Enforcing Contextual Integrity With Exposure Control*

Mainack Mondal
University of Chicago
Chicago, USA
mainack@uchicago.edu

Blase Ur
University of Chicago
Chicago, USA
blase@uchicago.edu

ABSTRACT

The normative model of contextual integrity (CI) equips individuals to reason about privacy requirements and violations in online systems. However, a subsequent step is the enforcement of CI in online systems via privacy-management mechanisms. In this work, we first investigate the suitability of access control, the dominant privacy management model in online platforms, in filling this role. We argue that access control is insufficient for enforcing CI because it does not consider the set of expected recipients for a piece of content. To that end, we identify the privacy model of exposure control as an extension of access control to better enforce CI. We discuss the effectiveness of exposure control in better enforcing CI and describe a generic prediction-based framework for controlling exposure in online systems.

CCS CONCEPTS

• **Security and privacy → Social aspects of security and privacy**; **Usability in security and privacy**;

KEYWORDS

contextual integrity, social media, online privacy, exposure control

ACM Reference Format:
Mainack Mondal and Blase Ur. 2018. Enforcing Contextual Integrity With Exposure Control. In *Proceedings of Symposium on Applications of Contextual Integrity (CI-SYMPOSIUM'18).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Data privacy is a complicated, yet supremely important, research domain that has taken many forms over the past century. Warren and Brandeis defined privacy as the "right to be let alone" in their seminal 1890 article [24]. Since then, scholars have captured different dimensions of privacy in offline world via a number of definitions [1, 2, 13, 25]. The explosion in recent years of personal data sharing in online platforms, especially on platforms like Facebook and Twitter, has renewed the discussion on privacy due to the scale of data sharing (billions of pieces of content shared daily), the type of shared information (e.g., personal opinions and pictures), and individuals' privacy concerns.

Consequently, researchers have attempted to unpack the nuances of online privacy by proposing new theories [18, 20, 22]. In particular, Nissenbaum proposed the normative model of contextual integrity (*CI*) [18]. CI captures privacy requirements and violations based on context-sensitive norms and flows of information. In a nutshell, CI models how data should be transmitted based on the normative appropriateness of the flow of information. This theory is effective in capturing nuanced and subjective aspects of online privacy. However, while CI presents an understanding of what is or is not a privacy violation, a subsequent step is to actually build mechanisms for preserving privacy by enforcing CI.

Today, the dominant model for building privacy-management mechanisms is the model of access control. In the model of access control, users specify a list of entities who are allowed to, or denied from, viewing particular content. Leveraging real-world case studies, we argue that although access control is effective as a first step toward building privacy-preserving mechanisms, it fails to capture violations of CI in crucial scenarios. Specifically, the model of access control requires the concrete enumeration of recipients before sharing data, a task which incurs significant cognitive burden for users. However, failure to enumerate the expected recipients exactly and precisely can result in violations of CI.

In the context of enforcing CI, we then consider exposure control, Mondal et al.'s recently proposed extension of the access control model [15]. Exposure control captures the notion of expected recipients (termed the "expected exposure" of information) and contrasts it with the actual recipients of the information to capture privacy violations. We argue that exposure control is suitable for building more usable CI enforcement mechanisms.

Finally, we present a generic prediction-based framework to control exposure in online platforms. In this framework, we point out key challenges toward building privacy mechanisms to control the exposure of content.

2 BACKGROUND

We provide brief background on two relevant domains: the theory of contextual integrity and the model of access control.

2.1 Privacy as contextual integrity

Contextual integrity, or CI, is a theory of privacy developed by Nissenbaum [4, 18]. CI provides a normative model of privacy, i.e., CI considers what are the normal and expected behaviors related to data flows. In this work, we focus on CI in the context of online data sharing. CI presents four descriptive claims [18]:

(1) Privacy is preserved by *appropriate* flows of information.
(2) Appropriate flows of information conform to *contextual information norms*.

*Portions of this work appeared earlier as part of Mainack Mondal's Ph.D. thesis [14] and a USEC 2014 paper [15].

CI-SYMPOSIUM'18, September 2018, Princeton, NJ USA

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

(3) Each contextual information norm consists of five independent parameters: data subject, sender, recipient, information type, and transmission principle.
(4) Conceptions of privacy are based on dynamic ethical concerns that evolve over time.

