Towards Implicit Contextual Integrity

(Position Paper)

Natalia Criado
School of Computer Science
Liverpool John Moores University
Liverpool, UK
n.criado@ljmu.ac.uk

Jose M. Such
School of Computing and Communications
Lancaster University
Lancaster, UK
j.such@lancaster.ac.uk

ABSTRACT

Many real incidents demonstrate that users of Online Social Networks need mechanisms that help them manage their interactions by increasing the awareness of the different contexts that coexist in Online Social Networks and preventing users from exchanging inappropriate information in those contexts or disseminating sensitive information from some contexts to others. Contextual integrity is a privacy theory that expresses the appropriateness of information sharing based on the contexts in which this information is to be shared. Computational models of Contextual Integrity assume the existence of well-defined contexts, in which individuals enact pre-defined roles and information sharing is governed by an explicit set of norms. However, contexts in Online Social Networks are known to be implicit, unknown a priori and ever changing; users' relationships are constantly evolving; and the norms for information sharing are implicit. This makes current Contextual Integrity models not suitable for Online Social Networks. This position paper highlights the limitations of current research to tackle the problem of exchanging inappropriate information and undesired dissemination of information and outlines the desiderata for a new vision that we call *Implicit* Contextual Integrity.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Security

Keywords

Contextual Integrity, Privacy, Online Social Networks, Norms

1. INTRODUCTION

Online Social Networks (OSNs) have been a source of privacy concerns and issues since their early days [19]. These privacy concerns have increased along the past decade due to many real privacy incidents being echoed in the media and users being more aware of potential privacy issues [11, 35]. Yet there is a lack of effective privacy controls that allow users to satisfactorily manage their privacy in OSNs [34]. In particular, the exchange of inappropriate information and the undesired dissemination of sensitive information across OSNs are very common and represent one of the major concerns for users. These inappropriate exchanges and undesired disseminations have not only caused serious privacy incidents — e.g., users have lost their jobs [28], have been outed and faced threats to sever family ties [17], have ended their marriages [32], etc. — but also facilitated other activities such as social phishing [21], identity theft [3], cyberstalking [25], and cyberbullying [30].

Some voices argue that this is mainly due to the fact that users are no longer able to share information differently for different contexts or spheres of life (friends, work, etc.) in the cyber world, as they would usually do in the physical world [40]. There are many examples in which this is of crucial importance: photos that depict users in embarrassing situations, indecorous comments, events that reveal some political affiliations, etc. In all these examples, the specific context determines whether or not the exchange of information is appropriate — e.g., one may be willing to share her political affiliations with friends but not with workmates.

Contextual integrity [27] is a modern privacy theory that expresses the appropriateness of information sharing based on the contexts in which this information is to be shared. In particular, contexts are defined considering a set of individuals playing particular roles and a number of norms that govern information sharing among them. Contextual integrity is said to be maintained — meaning that there are no privacy breaches — whenever these norms for information sharing are upheld. Norms for information sharing have two main purposes: (i) determine what information is appropriate to mention in a particular context, and (ii) dictate what information can be transmitted from one party to another or others according to the roles enacted by these parties within and across different contexts.

Computational models of contextual integrity have been recently proposed in the related literature [1, 24]. Following contextual integrity theory, they assume the existence of well-defined contexts, in which individuals enact pre-defined roles and information sharing is governed by an explicit set of norms. However, contexts in OSNs are "implicit, ever changing and not a priori-known" [12]. In particular, norms for information sharing are known to be *implicit* in OSNs [29, 39], i.e., they define the behaviour that is consistent with the most common behaviour. Moreover, roles are dynamic and may not be known a priori — i.e., relationships among individuals in OSNs are constantly evolving [5]. All of these reasons make explicit contextual integrity and the computational models based on it not suitable for OSNs.

In this paper, we argue that a new computational paradigm for Contextual Integrity is needed such that it supports implicit norms for information sharing and contexts as well as dynamic and not-a-priori-known roles. We call such an approach *Implicit* Contextual Integrity.

2. PROBLEM STATEMENT

This paper tackles the following two privacy threats in OSNs:

2.1 Inappropriate Information Exchange

Each context has its own appropriateness norms that determine which information can be mentioned inside each context. For example, one may not mention her political views in a work context, but she may do so in a family context [40]. New models of implicit contextual integrity can use the information that users of OSNs exchange with other users in one context (or community) to infer the appropriateness norms of this context. Specifically, the information that is frequently mentioned by the members of a context can be considered as appropriate whereas information that is never or rarely mentioned can be considered as inappropriate. For instance, if most people do not mention their political views at work, it could be inferred this is not an appropriate topic to exchange in a work context.

