Privacy Preserving Decentralized Identity Validation for Geo-social Networks over MANET

Leila Bahri Barbara Carminati Elena Ferrari

Ngoc Hong Tran

firstname.lastname@uninsubria.it
STRICT Social Lab
Insubria University
ITALY

ABSTRACT

Mobile phones, and more specifically smart gadgets, have known a rapid proliferation over the past years in terms of their adoption and usage. Their prices have also known noticeable declines making the ownership of a smart-phone at the ability of all pocket sizes. This has created tremendous potential for the design and creation of services that users can consume through their smart-phones and that would improve their daily lives tasks. In this work, we focus on the potential of using smart-phones in geographically bounded areas, such as shopping malls, museums, conference venues, etc, to establish collaborative ad-hoc networks over MANET. These networks are meant to allow for the provision of P2P exchange of information and help between visitors of such places to improve their visiting experience. We discuss how such a network could be designed and we focus on two main challenges: 1. identity validation over the network to ensure the worthiness of provided information, and 2. privacy preservation both against personal information inference from provided information and over the p2p overlay.

1. INTRODUCTION

Bounded spaces that gather heterogeneous groups of people who might be strangers to each other, but who are all bordered within a defined geographical space, such as a shopping mall, a supermarket, an exposition, a museum, etc, represent interesting potentials for social collaboration. In fact, such spaces typically group different services or goods offered to their visitors and strive to provide them with a nice visiting experience. However, visitors of such spaces might not always have the time to discover all what is offered, might not be informed about all what the space has to offer, or might be targeting specific activities and/or goods and do not know where they are located or how interesting they are.

Within such settings, people can collaborate to help each other to create and to have a better service experience. For instance, if a visitor of a museum does not know which sections would interest her the most, it would be of great help for her to get the opinion of other visitors who are scattered across the different sections of the museum. More importantly, the opinion of those visitors sharing her interests would be of higher value, even compared to the information provided by the museum management in pamphlets or maps. Given the wide adoption of smart-phones nowadays and their rapid proliferation, exploiting these devices to offer such a collaborative service over an ad-hoc network set up within the boundaries of the aimed space is a promising opportunity.

Some research works have already suggested the exploitation of geo-localization in bounded spaces to offer better assistance or better services to users. For instance, location based services by PointInside[1] have been deployed in the American hypermarket chain, Meijer[2]. Inside each of their retail stores, 26 hot spots are used to receive customers' locations in order to provide them with on-time assistance about the location of the goods they want to locate. The system receives user location then processes it and looks for a solution to be sent to the customers. Whereas, Apple recently introduced a set of small sensors, called iBeacon,[3] that can be placed around their retail stores to track and communicate with customers who use iPhones. Then, iBeaconInsider[4] solution has been developed over Apple's iBeacon standard and can serve both iOS and Android mobile devices. Moreover, some developed mobile geofencing applications that send promotions, coupons, or ads to users by SMS right when they enter a specific geographical area. As an example of geofencing application, Locaid[5] solution about a location-aware SMS marketing campaign for Eas-

HOTPOST'15 June 22-25, 2015, Hangzhou, China

DOI: http://dx.doi.org/10.1145/2757513.2757520.

[1]PointInside: http://www.pointinside.com/#
[2]Meijer: http://www.meijer.com/
[3]iBeacon: https://developer.apple.com/ibeacon/
[4]iBeaconInsider: http://www.ibeacon.com/what-is-ibeacon-a-guide-to-beacons/
[5]Locaid: http://www.loc-aid.com/

ton Mall[6]. Moreover, one of the outdoor location based services is *Last.fm Festival*[7] that suggests users a list of music festivals near them when users request and they use GPS technique.

Without neglecting the importance of these works, the approach that they adopt suffers from two limitations. First, users are being tracked by a central entity that collects information about their actions and behavior and that might use it for profiling purposes. This might constitute a serious privacy concern to the users. Second, the information and advice are disseminated by the service or goods providers. This might result in corrupted advice as the information is stained by a profit-based spirit. That is, disseminated information is mostly advertising oriented and the relationship between users and information providers is unidirectional only (as users are passive receiving entities) and is knit to monetary profit.

