　通常、サービス顧客のアカウントの確認はサービス事業者が行っている。例えば、銀行などはその顧客に身分証明となる文書を要求し、顧客の名前の真正性を確認している。しかし，サービス事業者が、偽のアカウントが発生するリスクを担保しない場合、このような確認は行われない。例えば、Twitterでは、一部の有名人のユーザに対して本人確認したことを意味する青いチェックマークを表示するが、一般的なユーザに対してはこのような表示はなされず、ユーザの真正性を判断する方法は与えられていない。Facebook等ではユーザ確認はまったく行われていない。この問題を解決し、多数のユーザの真正性を確認するアプローチとして、我々はWeb-of-Trust方式と信頼出来る第三者機関を動的に選択する方式とを組み合わせたハイブリッド公開鍵基盤を開発した。本論文では、本アプローチで用いられている選択方式について、Twitter、OpenSSL、及びBitCoinを組み合わせて実現する例を挙げて説明する。

Providing user account validation in anchor-less networks

Marat Vyshegorodtsev　　Daisuke Miyamoto　　Yasushi Wakahara

The University of Tokyo
2-11-16 Yayoi, Bunkyo-ku, Tokyo, Japan 113-8658
kamimachi@cnl.t.u-tokyo.ac.jp, {daisu-mi, wakahara}@nc.u-tokyo.ac.jp

Abstract – The service provider usually does the account validation itself, e.g. banks request customers their identification documents to ensure the validity of the names provided. However, when a service provider does not bear the risk of fake accounts, it usually does not provide such an entity validation service. For example, Twitter provides account validation ("the blue check mark") only for celebrities, so that the regular users are not given any instrument to decide on the validity of their peers. Other services, such as Facebook, do not provide user validation at all. To cope with this problem, we have developed a new approach to provide massive user validation via hybrid key infrastructure combining web-of-trust and dynamically selected trusted third parties. In this paper we describe the selection method used in the approach and give an example of a mash-up system based on Twitter, OpenSSL and BitCoin

1. Introduction

With the growth of social networks it became possible to build independent web-of-trusts (WoTs) by performing simple graph connectivity analysis. However, the problem of connecting mutually trusted peers on a global scale cannot be solved without some trusted third parties (TTPs). A TTP mediator selection algorithm comes with a trade-off: the fewer mediators are, the validation

process becomes more expensive and harder to operate. The more mediators we add to the system, the less secure our decision is. Different research studies (EigenTrust [1], XRep [2], PowerTrust [3]) are trying to add some kind of reputation systems generally based on evidence observations and weighted voting to tolerate a large number of TTPs. Such approaches pursue the creation of strictly technical imbalance between malicious and benign behavior patterns and do not consider social aspects of the entity validation process. Therefore, to perform an attack it is merely required to follow the feedback algorithm to provide as many positive feedback reports as needed.

The reputation systems backed up by some financial mechanisms (eBay 96 [4]) are stronger against the identity spoofing and Sybil attacks. Each change in the reputation score has a cost, thus malicious behavior causes financial losses, which contradicts the primary goal of the attacker. However, such reputation systems imply mutual evaluation, i.e., the seller evaluates the buyer and vice versa. In the user identity validation process there are two counterparts evaluating the seller: the destination peer and a user of the peer's service, thus mutual evaluation can hardly be implemented.

In this paper we propose a TTP selection algorithm for large scale trust networks that is also guaranteed by the financial instruments. We propose to make the validation process into the service, thus create a P2P market for the user certification. We describe a proof-of-concept system for Twitter, which combines X.509 (OpenSSL) and BitCoin (the reputation scoring basis).

The rest of this paper is organized as follows: in section 2 we discuss the problem of establishing trust in large scale networks, in section 3 we discuss the differences between TTPs' behavior and propose an evaluation method to support the selection process. In section 4 we discuss different types of known attacks on the reputation systems in regard of our reputation score. In section 5 we show a practical implementation example of such a system based on the proposed method.

2. Trust in large scale anchor-less networks

In large social networks such as Twitter or Facebook users are given the following instruments to decide on the trustability of a peer:

1.Total number of connections (followers, friends).

2.Number of shared connections (number of friends also subscribed to the target peer).

Both of these instruments cannot be considered secure for the following reasons:

1. The total number of connections can easily be manipulated by creating bot accounts (Sybil attack)
2. The number of shared connections could be zero if a user does not share any interests with his close friends or he is just new to a social network. Also, the trustability criteria between the user and his friends might be different. For example, an official account of Kremlin (@KremlinRussia) has almost the same number and consistency of followers as a humorous fake account @KermlinRussia. A non-experienced user can easily make a mistake about the origin of the latter account, because most of his friends would read the humorous account together with the real one, thus the decision cannot be solely based on the number of friends following the account.