CI provides a framework to argue about the violation or preservation of privacy. Crucially, it allows for the evolution as well as alteration of informational norms, such as the changes in norms caused by the advent of online social media. The CI framework states that for any given social context, there are informational norms defining appropriate flows for various types of information. Breaking these norms results in a privacy violation.

A prime example of the effectiveness of CI is to explain the phenomenon of "privacy in public." Users of online social media might upload personal content (e.g., a Facebook profile picture) that, based on its privacy settings, is accessible to everyone in the internet. However, users might feel their privacy is violated when a third-party crawler collects this "public" information and shares it in easily accessible form [10]. According to CI, privacy is violated in this case because the data collection does not conform to the transmission principle; the user did not consent for their content to be included in this third-party data collection.

Thus, even if content is shared with the public, there is some notion of privacy of that information. CI gave rise to a flurry of research in recent times to formalize CI [4], apply CI notions to understand privacy in online systems [19], and develop crowdsourced mechanisms to build contextual norms [21]. Very recent work surveyed the computer science literature to identify the main themes on how contextual integrity is captured in computational systems research [5]. In contrary, we aim to understand the effectiveness of existing data-management models in enforcing CI.

2.2 The model of access control

In online social media and other online systems, the majority of the available privacy tools (e.g., Facebook's audience-selector tool [8]) can be modeled by a simple *access control* model. Access control requires that one enumerates the users, groups, or roles who are, or are not, permitted to access information. However, the popularity of online social media has led to a renewed discussion about whether access control is a satisfactory model for user privacy. For example, in the "privacy in public" scenario, access control is not violated (based on its privacy setting, the content was shared publicly) despite subjective impressions that privacy was violated.

In this work, we are interested in the data-management mechanisms that online systems provide users to manage access to their content, and thus we focus on access control. Other work [3, 9] focuses on the orthogonal concern of protecting users' content from the site operator.

3 ACCESS CONTROL FOR ENFORCING CI

We examine a generic online social media site as a proxy for online systems broadly. Our scenario is that a user (or *uploader*) is uploading their content to a social media site.[1] The uploader is concerned about the privacy of their own content. In the access control model, they manage privacy by allowing or denying access to others by explicitly choosing whether to include those users in an access control list. The uploader can also change the access control settings of their uploaded content at any point of time in the future. We start investigating the effectiveness of access control in enforcing CI via two real-world case studies:

1. When Facebook introduced the *News Feed*—a feature that automatically presents updates from friends when a user logs in, as opposed to requiring the user to visit the friends' pages—users objected strongly and accused Facebook of privacy violations [6]. The News Feed did not change the access control policy; all users who could view content through the News Feed previously had access to it. However, the change from a pull mechanism to a push mechanism resulted in a change in the transmission principle (all logged-in friends received the data, not only the ones who explicitly visited the uploader's page). This change in transmission principle resulted in an effective change in actual recipient set, i.e. the set of users who actually viewed the content. Thus, according to CI, the News feed has violated a contextual informational norm.

2. The data aggregator *Spokeo* links together public information from different services (e.g., government databases, social media). While each individual piece of content that Spokeo aggregates is publicly available, users may feel that their privacy is violated when this information is linked. For example, Spokeo cross-references addresses with property records, allowing others to quickly estimate someone's wealth. In this case, there is again a violation of contextual informational norms in the form of changed transmission principle (storing social media content automatically for further analysis). This results in a mismatch between social media contents' expected recipients (social connections) and actual recipients (a data aggregator).

In each case, CI illuminates a potential privacy violation due to a change in transmission principle and consequent change in the *actual* recipient set, or the set of users who actually view the information. Concerningly, however, access control does not capture these privacy violations.

Ineffectiveness of access control to enforce CI: Access control captures a very basic notion of contextual integrity. Given a piece of content, access control ensures privacy only when all five parameters of informational norms are fully specified by users a priori via access control lists (list of users who are allowed/denied to access content). However, enforcing contextual integrity requires additional mechanisms to help users express privacy preferences beyond those captured by access control. For example, earlier work [19] revealed that people have an implicit idea about how information should flow in a platform to preserve CI (e.g., a user's location, although publicly posted, should perhaps only be revealed to the people who are geographically close to the user). This observation implies that users internally have an expected set of recipients.