To overcome these limitations, we suggest in this work a system that allows for intra-collaboration between users to provide each other with the advice and information they are interested at in a need-based approach. That is, a user requests information when she needs it and other potential helpers answers in a collaborative way. Indeed, people within such spaces can, for instance, request on-the-spot information from their geo-fellows and get timely and precise feedback just by using their smart-phone without having to move in search for only a piece of information. This can be allowed by basing on geo-localization of other fellow users in combination with other basic background information they provide about themselves (e.g., a basic profile advertising the interests of each user) to create an ad-hoc network for such social collaboration. To the best of our knowledge, there is no other system yet that provides a platform for ad-hoc collaborative social networking between strangers in bounded spaces that allows them to exchange information and advice and that ensures identity validation in a privacy preserving manner.

To design such a collaborative geo-social network, we suggest exploiting Mobile Ad-hoc NETwork (MANET) [2] as MANET's properties seem to be answering most of our applications requirements, such as, availability, cost saving, self-organized, collaborative, mobiel-phone based, and infrastructure-less architecture. Users can quickly set up a MANET in need, by Wi-fi or Bluetooth connections, without going through any central coordinator often asking for more security and performance requirements. This progress brings users convenience in getting information on time and they do not spend much time waiting or pay for any service cost. Moreover, MANET turns out to be one of the most potential future research trends that is continuously developed as predicted in [17], [24]. Applicability of MANET is extremely large, for instance, plenty of MANET applications have been developed on tactical networks, emergency, as well as education, context aware services, entertainment, or military services [15]. Hence, MANET gets more popular in real life today. In particular, MANET applications become effective when they are deployed in places having a high density of mobile users, such as, hospitals, shopping malls, shopping streets, trade buildings, etc. This goes in line with the scenarios we are targeting in this paper for collaborative geo-social networks in bounded areas using mobile-phones.

Designing a system that allows for intra-collaboration between users in a bounded area over MANET is expected to overcome the limitations discussed above, however it does also introduce new security and privacy issues. The first one derives from a security and trust perspective. In fact, users would need an assisting measure to know to which extent the information they received is valid and its source is trustworthy. Moreover, people when both requesting information or providing it might require to preserve their privacy both against inference of personal information from their exchanged messages and against adversaries in the p2p overlay. In this paper, we focus on these two challenges and we discuss how they could be addressed.

The remaining of the paper is organized as follows. Section II explains the system through an accompanying scenario and discusses the aforementioned challenges, whereas Section III summarizes the paper and discusses future work.

2. IDENTITY VALIDATION AND PRIVACY RELATED CHALLENGES FOR GEO-SOCIAL NETWORKS

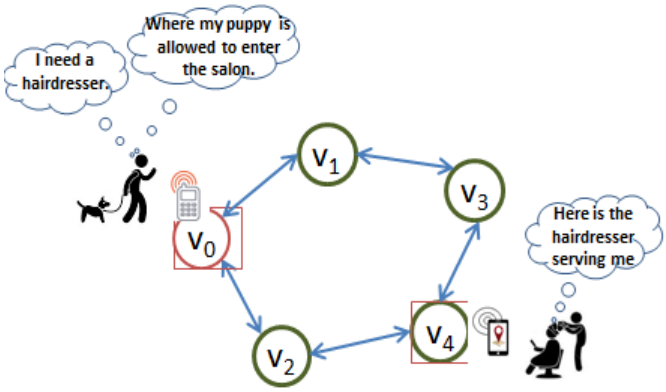

Figure 1: An overview architecture illustrated through an example

For illustration purposes, we carry on our discussions considering the scenario of a shopping mall that gathers shops offering different goods, restaurants, supermarkets, and other services to a heterogeneous group of people who might be strangers to each other but who are all geographically bound to the space of the mall. These people visit the mall for different purposes. For example, some might be interested in shopping for specific goods (e.g., clothes or babies care products), others might be there to benefit from a specific service (e.g., looking for a hairdresser), while others might be there just to spare some free time. Generally, shopping malls provide maps and guiding information to their customers, however people still might find it hard to browse through the mall and find what they are looking for. Furthermore, some customers might be interested at specific goods and might need to compare their prices or their availability across different stores in the mall. Besides, customers might not want to fall in the trap of the advertising campaigns of some specific brands simply because it happened

[6]Easton Mall: http://www.eastontowncenter.com/
[7]Last.fm Festival: http://www.last.fm/festivals

that they got through their stores first. Therefore, the experience of shopping mall customers could be hugely improved given better assistance and easier access to information in terms of both timeliness and relevance to their needs.