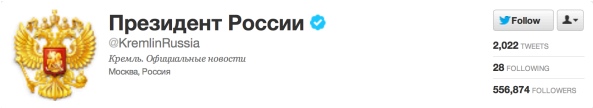

Figure 1 – an official twitter of the President of Russia

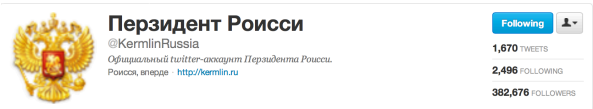

Figure 2 – a humorous account of the "Persident" of Russia

The difference between the accounts shown above is the "blue check mark", which is an official seal of Twitter validation. Such a mark can only be obtained by celebrities or very famous people due to large number of Twitter

users that Twitter would not be able to process in a reasonable amount of time.

If a validation process could be performed in a web-of-trust manner, then the operational cost of validation process can be distributed among additional trusted third parties (i.e. not Twitter or Facebook itself). If the validation process could be distributed completely, then at maximum we would obtain a number of TTPs equal to the number of users of the system. Then, the following problem arises: How would an arbitrary user select a TTP for certificate the verification of an arbitrary peer?

In general there are 4 cases possible:

1. A peer has been validated by a TTP, but neither the peer nor the TTP has any trust connections with the user. The user has to add the TTP of the peer to his web-of-trust or ask the peer to get validated through a TTP the user can trust.

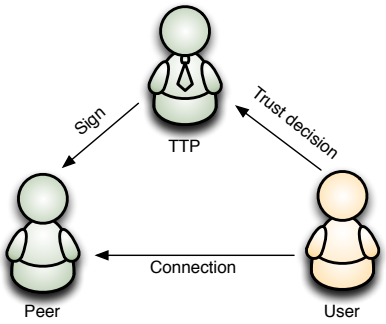

Figure 3 – Peer is verified, and user is deciding on trust

2. A user has trust connections with some TTPs, but none of them have validated the peer. The user may have to select some additional TTPs to include the peer into his web-of-trust or ask the peer to get entity validation from one of the TTPs trusted by the user.

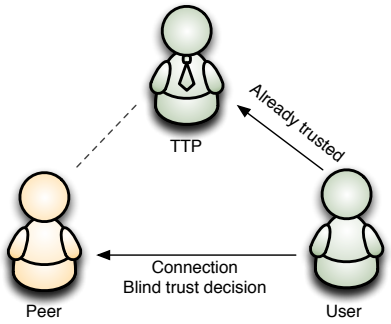

Figure 4 – User and peer do not share user's TTP

3. A user and a peer do not have any connections with any TTPs. Such a case can be considered as the worst (e.g. user used a web search to find a twitter account of some particular person. He clicked on the first result and need to decide whether this is a person he is looking for).

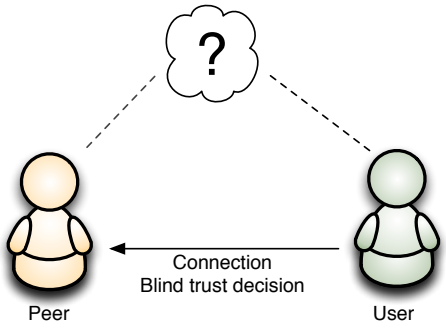

Figure 5 – User and peer do not share any TTPs

4. A user and a peer have at least one common and connected TTP. In this case no future action is required, since a trust connection already exists (e.g. the peer is validated by Twitter and user trusts Twitter).

Below we consider the cases from 1 to 3 and show how user verification could be performed in all of them.

3. Types of TTPs and proposed evaluation method

We divide the TTPs into two categories: community-driven and profit-driven. The first are focusing on entity validation of their friends or neighbors, mainly for free. The second are commercial notaries similar to commercial certificate vendors in SSL. For

profit-driven TTPs we propose an evaluation method that will ease the selection of a TTP, when an untrusted or semi-trusted connection occur.

3.1. Community-driven TTPs

The community-driven peers represent the feature implemented in most of the social networks called “mutual friends”. For example, a famous blogger on Japan (e.g. @japanreports) will provide signatures for other bloggers on Japan, but only for those living in Japan themselves, i.e., not for those posting messages using news from abroad. He can establish a validation procedure of calling the phone starting with a +81 country-code for Japan. For this particular case such a validation would be more than sufficient. Such a TTP may or may not charge the service fee. Such signatures would improve the quality of the community content, but would not guarantee the identity of the signed peers.

These TTPs do not require any kind of reputation score, because their use-case implies that they are well-known. Therefore, the selection method is per-user subjective and applied based on their experience.