To this end, we ask if there is an extension of access control which, on top of user specified list of recipients, also take the difference between *expected* and *actual* recipient set in consideration. This question brings us to the model of exposure control.

[1] In this work, we are interested in how users can protect the privacy of content they themselves upload. We consider how content uploaded by one user could violate the privacy of another user out of scope.

4 EXPOSURE CONTROL FOR ENFORCING CI

We start with a brief description of exposure control, previously defined by Mondal et al. [15]. Let I be an item of information (e.g., that Alice's date of birth is Jan 1, 1980). Informally, I's exposure is defined to be the set of principals we expect to eventually learn I. The exposure set includes principals who learn I directly from Alice or indirectly from a third person with knowledge of I, and those who infer I from other knowledge available to them. The exposure of an item of information may change over time. For instance, when a little-known website is listed on Slashdot, the set of users likely to discover the information contained in it increases dramatically and unexpectedly.

In the model of exposure control, a user is more likely to feel that her privacy is violated if she is surprised by the fact that certain people have learned the information. Specifically, a user has some expectation about the set of people who know or are likely to learn an item of information. Users tends to feel their privacy is violated if the actual exposure of an item (i.e., actual recipients) includes many more people than the expected exposure (i.e., expected recipients) [7]. Note that the expected exposure of a piece of content is an estimation of the expected recipient set. This expected exposure set can either be explicitly specified by user (e.g., by enumerating in an access control list every user expected to view a piece of content) or in a more realistic case can be predicted based on the past history of user actions in an online system (e.g., the number of likes, shares, comments, or views obtained in the past).

4.1 Enforcing CI by controlling exposure

In order to demonstrate the effectiveness of exposure control, we revisit the two case studies from Section 3.

1. The notion of exposure captures the changes caused by the introduction of the Facebook News Feed. Prior to its introduction, the exposure of an item I on Alice's profile was the number of unique users who visit Alice's Facebook page during I's lifetime, which could be much smaller than the set of users N_I with permission to access I, particularly if the content was publicly visible. With the News Feed, in contrast, I's exposure potentially includes all of Alice's friends, plus any user in N_I who Facebook deems potentially interested in I. I is pushed to these users, who will learn I serendipitously the next time they log into Facebook.

2. Spokeo aggregates people's personal information, including their name, address, data of birth, income, property value, and family tree from different online sources, making it available and searchable under the person's name and place of residence. By making it far easier to learn this information, exposure is increased.

Exposure control captures privacy violations in both of these cases where access control fell short. Exposure control can capture these violations because it considers differences between the uploader's expected set of recipients and the actual set of recipients.

These case studies demonstrate that exposure control can help to capture privacy violations not captured by access control. The reason is that access control requires users to extensively enumerate recipients and concretely specify the transmission principle. However, given the enormous amount of data shared online and the ever-changing nature of social relationships, this is a daring and often impossible task for users. Exposure control proposes to model the user's expected exposure for a piece of content. Both the expected exposure set and the informational norm parameters (specifically recipient and transmission principle set by a user) are often rooted in the same background knowledge, namely past user experience. Intuitively, controlling exposure provides a more effective way than access control to enforce contextual integrity.

Furthermore, exposure control captures the idea of "privacy in public." CI points out that there are implicit informational norms in users' mental models and hence there are appropriate information flows associated with public information. Capturing this same notion, the exposure control model points out that even for public content (which anybody can access), the user will have an expected exposure set in mind.

While the model of exposure control extends access control to better capture and enforce CI, it is not a silver bullet and has a number of limitations. Specifically, there are aspects of CI that are still not captured by exposure control. Exposure control leverages users' past behavior as a proxy for a user's mental model, yet there are implicit contextual norms that are deeply rooted in social processes. For example, suppose a social media user is not aware that they are being stalked by someone they know and therefore include the stalker in their expected exposure set. However, stalking violates social norms and thus violates contextual integrity.

5 MANAGING PRIVACY VIA EXPOSURE

A key challenge of exposure control is how to compute content's expected exposure automatically. Given that expected exposure is rooted in users' mental models, our intuition is that past interaction history can serve as a proxy. In this section, our goal is to propose a general prediction-based methodology that could be broadly applied to control the exposure of users' information in a variety of online systems.