To achieve this, we exploit MANET to create an ad-hoc network for visitors of the shopping mall that they can join using their mobile-devices only. This network allows for transferring requests and responses under a data structure, called *token*, between two collaborating users. Tokens move through intermediate nodes, that is nodes making part of the network apart from the requester and the provider, in a voluntary collaborative spirit. To demonstrate this, let us consider Example 1 that is also illustrated on Figure 1.

Example 1. Let us consider that a MANET ad-hoc network is established in a shopping mall by collaborating joined nodes, v_0, v_1, v_2, v_3, and v_4. Assume that person v_0 is visiting the shopping mall with her puppet. v_0 wants to go to a hairdresser and is inquiring for advice on a good one in the mall. Besides, she also needs the salon to be admitting accompanying pets. v_0 formulates her request in a message (i.e., token) that she launches in the network. In the meantime, person v_4 is at a hairdresser's salon. Let us assume that v_4 receives v_0's request via the path (v_0, v_2, v_4) and that she wants to answer it. There are three issues here:

- When v_0 launches her message in the network, she might not want to disclose her identity; however she still needs to get reliable feedback.
- When v_4 answers the request, she also might not want to get identified; however her trustworthiness and the reliability of her feedback should be verifiable by v_0.
- Both the requester v_0 and the responder v_4 might not want to divulge information about their identities to intermediate nodes in the path, that is, to v_2.

2.1 Trustworthiness estimation and identity validation

Identity validation is a major concern in online social networks (OSNs) in general. Several researchers have paid due attention and effort to the issue of validating users' identities [25, 8, 23, 12], detecting fake profiles [28, 16, 14, 5], or enhancing the safety of OSN environments as a whole. In one of our previous works, we have suggested a model for identity validation based on profile information only [1]. The model focused on learning correlations between profile attributes that carry significant information in terms of validating a profile's trustworthiness. More precisely, we have designed our identity validation model over two phases. The first phase consisted at learning the correlations between attributes that would serve from an identity perspective. For example, attributes (*Job*, *Education*) could be said to be correlated as a trustworthy profile is expected to show homogeneous values for these two attributes. As such, we have designed the learning phase exploiting a supervised and feedback based approach by which a group of trusted users provided feedback on the correlations between attribute values they think could help them estimate the trustworthiness of a profile. The second phase consisted at exploiting these learned correlations to assist groups of selected raters to evaluate the trustworthiness of new target profiles. The results of experiments in [1] showed that the suggested method is efficient in assigning reliable estimated trustworthiness values for target profiles.

The approach of basing the estimation of the trustworthiness of a profile, or of an online user, on attribute correlations, like suggested in [1], seem to be a good fit for the identity validation needs in our suggested geo-social-networks over MANET. Indeed, if we assume that our users in the shopping mall, as on Example 1, provide some basic background information about themselves upon joining the network, such as their gender and interests, then this information can be validated based on its conformity to their locations. For instance, referring to Figure 1, if user v_4 specifies that she is a woman and wants to answer the request of v_0 regarding hairdressers, then one possible validation of v_4's reliability is by confirming the homogeneity between her detected location and her claimed profile values (i.e., a woman sitting at a hairdresser). More precisely, location coordinates are retrievable over MANET when a user interacts. These coordinates can be looked up in a public repository, such as Google Maps[8] for example, to retrieve information about the category and the type of the location they refer to. Therefore, if v_4 claims to be a woman being served by a hairdresser, the trustworthiness of this information can be validated if the location she is at refers to a hairdresser for women.

However, such a solution would also incur unwanted privacy concerns. For example, user v_4 might not appreciate being identified as the exact person who is sitting at an identified hair-salon. Such a problem can be avoided by validating information at generalized levels only. For instance, as discussed beforehand, the system can only validate the category of the user's location (i.e., a hairdresser for women) without revealing the exact location or any further information about the place. Nevertheless, this would not completely solve the problem especially given some special situations like when there is only one hairdresser for women in all the shopping mall and only one customer being served there at that time. We further discuss such privacy concerns and how they can be addressed over the following subsection.