Besides the appropriateness norms of each context, there are situations in which people decide to exchange information that may be seen as inappropriate. One of the main reasons that explain this is the creation and reciprocation of close relationships [18]. Indeed, there are empirical studies that demonstrate the fact that reciprocated communication is the dominant form of interaction in OSNs [8]. Accordingly, models of implicit contextual integrity can take into account the appropriateness of the information that has been exchanged with each user to determine when inappropriate information is being exchanged to reciprocate a close friend or to create a close relationship.

2.2 Undesired Information Dissemination

Dissemination occurs when information disclosed in one context travels to another context. That is, dissemination is inter-context disclosure while exchange (as stated above) is intra-context disclosure. Obviously, if the information to be disclosed is already known in the contexts were it may be disclosed, then the disclosure of this information in these contexts cannot entail any new privacy risk. However, disseminations may potentially be undesired and hazardous when they entail the disclosure of sensitive information that was previously unknown in a context [33]. Indeed, studies on regrets associated to users' posts on OSNs highlight the fact that revealing secrets of others is one of the main sources of regret [40]. For instance, there was a recent case in which the sexuality of a person was leaked from her friends context to her family context where her sexuality was previously unknown, causing her being outed and facing threats to sever family ties [17].

A first-line defence against undesired disseminations may be avoiding sharing sensitive information in contexts in which there are people that could disseminate the information to other contexts in which this information is previously unknown. Whether these people decide to disseminate the information or not may depend on the relationship they have to others. That is, people usually have confidence relationships with others with whom they decide to share sensitive information expecting them to keep it secret. One can share some of her deepest secrets with her husband but this may not mean her husband would disseminate this information to other contexts. Thus, models of implicit contextual integrity could take into account the knowledge of the information that has been exchanged with each user to determine when sensitive information is being exchanged to reciprocate a trusted friend or to create/maintain trust relationships.

3. LIMITATIONS OF RELATED WORK

In this section we discuss why current approaches are not enough to deal with Inappropriate Information Exchange and Undesired Information Dissemination in OSNs.

3.1 Contextual Integrity Modelling and Reasoning

Previous work on computational models of contextual integrity proposed mechanisms for modelling and reasoning about contextual integrity principles. For example, Barth et al. [1] formalized some aspects of contextual integrity assuming that there is a set of explicitly defined norms that determine what is permitted and forbidden, that the interactions take place in well-known contexts, and that interaction participants play a specific role in each interaction. In a more recent work, Krupa et al. [24] proposed a framework to enforce norms for information sharing in an electronic institution where norms, contexts and roles are explicitly defined. While these approaches seem appropriate for the kind of domains described in [1] and [24], in OSNs there are not well-known contexts, there is not an explicit definition of the roles played by users and the exchange of information is governed by implicit norms for information sharing. Note that these implicit norms for information sharing define the behaviour that is consistent with the most common behaviour. In contrast, explicit norms for information sharing define behaviour that is normative (i.e., moral).

3.2 Access Control Models for OSNs

The suitability of traditional access control models such as role-based access control (RBAC) for OSNs has been recently challenged on the basis that they cannot capture the inherent social nature of OSNs, such as social relationships and distance among users. To address this limitation, there is a new paradigm that precisely emphasises the social aspects of OSNs. Access control models in this new paradigm are commonly referred to as Relationship-based Access Control (ReBAC) [14, 16, 15, 4, 7, 6, 9]. ReBAC models utilise a variety of features or aspects to characterise users' relationships and define access control decisions based on them. While ReBAC models represent a better framework than other traditional access control approaches to develop tools for defining and enforcing access control policies in OSNs, access control on its own is unlikely to be the complete and definitive solution for an appropriate privacy management in OSNs, as users need awareness about access control decisions to fully understand the consequences of their access control policies [22, 26]. For instance, access control models are known to fail to prevent unintended disclosures [2].

3.3 Disclosure Decision-Making Mechanisms

In the related literature, the use of software endowed with disclosure decision-making mechanisms is not new. For example, several authors [37, 23] proposed mechanisms for computing the privacy-benefit trade-off of information disclosures in online interactions. The aim is to only disclose information when this trade-off renders appropriate results, i.e., where the utility of a particular disclosure is worth the privacy risks/consequences involved by performing the disclosure. However, these mechanisms have difficulties to deal with scenarios where the direct benefit of disclosing a piece of information is a priori unknown or difficult to express in economic terms, such as OSNs, in which disclosures are mostly driven by social factors [20]. In a more recent work, Such et al. [36] proposed a mechanism for entailing agents with capabilities to select the personal attributes of their users to be disclosed to other agents during interactions considering the increase on intimacy and privacy loss a disclosure may cause. However, this mechanism does not consider that the appropriateness of disclosures may vary from context to context, nor does it consider information disseminations.