3.2. Profit-driven TTPs

The profit-driven TTPs are supposed to provide their services on a global scale with some established security validation procedures. Such TTPs can allow interconnection between two completely unknown peers. For example, in scenario 3 described in the section 4, the peers don’t have any single connection between them. The peer, which is taking the risk of being mistaken or spoofed, e.g., some celebrity that is not very famous yet, pays for a validation procedure to a TTP. The peer attempting the connection (e.g. the fan of the celebrity) can see the signature of that TTP and makes the trust decision based on his opinion on this TTP.

The workflow for the profit-driven TTPs is as follows:

1. A destination peer (celebrity) selects a commercial TTP based on its own experience and users’ demand. It could be some security company or very trustworthy entity, e.g., a city ward office (“yakusho:” in Japan), should they provide such a service.
2. The TTP is paid by the peer using some verifiable way (discussed later).
3. The payment system receives the transaction handling fee. The amount of this fee is added to the reputation score of the TTP.
4. TTP issues a signed certificate for a peer.
5. A user connects to the peer and validates the signature. The user makes trust decision based on the reputation score of the TTP.

The workflow above covers the cases 1–3 described in section 2: when no TTP is present, the destination peer selects a best fit TTP based on his own criteria and the reputation score. When TTP is already chosen, the user can expand his own trust network using the reputation score as a decision parameter.

Our approach is focused on the profit-driven TTPs as the service providers for the peers and users, which do not have any shared trust.

3.3. Proposed evaluation method for the profit-driven TTPs

Our proposed approach is focused on the profit-driven TTPs as the service providers for the peers and users, which do not have any shared trust.

In general, a destination peer can have multiple TTPs verifying its identity, thus multiple signatures associated with the public key of the peer. When a user performing the certificate verification fetches the certificate from the certificate storage server, the scores of all signatures are evaluated. The following evaluation methods are suggested:

1. Calculate the absolute position among all profit-driven TTPs, i.e., the TTPs below a certain threshold can be filtered out. The threshold can be defined absolutely (e.g. top 20 TTPs) or relatively (e.g. top 20% of TTPs).
2. Perform the time series analysis of the score. For example, a user might want to

exclude very new TTPs or the TTPs with the long spans of inactivity. If no profit was generated, it may indicate the loss of trust at that period. The time series of the reputation score should be decomposed by the rates of change to indicate the relative difference between each TTP used in the peer's certificate. The most indicative components is the non-deterministic noise which could indicate irregular fluctuations in the reputation score (e.g., when a TTP was compromised or performed some malicious actions).

The final selection decision is performed on the user's behalf and assisted by the client software, e.g., the web browser. It is possible to automatize the selection process completely by combining the two methods described above: first, pre-filter the TTPs by the absolute criterion, then decompose the score function and compare the rest of TTPs against each other. The best candidate is automatically included into the software key chain.

4. Security considerations

In this section we discuss the attacks possible in the reputation scoring system used in our approach and demonstrate high protecting capability against these attacks.

4.1. Sybil attack

A TTP wishing to improve its reputation score may create fake identities to perform the following actions: demoting other TTPs with negative feedback votes, promoting a malicious TTP with positive feedback votes, ignoring the bad behavior of a malicious TTP in statistical systems.

In the system that we are proposing the reputation score is not based on voting or evidence collection, but on the financial characteristic of a TTP. Simply, if a TTP receives a lot of money, then it pays a lot of handling fees, therefore increases the reputation score. If a malicious TTP had attempted to generate fake identities and inflate its own reputation score, it would require to spend a lot of money on transaction fees, hence make the attack financially unreasonable.

4.2. Temporal reputation score inflation

A malicious TTP may collect some money to inflate its reputation score temporarily to perform an attack. For example, if some user does not have a registered account yet, then an attacker can create a fake user account and a TTP account. Using the money collected the TTP would inflate the reputation score of its account and use the verified peer account for cheating. It might be critical in case of politicians during the elections period or sportsmen during the Olympic games [5].

Such an attack should be mitigated on the user side by using the timing filters on the reputation score, e.g. some statistical regression analysis methods. When a user detects the "inflation", i.e., significant increase in the reputation in a short period of time, it should not trust such a TTP.

4.3. Compromise of the payment system

The core component of the proposed reputation system is the payment system collecting the handling fees, thus forming the reputation score. If the payment system becomes compromised or malicious, then all underlying systems will be compromised as well. This is the reason why the transactions must be verifiable and open-public. Only few payment systems can allow this level of transparency in real time.

In this proposal we suggest using Bitcoin [6, 7, 8], a distributed electronic currency. There are two main reasons for that:

1. Bitcoin transactions are all disclosed and open-public, hence easily verifiable.
2. Transaction handling fees are paid to the nodes on the network, which have succeeded in the calculation of proof-of-work. Since it is nearly impossible to predict which node will succeed in calculation, any malicious manipulations with the transaction handling fee payments are nearly impossible too.

Use of Bitcoin allows us to avoid the control of the reputation scoring system by a single entity (such as in eBay 96) and maintains the financial basis of the reputation scoring.