2.2 Privacy challenges in identity validation

Privacy preservation and security insurance often come in trade-offs against each other. In national security and human rights terms, security vs. privacy has been and still is an open debate [10]. Privacy advocates call for the right of humans to fully own and to fully preserve their personal information, while security guards assure that the establishment of security to offer safe environments cannot be done without compromising some personal privacy [29]. This dilemma carries on to the realms of the online world as well where it is even more seriously pronounced [29, 18]. It then comes naturally that we face the challenge of privacy against safety in our scenario of geo-social networks as well. In fact, validating identities serves for the insurance of a safer and more reliable communication and advice exchange environment; thus it represents some level of security insurance. On the other hand, doing so seem to require some privacy sacrifices. These can be categorized into two main groups based on how they happen: 1. privacy breaches that happen as a result of personal data inference from provided information, and 2. privacy leakages that happen while personal information is transmitted over the underlying network.

[8] maps.google.com

2.2.1 Identity disclosure and personal information inference

Though protection against personal information inference could not be fully met yet, researchers have paid great efforts to suggest and to design models and solutions for privacy preservation over personal data either under collection, processing, or analysis. One of the major disciplines under this goal is based on syntactic anonymization techniques [9] such as k-anonymity [27], l-diversity [21], t-closeness [20], β-likeness [4], etc. These techniques aim at preserving privacy through insuring the anonymity of every distinct record in the data under hand, majorly by obfuscating it within the crowd (i.e., the other records in the dataset). For example, the principle of k-anonymity relies on making every record in a k-anonymized dataset indistinguishable from at least $k-1$ other records w.r.t some defined identity disclosing attributes [27]. In our scenario, for example, k-anonymity can be applied given that there are k hairdressers for women in the shopping mall, or there are k different customers at the hairdresser's who are connected to the network.

Although such syntactic anonymization based techniques have largely served privacy preservation across different scenarios and application domains, they do not ensure complete guarantees on privacy. Indeed, these techniques have been proved to be vulnerable to some inference attacks especially when attackers have access to some auxiliary information that is not even available in the targeted dataset [9]. Consequently, newer research on privacy suggested the concept of differential privacy that shifts from providing absolute guarantees about disclosure to relative ones [9]. More precisely, differential privacy is based on adding noise to queries on the targeted data to obfuscate the exactness of the real content [9]. While some researchers believe that differential privacy is the answer to the new requirements on data privacy, others still find it immature to completely replace the well established syntactic alternatives [9]. That said, we believe that the two approaches are equally important in the sense that each of them answers or fits different requirements and different scenarios.

For our scenario in hand, the syntactic techniques seem to better fit with our requirements and with the nature of the data we deal with. In fact, this data consists at limited background information about members of the network who are most likely strangers to each other with low probabilities of having access to any auxiliary data. For this, we orient our focus towards syntactic anonymization techniques and we discuss the benefits and the challenges emerging from the specificity of our targeted scenarios.

Syntactic anonymization techniques have first been designed for static collections of data; however, with the emergence of situations and scenarios requiring the anonymization of dynamic and real time data (such as online data-streams, geo-localization information, etc), these techniques have been revised to fit such scenarios as well [3, 13]. More closely related to our scenario, works such as [22] or [19] have already considered anonymization techniques via generalizations for real time geo-localization data. The main challenges with anonymization of such data come from the fact that it changes both quickly and sporadically as users are in continuous movement. However, in our considered scenarios of bounded spaces, the movement space remains limited and the number of people per sub-area in the space is also expected to be high enough to allow for applying such anonymization techniques. Moreover, other criteria might be considered such as the distance between information requester and responder to better protect against identity disclosure.

2.2.2 Over the underlying network

In this paper, we assume to exploit MANET for transferring requests and responses, as *tokens*, between two collaborating users. These tokens contain the content of messages exchanged between communicating end nodes (i.e., the request or the answers text) as well as other data for trustworthiness and reliability estimation of the engaged nodes. Tokens move through intermediate nodes, that is other nodes in the network engaging in the transmission of the exchanged tokens between the requester and the provider, in the sense that a low probability of knowing each other exists among all voluntary participants. Besides, the fact that such an exchange of information can happen depends much on voluntary spirit and credibility of participants. This brings about a high risk to data security and to data integrity, and requires an assurance that intermediate nodes have not to learn the information inside the token, as well as, cannot alter the token's content. For instance, such an assurance can be to at least be able to detect when a token is counterfeit, pop up some information message, and drop it out, so that the bandwidth consumption is more effective.