5.1 Predicting exposure

Modeling and predicting the growth in popularity of information like Facebook photos, Twitter posts, or YouTube videos [11, 12, 23] has received significant research attention. We believe that these prediction techniques can be leveraged to predict content's expected exposure. These studies use empirical data of how information became popular in the past to build models for information propagation that can predict the future popularity of similar information. The prediction models vary from very simple models that extrapolate from the historical growth in popularity of a single piece of information to more complex models that take into account factors including attributes of the information (e.g., quality, type, and length of a video), historical data about the spread of similar content, and the effectiveness of dissemination channels (e.g., personalized recommendations or search results).

5.2 Empowering users to control exposure

Providing users with accurate exposure estimates for their information does not by itself eliminate the risk of privacy violations. System designs need to enable users to tune the exposure to the values they desire and also build mechanisms to alert users when the actual exposure diverges significantly from the predicted exposure. Below, we propose mechanisms to achieve these goals.

Tuning exposure: When a user finds that the predicted exposure of information is different from what they desire, a user could tune the exposure in several ways. First, they could enable or disable one or more dissemination channels. For example, on Facebook, they could opt out of being part of "directory or graph search."

Second, users can resort to more expansive or restrictive access controls to change the exposure of information. For example, to increase the exposure of information originally shared with social media friends, a Facebook user might choose to make it available to friends of friends. To decrease exposure, the user might choose to make it available to only a subset of friends. Since exposure control is an extension of access control, by changing access controls, the user can expand or contract the list of possible viewers and thereby change the list of expected viewers.

Limiting divergence from predictions: Even after a user tunes the exposure to match their expectations, unanticipated events (e.g., a post going viral) might cause the actual exposure to deviate significantly from the predictions. Intuitively, it is extremely difficult for any model to predict these anomalies in advance.

To minimize privacy violations in such scenarios, systems could adopt *tripwires* that automatically make content inaccessible whenever its actual exposure deviates significantly from the predicted exposure, notifying the user of this divergence. Upon notification, users can explicitly choose to keep the information inaccessible or re-enable access to the information (and readjust the tripwires). Alternatively, systems could allow users to specify tripwires that upper-bound the views (e.g., no more than 10 views per day or 50 views in total). These tripwires could therefore play a role in enforcing CI's transmission principle even when content gains unexpectedly large amounts of attention.

We expect that tripwire mechanisms can be easily enabled in current systems like YouTube or Facebook. In fact, YouTube already allows users to limit the total number of views to their videos to a preset value of 50 (effectively providing a limited form of exposure control) [26].

Aside from the challenges associated with predicting and controlling exposure, additional future research directions include building interfaces to easily convey the predicted exposure of a set of content and reducing the overhead of fine-tuning exposure. There are already a few systems that control exposure in specific contexts [16, 17]. However, addressing these challenges in the general case would open new avenues for building online platforms that could better enforce notions of CI.

6 CONCLUSION

In this work, we discussed the effectiveness of access control, the traditional model for managing privacy, for enforcing CI in online platforms. Through case studies, we point out the inadequacy of access control, leading us to recommend exposure control as a candidate for better enforcing CI. We provide an initial implementation framework for controlling exposure in online world. Consequently, we identify the key challenges to build exposure-control-based management systems for better enforcing CI. These challenges provide concrete future research directions to build practical CI-enforcing systems in online platforms.

REFERENCES

[1] Irwin Altman. 1975. *The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding.* Brooks/Cole Pub. Co.

[2] Irwin Altman. 1977. Privacy Regulation: Culturally Universal or Culturally Specific? *Journal of Social Issues* 33, 3 (1977), 66–84.

[3] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. 2009. Persona: An Online Social Network with User-defined Privacy. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM'09).* Barcelona, Spain.

[4] Adam Barth, Anupam Datta, John C. Mitchell, and Helen Nissenbaum. 2006. Privacy and Contextual Integrity: Framework and Applications. In *Proceedings of the 27th IEEE Symposium on Security and Privacy (IEEE S&P'06).* Oakland, CA, USA.

[5] Sebastian Benthall, Seda Gürses, and Helen Nissenbaum. 2017. Contextual Integrity Through the Lens of Computer Science. *Found. Trends Priv. Secur.* 2, 1 (2017), 1–69.