3.4 Norm Learning

Norm learning [13] is the process of learning how to behave in a specific situation. In case of OSNs, norms for information sharing are implicit (i.e., there is not an explicit definition of what is sensitive or inappropriate), and supervised machine learning algorithms cannot be used to infer norms for information sharing. In the existing literature, *social learning* [10] of norms is defined as the process of inferring implicit social norms concurrently over repeated interactions with members of the social network. In most of the proposals on social learning, norms are inferred by analysing the outcomes of interactions and normative decisions in terms of utility [31]. As previously mentioned, in OSNs the benefit of exchanging information may be difficult to be determined in economic terms. In other proposals, norms are inferred by analysing explicit normative signals such as punishments, sanctions and rewards [38]. These approaches cannot be used in OSNs since implicit norms for information sharing are product of informal social control that is rarely stated explicitly (e.g., sanctions) to unfriendly individuals. Other approaches [13] use imitation as a mechanism for learning social norms. In these proposals, the norms are inferred from the public behaviour exhibited by the majority of the members of the social network (or the majority of the members within an observation radius). A main drawback of imitation approaches is that all members are equally considered; i.e., they do not consider the existence of different social contexts with different social norms and the fact that users engage in relationships of different nature and strength. These unsupervised machine learning approaches are unsuitable to be applied to ONS.

4. IMPLICIT CONTEXTUAL INTEGRITY

We propose that a new computational model of *implicit* Contextual Integrity for OSNs should be built. To be applicable in mainstream OSN infrastructures, this model should only utilise the information that is currently available to users of OSNs and their applications —e.g., the tweets posted by users the following relationships, etc.

Our vision is to include such kind of model in what we would call *Information Assistant Agents* (IAAs), which are agents that act as proxies to access the OSN. IAAs should be capable of learning contexts and their associated norms for information sharing even if these are implicit or unknown a priori with the aim of helping users to avoid inappropriate information exchanges and undesired information disseminations. In particular, each IAA monitors the information exchanges of its user and based on this it infers: (i) the different contexts in which information sharing is to happen; (ii) the relationships between its user and the individuals in each context; and (iii) the norms for information sharing of each context. If IAAs detect a potential violation of the norms for information sharing, they should alert their users, who should have the last word on whether sharing the information or not.

REFERENCES

[1] A. Barth, A. Datta, J. C. Mitchell, and H. Nissenbaum. Privacy and contextual integrity: Framework and applications. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 184 – 198, 2006.

[2] M. S. Bernstein, E. Bakshy, M. Burke, and B. Karrer. Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 21–30, 2013.

[3] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the International Conference on World Wide Web*, pages 551–560, 2009.

[4] G. Bruns, P. W. Fong, I. Siahaan, and M. Huth. Relationship-based access control: Its expression and enforcement through hybrid logic. In *Proceedings of the ACM Conference on Data and Application Security and Privacy*, pages 117–124. ACM, 2012.

[5] M. Burke and R. E. Kraut. Growing closer on facebook: changes in tie strength through social network site use. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 4187–4196, 2014.

[6] B. Carminati, E. Ferrari, and A. Perego. Rule-based access control for social networks. In R. Meersman, Z. Tari, and P. Herrero, editors, *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, volume 4278 of *Lecture Notes in Computer Science*, pages 1734–1744. Springer Berlin Heidelberg, 2006.

[7] B. Carminati, E. Ferrari, and A. Perego. Enforcing access control in web-based social networks. *ACM Transactions on Information and System Security*, 13(1):6:1–6:38, Nov. 2009.

[8] J. Cheng, D. M. Romero, B. Meeder, and J. M. Kleinberg. Predicting reciprocity in social networks. In *Privacy, Security, Risk and Trust, IEEE International Conference on Social Computing*, pages 49–56, 2011.

[9] Y. Cheng, J. Park, and R. Sandhu. A user-to-user relationship-based access control model for online social networks. In N. Cuppens-Boulahia, F. Cuppens, and J. Garcia-Alfaro, editors, *Data and Applications Security and Privacy XXVI*, volume 7371 of *Lecture Notes in Computer Science*, pages 8–24. Springer Berlin Heidelberg, 2012.

[10] N. Criado, E. Argente, and V. J. Botti. Open issues for normative multi-agent systems. *AI Communications*, 24(3):233–264, 2011.

[11] B. Danah and E. Hargittai. Facebook privacy settings: Who cares? *First Monday*, 15(8), 2010.