To deal with these issues, a naive idea is that we can adopt signature algorithms [11] (i.e., RSA, ECDSA, etc.) to guarantee data integrity. Hence, the token can be removed when one node realizes that is a transmuted token. We can also make the intermediate nodes blind while cooperating by applying asymmetric cryptography algorithms (i.e., RSA, ECC, etc.) [26] to the tokens so as only the requester can read the answer encrypted with his/her public key made by the provider. This can secure the token against unauthorized nodes. Such security solutions can help face with the honest-but-curious attack, where adversaries comply properly to the protocol, but also try to infer extra information from the token. However, they seem not adequate to cope with malicious attacks by which adversaries impersonate themselves as other honest users pretending to falsify the data or detour the protocol.

In addition to this, still other problems should be considered, since participants use mobile devices and mobile networks to contact each other. This means that disadvantages of mobile devices (such as, low memory, low energy, low power, etc.) and of mobile networks (such as, low bandwidth, noise, etc.) do not allow whatever of the above mentioned algorithms to be used. Moreover, MANET itself still has its own weak points, that is, nodes can go online and be down very frequently, and they often move, making the network structure change regularly.

Sharing those same issues related to the security of exchanged tokens over mobile networks and using mobile devices, we have proposed, in [7] and in [6], a secure protocol to preserve the confidentiality of the exchanged data and the privacy of mobile communicating users. We have achieved this by adopting homomorphic encryption algorithms adapted to mobile devices and to mobile networks. Yet, that solution is not fully matching the new requirements of this paper, mostly the ones coming from the deployment of MANET and its resulting inconveniences as mentioned above. For this, our previous work should be further revised

so as to be applied more effectively and more appropriately. More precisely, the new settings of our considered scenario of geo-social networks over MANET introduces two main new challenges. The first one is with respect to the new issues and requirements resulting from using MANET as the underlying architecture. Whereas the second one is related to the fact that, in contrary to our work in [7] and in [6], we do not have access to any other additional information about the collaborating users. In fact, the solution in [7] and in [6] relies on computing trust values between collaborating nodes. These trust values have been based on some given social network connections different from the ad-hoc network we are targeting in this paper. Moreover, in our targeted scenario, the requester and the responder do not know each other and might require to keep anonymous to each other during the transaction process; that is, during the exchange of information request and information provision tokens. Therefore, new revisions to the work should be applied based on the new needs and on the new available information.

3. CONCLUSIONS AND FUTURE WORK

Offering network platforms for ad-hoc geo-socializing between people in bounded spaces for information and advice sharing is promising in improving people's experiences and providing them with easier access to information. However, as we have discussed, this represents real challenges with regard to identity trustworthiness for the estimation of advice and information reliability. Ensuring these two can exploit the validation of correlations between some profile attributes and geo-localization information; however, this results in privacy concerns both with regard to identity disclosure and to personal information inference.

In this paper, we have discussed the challenges related to a privacy preserving identity validation over geo-social networks deployed over MANET in bounded spaces for purposes of information and advice sharing. Though there are available techniques for both identity validation and for privacy preservation against personal information inference, either at message receiver side or over the underlying network, the scenario of geo-social networks introduces new requirements that result in new challenges.

We plan to carry on this proposal to address these challenges with respect to the specificity of the targeted goal and scenario; that is, allowing people in bounded spaces to exchange information and advice over safe ad-hoc geo-social networks that provide both identity trustworthiness measures and privacy preservation guarantees.

Acknowledgment

This work is partially supported by the iSocial EU Marie Curie ITN project (FP7-PEOPLE-2012-ITN).

4. REFERENCES

[1] L. Bahri, B. Carminati, and E. Ferrari. Community-based identity validation on online social networks. In *ICDCS'14*, pages 21–30. IEEE, 2014.

[2] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic. *Mobile Ad Hoc Networking*. Wiley-IEEE Press, 2004.