[6] Danah Boyd. 2008. Facebook's Privacy Trainwreck. *Convergence: The International Journal of Research into New Media Technologies* 14, 1 (2008), 13–20.

[7] delete-tweets [n. d.]. Deleting My Teenage Tweets: A Student Journalists Perspective. http://ajr.org/delete-tweets/. (Last accessed on June 2017).

[8] Facebook. [n. d.]. Basic Privacy Settings & Tools. https://www.facebook.com/help/325807937506242. (Last accessed on June 2018).

[9] Saikat Guha, Kevin Tang, and Paul Francis. 2008. NOYB: Privacy in Online Social Networks. In *Proceedings of the 1st ACM workshop on Online social networks (WOSN'08).* Seattle, WA, USA.

[10] Hacker proves Facebook's public data is public [n. d.]. Hacker proves Facebook's public data Is public. https://techcrunch.com/2010/07/28/hacker-proves-facebooks-public-data-is-public/. (Last accessed on June 2017).

[11] Liangjie Hong, Ovidiu Dan, and Brian D. Davison. 2011. Predicting Popular Messages in Twitter. In *Proceedings of the 20th International World Wide Web Conference (WWW'11).* Hyderabad, India.

[12] Kristina Lerman and Tad Hogg. 2010. Using a Model of Social Dynamics to Predict Popularity of News. In *Proceedings of the 19th International Conference on World Wide Web (WWW'10).* Raleigh, NC, USA.

[13] Stephen T. Margulis. 2003. On the Status and Contribution of Westin's and Altman's Theories of Privacy. *Journal of Social Issues* 59, 2 (2003), 411–429.

[14] Mainack Mondal. 2017. *Understanding & controlling user privacy in social media via exposure.* Ph.D. Dissertation. Saarland University and MPI-SWS.

[15] Mainack Mondal, Peter Druschel, Krishna P. Gummadi, and Alan Mislove. 2014. Beyond Access Control: Managing Online Privacy via Exposure. In *Proceedings of the Workshop on Useable Security (USEC'14).*

[16] Mainack Mondal, Johnnatan Messias, Saptarshi Ghosh, Krishna P. Gummadi, and Aniket Kate. 2016. Forgetting in Social Media: Understanding and Controlling Longitudinal Exposure of Socially Shared Data. In *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS'16).*

[17] Mainack Mondal, Bimal Viswanath, Allen Clement, Peter Druschel, Krishna P Gummadi, Alan Mislove, and Ansley Post. 2012. Defending against large-scale crawls in online social networks. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT).*

[18] Helen Nissenbaum. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford University Press.

[19] Xinru Page. 2013. Contextual Integrity and Preserving Relationship Boundaries in LocationSharing Social Media. In *Proceedings of the Measuring Networked Social Privacy Workshop.* San Antonio, Texas, USA.

[20] Leysia Palen and Paul Dourish. 2003. Unpacking "Privacy" for a Networked World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03).* Fort Lauderdale, Florida, USA.

[21] Yan Shvartzshnaider, Schrasing Tong, Thomas Wies, Paula Kift, Helen Nissenbaum, Lakshminarayanan Subramanian, and Prateek Mittal. 2016. Learning Privacy Expectations by Crowdsourcing Contextual Informational Norms. In *Proceedings of the Fourth AAAI Conference on Human Computation and Crowdsourcing, HCOMP.*

[22] Daniel J. Solove. 2006. A Taxonomy of Privacy. *University of Pennsylvania Law Review* 154, 3 (2006), 477–560.

[23] Gabor Szabo and Bernardo A. Huberman. 2010. Predicting the Popularity of Online Content. *Communications of ACM* 53, 8 (2010), 80–88.

[24] Samuel D. Warren and Louis D. Brandeis. 1890. The Right to Privacy. *Harvard Law Review* 4, 5 (1890), 193–220.

[25] A.F. Westin. 1970. *Privacy and Freedom.* Bodley Head.

[26] youyube exposure [n. d.]. How do I share a private YouTube video with someone? https://webapps.stackexchange.com/questions/7588/how-do-i-share-a-private-youtube-video-with-someone. (Last accessed on October 2017).