5. Twitter example

As a real world example of the proposed reputation scoring system we suggest a mash-up of Bitcoin and OpenSSL with a keyserver. At the Twitter part we store the certificate fingerprint in the “Bio” profile field.

A user, who wants to be validated, creates a OpenSSL key pair and generates a X.509 certificate signing request (CSR) with common name equal to the twitter handling name, and all other fields correspondingly (real name, address, etc.). Also, the user creates the Bitcoin transaction and includes its ID number into the certificate too.

Then the user sends this CSR to the chosen TTP and pays the Bitcoin transaction with the amount and transaction handling fee specified by the TTP. The user and the TTP exchange oAuth tokens to verify each other identities, then the TTP performs extended validation.

The received certificate is stored to the keyserver. It may be some common shared trusted server or a private server (specified in the “Homepage” profile field in Twitter). The fingerprint of the certificate is stored in the “Bio” field.

The keyserver pre-calculates the reputation score and keeps all certificates indexed by the TTP ID. When a checking user accesses the keyserver, he might request the pre-calculated reputation score value per TTP, or the whole transaction log with this TTP to calculate the reputation score manually.

Given the reputation score, the user makes the trust decision assisted with the client software.

6. Conclusion

We have developed a new approach which addresses the problem of trust decision-making in anchor-less networks. It may be applied in any type of network where the strict policies can not be effectively applied due to the lack of the administrative resources. It improves the pure web-of-trust by adding the profit-driven TTPs with reputation scoring based on the transaction fee handling.

The security of the proposed reputation scoring system is solely based on economic imbalance between malicious and benign behavior.

The new reputation scoring system has the following features:

1. It does not have a single centralized trust anchor.
2. It applies an economic approach, hence mitigates the impacts of attacks on the reputation scoring system.
3. If implemented using Bitcoin, the system is as scalable as Bitcoin is. The reputation scoring data is easily verifiable.

The confidence level of the proposed system can be considered higher, because the feedbacks are easier to verify and harder to fake than in conventional systems such as EigenTrust, XRep or PowerTrust. The proposed concept is similar to eBay 96, but does not imply any single entity to control the financial subsystem, hence it is more secure and scalable.

Future work

The directions of the future work are as follows:

Automate the selection method completely, so that a user will not need to perform any manual actions. An economic research study on the time series analysis in this market is required.

To decentralize the process completely it is required to study how the keyserver can be removed from the architecture. Though, the most effective way would be achieved if the social network could build keyservers into their service.

It is also required to prove the feasibility of the system by running the live experiments.

References

[1] S. Kamvar, M. Schlosser, and H. Garcia-Molina, “The EigenTrust algorithm for reputation management in P2P networks,” In Proceedings of the 12th International Conference on the World Wide Web (WWW’03), 2003.

[2] R. Zhou and K. Hwang, "PowerTrust: A Robust and Scalable Reputation System for Trusted Peer-to-Peer Computing," IEEE Transactions on Parallel and Distributed Systems, vol. 18, no. 4, pp. 460-473, 2007.

[3] R. Aringhieri et al., “Fuzzy techniques for trust and reputation management in anonymous peer-to-peer systems,” J. Am. Soc. Inf. Sci. Technol. 57, 4 (Feb.), pp. 528–537, 2006.

[4] D. Houser and J. Wooders, “Reputation in auctions: Theory, and evidence from eBay,” J. Econom. Manage. Strat. 15, 2 (June), pp. 353–369, 2006.

[5] Mary Pilon. “Twitter Comment Costs Greek Athlete Spot in Olympics,” NYTimes, 25 July 2012. [Online] Available: http://www.nytimes.com/2012/07/26/sports/olympics/twitter-comment-costs-greek-athlete-spot-in-olympics.html [Accessed 20 August 2012]

[6] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," CRFDL Computing, 2009.

[7] "Namecoin DNS - DotBIT Project," [Online]. Available: http://dot-bit.org.

[8] D. Barok, "Bitcoin: censorship-resistant currency and domain system for the people," 19 July 2011. [Online]. Available: pzwart3.wdka.hro.nl. [Accessed 1 December 2011].

[9] A. Back, "HashCash," 1997. [Online]. Available: http://hashcash.org. [Accessed 21 July 2012].

[10] A. Juels and J. Brainard, "Client Puzzles: A Cryptographic Countermeasure Against Connection Depletion Attacks," in Proceedings of NDSS '99, 1999.

[11] K. Hoffman, D. Zage and C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," ACM Computing Surveys (CSUR), vol. 42, no. 1, pp. 1- 31, 2009.

[12] F. G. Mármol and G. M. Pérez, "Providing trust in wireless sensor networks using a bio-inspired technique," Telecommunication Systems, vol. 46, no. 2, pp. 163-180, February 2011.