[12] G. Danezis. Inferring privacy policies for social networking services. In *Proceedings of the ACM workshop on Security and artificial intelligence*, pages 5–10, 2009.

[13] J. M. Epstein. Learning to be thoughtless: Social norms and individual computation. *Computational Economics*, 18(1):9–24, 2001.

[14] P. Fong, M. Anwar, and Z. Zhao. A privacy preservation model for facebook-style social network systems. In M. Backes and P. Ning, editors, *Computer Security âĂŞ ESORICS 2009*, volume 5789 of *Lecture Notes in Computer Science*, pages 303–320. Springer Berlin Heidelberg, 2009.

[15] P. W. Fong. Preventing sybil attacks by privilege attenuation: A design principle for social network systems. In *IEEE Symposium on Security and Privacy*, pages 263–278, 2011.

[16] P. W. Fong. Relationship-based access control: Protection model and policy language. In *Proceedings of the ACM Conference on Data and Application Security and Privacy*, pages 191–202, 2011.

[17] G. A. Fowler. When the most personal secrets get outed on facebook. http://online.wsj.com/articles/SB10000872396390444165804578008740578200224, Accessed: Nov, 2014.

[18] K. Greene, V. J. Derlega, and A. Mathews. Self-disclosure in personal relationships. *The Cambridge handbook of personal relationships*, pages 409–427, 2006.

[19] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.

[20] D. J. Houghton and A. N. Joinson. Privacy, social network sites, and social relations. *Journal of Technology in Human Services*, 28(1-2):74–94, 2010.

[21] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[22] L. Kagal and H. Abelson. Access control is an inadequate framework for privacy protection. In *W3C Privacy Workshop*, pages 1–6, 2010.

[23] A. Krause and E. Horvitz. A utility-theoretic approach to privacy and personalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 8, pages 1181–1188, 2008.

[24] Y. Krupa and L. Vercouter. Handling privacy as contextuxal integrity in decentralized virtual communities: The privacias framework. *Web Intelligence and Agent Systems*, 10(1):105–116, 2012.

[25] A. Lyndon, J. Bonds-Raacke, and A. D. Cratty. College students' facebook stalking of ex-partners. *Cyberpsychology, Behavior, and Social Networking*, 14(12):711–716, 2011.

[26] M. Mondal, P. Druschel, K. P. Gummadi, and A. Mislove. Beyond Access Control: Managing Online Privacy via Exposure. In *Proceedings of the Workshop on Useable Security*, pages 1–6, 2014.

[27] H. Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79(1):119–158, 2004.

[28] G. H. Pike. Fired over facebook. *Information Today*, 28(4):26–26, 2011.

[29] K. Raynes-Goldie and D. Fono. Hyperfriends and beyond: Friendship and social norms on livejournal. *Internet Research Annual*, 4:8, 2006.

[30] M. C. Ruedy. Repercussions of a myspace teen suicide: Should anti-cyberbullying laws be created. *North Carolina Journal of Law & Technology*, 9:323–346, 2007.

[31] S. Sen and S. Airiau. Emergence of norms through social learning. In *Proceedings of the International Joint Conference on Artifical Intelligence*, pages 1507–1512, 2007.

[32] J. Stevens. The facebook divorces: Social network site is cited in 'a third of splits'. http://www.dailymail.co.uk/femail/article-2080398/Facebook-cited-THIRD-divorces.html, Accessed: Nov, 2014.

[33] L. J. Strahilevitz. A social networks theory of privacy. In *American Law & Economics Association Annual Meetings*, pages 919–988, 2005.

[34] K. Strater and H. R. Lipford. Strategies and struggles with privacy in an online social networking community. In *Proceedings of the HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction*, pages 111–119, 2008.

[35] F. Stutzman, R. Gross, and A. Acquisti. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of Privacy and Confidentiality*, 4(2), 2013.

[36] J. M. Such, A. Espinosa, A. Garcia-Fornes, and C. Sierra. Self-disclosure decision making based on intimacy and privacy. *Information Sciences*, 211(0):93 – 111, 2012.

[37] S. van Otterloo. The value of privacy: optimal strategies for privacy minded agents. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pages 1015–1022, 2005.

[38] D. Villatoro, G. Andrighetto, J. Sabater-Mir, and R. Conte. Dynamic sanctioning for robust and cost-efficient norm compliance. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 414–419, 2011.

[39] M. Vorvoreanu. Perceptions of corporations on facebook: An analysis of facebook social norms. *Journal of New Communications Research*, 4(1):67–86, 2009.

[40] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. I regretted the minute i pressed share: A qualitative study of regrets on facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 10, 2011.