[3] J. Cao, B. Carminati, E. Ferrari, and K. L. Tan. Castle: A delay-constrained scheme for k s-anonymizing data streams. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 1376–1378. IEEE, 2008.

[4] J. Cao and P. Karras. Publishing microdata with a robust privacy guarantee. *Proceedings of the VLDB Endowment*, 5(11):1388–1399, 2012.

[5] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro. Aiding the detection of fake accounts in large scale social online services. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pages 15–15. USENIX Association, 2012.

[6] B. Carminati, E. Ferrari, and H. N. Tran. Smartpay: a lightweight protocol to enforce trust preferences in mobile person-to-person payments. *ASE Science Journal*, 2(3), 2013.

[7] B. Carminati, E. Ferrari, and N. H. Tran. Enforcing trust preferences in mobile person-to-person payments. In *Social Computing (SocialCom), 2013 International Conference on*, pages 429–434. IEEE, 2013.

[8] P. Chairunnanda, N. Pham, and U. Hengartner. Privacy: Gone with the typing! identifying web users by their typing patterns. In *PASSAT'11*. IEEE, 2011.

[9] C. Clifton and T. Tassa. On syntactic anonymity and differential privacy. In *ICDE Workshops*, pages 88–93, 2013.

[10] T. E. Debates. Privacy and security: this house believes security in the modern age cannot be established without some erosion of individual privacy, 2015.

[11] P. Gallagher, D. D. Foreword, and C. F. Director. Fips pub 186-3 federal information processing standards publication digital signature standard (dss), 2009.

[12] O. Goga, H. Lei, S. H. K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira. Exploiting innocuous activity for correlating users across sites. In *WWW'13*. International World Wide Web Conferences Steering Committee, 2013.

[13] K. Guo and Q. Zhang. Fast clustering-based anonymization approaches with time constraints for data streams. *Knowledge-Based Systems*, 46:95–108, 2013.

[14] B.-Z. He, C.-M. Chen, Y.-P. Su, and H.-M. Sun. A defence scheme against identity theft attack based on multiple social networks. *Expert Systems with Applications*, 2014.

[15] J. Hoebeke, I. Moerman, B. Dhoedt, and P. Demeester. An overview of mobile ad hoc networks: Applications and challenges. *Journal-Communications Network*, 3(3):60–66, 2004.

[16] L. Jin, H. Takabi, and J. B. Joshi. Towards active detection of identity clone attacks on online social networks. In *CODASPY'11*. ACM, 2011.

[17] K. I. Lakhtaria. *Technological Advancements and Applications in Mobile Ad-Hoc Networks: Research Trends*. IGI Global, Hershey, PA, USA, 1st edition, 2012.

[18] F. Larry Magid. Online privacy and security is a shared reponsibility, 2013.

[19] H. P. Li, H. Hu, and J. Xu. Nearby friend alert: Location anonymity in mobile geosocial networks.

Pervasive Computing, IEEE, 12(4):62–70, 2013.

[20] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *ICDE*, volume 7, pages 106–115, 2007.

[21] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[22] S. Mascetti, D. Freni, C. Bettini, X. S. Wang, and S. Jajodia. Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies. *The VLDB Journal, The International Journal on Very Large Data Bases*, 20(4):541–566, 2011.

[23] G. Roffo, C. Segalin, A. Vinciarelli, V. Murino, and M. Cristani. Reading between the turns: Statistical modeling for identity recognition and verification in chats. In *AVSS'13*, pages 99–104. IEEE, 2013.

[24] G. R. S. New trends in mobile ad hoc network. *International Journal of Ethics in Engineering and Management Education (IJEEE)*, 1(4), 2014.

[25] M. Sirivianos, K. Kim, J. W. Gan, and X. Yang. Assessing the veracity of identity assertions via osns. In *COMSNETS'12*. IEEE, 2012.

[26] I. S. Specifications. Ieee standard specifications for public-key cryptography. *IEEE Std 1363-2000*, pages 1–228, Aug 2000.

[27] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[28] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. Sybillimit: A near-optimal social network defense against sybil attacks. In *Security and Privacy*. IEEE, 2008.

[29] C. D. Zaruu. Dilemmas of the internet age: privacy vs. security, 2014.