UNIVERSITÉ Concordia UNIVERSITY

CIVI 6991

Using urban big data in built environment research; a review of privacy risk assessment

Seyed Ali Enjavi Amiri

ID: 40094117

Supervisors: Dr. Mazdak Nik-Bakht, Dr. Mohamed Ouf

Summer 2020

Department of Building Civil and Environmental Engineering

Contents

Abstract

This report includes explanation of underlying aspects of tackling privacy problems in urban smart built environment. Then, it defines important components related to privacy which has an impact on it. It also offers a framework for researchers working in built environment to follow and reach a result indicating potential privacy risks associated with applications operating in the field which is posed on end-users using them.

Introduction

Until a decade ago, most urban-related data was generated using surveys, census, or basic administrative information (e.g., building permits, tax assessment rolls…etc.). These methods often provided aggregate high-level snapshots of city dynamics at specific moments in time that may quickly go out of date. Recently, new forms of urban-related data have emerged thanks to new technologies such as smartphones, RFID cards, motion sensors, drones, and social media among many others. These new forms of data fundamentally revolutionized the scope, frequency, and resolution of data generated from cities, and offer immense untapped possibilities to better understand the dynamics of our urban environment. Until now, there are extensive papers focusing on big data privacy issues of smart cities, however, there is a limited number of papers focusing on the privacy issue of built environmental research.

Problem Statement

A vast majority of the existing literature have small scales, focus on one aspect of the context, and overall, lack a solid and vivid explanation of what a smart built environment is capable of providing for cities around the world. Also, with the daily advancements in new technologies and new possibilities also comes new threats posed on privacy. Researchers who would want to work in this area and context are stepping into a tangled and complex web of privacy issues in numerous layers and aspects.

Objectives

Moving forward from the start of the research process, the objectives changed due to different reasons. An as-wide-as possible scoped literature browsing, listing, reviewing and narrowing was done to form a better scope and to find what was missing. Based on the gap and with the privacy related problems in mind, different approaches to address these privacy issues were studied and analyzed. And, in the end, a framework was developed to assess privacy risk of applications working in this environment.

Background

The task was to define a framework for assessing privacy risks of applications related to the field. Basically, breaking them down to smaller pieces and seeing how and why they might pose a threat to the privacy of end-users. Most guidelines crafted by researchers, authorities and regulators either focus on a small and specific scope of work (e.g. privacy risks associated with medical and health monitoring applications on smartphones) or discuss general and ordinal terms (e.g. ISO 31000:2018 Risk management — Guidelines). Hence, a clear, direct practice that addresses privacy risks within this paper’s scope is needed; A tool which quantifies risks correlated with different aspects of an application with tangible units resulting in a distinct and comparable outcome (Wagner & Boiten, 2018).

Primary audience of this framework is researchers. Especially, who work with urban data in built environment scope can follow this framework to see what privacy risks are associated with their area of research and are posed to end-users. They might find it useful to build upon and further develop and expand the framework to fit their need and area of work. On a broader level, developers and policymakers can also benefit the framework. Companies and organizations who want to develop a new application and deploy it in built environment can assess different potential architecture and direction for their product design based on its risk using the framework. Authorities and decision makers can use the framework to evaluate proposed options of implementing smart solutions in their city or business and facilitate the process of choosing.

We worked alongside with our teammates from the philosophy and communication departments and shared ideas, resources and feedback. A specific and more solid shared point between the framework section and the section on definition of trust is manifested here as we both emphasize on the importance of communication and educating the end-user as the main stakeholder.

It is important to emphasize the main stakeholder here; The end-user. The end-user must always be the focus and all efforts should lead to improving the applications in a way that refines user experience, well-being, security, etc. It is needless to say that users living and working in built environment are drastically diverse in terms of education, age, income, social class, gender, etc. All these factors affect the level of their digital literacy and knowledge. This concept is typically not considered in any level of application design, development and monitoring and also in broader levels of policy making. The assumption is the end-user possesses strong digital navigation and privacy protection skills and that she has good knowledge of collection, process and distribution aspects of an application. This however, is mostly not the case and evidence suggest that the digital literacy divide plays a big role in making a difference when it comes to privacy protection. It is absolutely crucial to be aware of this fact and always look for ways to communicate with users and raise their knowledge. After all, no matter how many risk mitigation measurements are taken, a user who lacks the necessary skills to work with an application can jeopardize her and potentially other people's privacy (Park, 2011).

As a good starting place and inspiration for many key components of the framework the OWASP list is mentioned and from that point we move to explaining the framework and then defining each component and finally reaching a table summarizing the whole process in a compact visual presentation. The effort here for the framework is to be general enough to encompass the vast scope and also to have related components. What is put in the definition of each component and the weights and impacts are accumulation of several and various sources and analysis.

OWASP top 10 privacy risks can be a good starting point and is worth mentioning here (Foundation, 2016). It mostly contains web application privacy risks covering both organizational and technical aspects. It provides ways to check potential risks and also possible countermeasures. Insufficient Data Breach Response, Non-transparent Policies, Terms and Conditions and Sharing of Data with Third Party are some examples from the list. Each risk has its unique checklist to go through and mitigation procedures. More examples and resources are also provided for more detailed investigation. This can be a thought-provoking process for researchers and help them to consider more aspects which might have been unnoticed.

Methodology

What follows is a list of key components needed to be evaluated based on an application to reach a good understanding of required risk assessment components. Some have simple and practical aspects that are easy to answer and some have deeper, more philosophical layers demanding more contemplation. These components are put together here based on a literature review set to look for items affecting the ideology and end goal of the framework. Based on their effect on the end risk profile a weight is assigned to each component. Depending on the nature of the subject, different risk impacts are explained and distributed to its subcomponents and alternatives. This will all add up to a final number defined as risk grade resulting in categorization of applications in low, mid and high-risk classes.

Data collector

A clear distinction, which is a pivotal element for further assessment, is the collector of data (Zoonen, 2016). In any case, the organization's characteristics should be evaluated through ways such as studying its privacy policy. With having government/private distinction in mind, it can be argued that if the government is the collector, in a transparent environment, obtaining solid rules and regulations stating detailed policies is easier than a reluctant private company competing with its rivals. Another aspect on the other hand, is consent which is typically rendered meaningless in a public place with sensors collecting data for the government. This can also be the case when using a public service or utilities provided by a private company. One cannot simply choose the impractical option of no longer using these facilities. Only when a user is evaluating the alternatives for personal use, like smart home appliances, can he choose to opt out of using a certain application and assert his consent

(Hornberger, 2014). Therefore, a weight of 1 is given to the data collector. The risk impact consideration is 1 for governments and 2 for private companies for the first level, with risk impact of 1 for personal use and 2 for public use as the second level. The final impact is the multiplication result of the two levels.

Data collection purpose

Art. 5(b) of GDPR describes two building blocks of purpose limitation of data and states personal data shall be: '[1] collected for specified, explicit and legitimate purposes and [2] not further processed in a manner that is incompatible with those purposes.' This is based on Article 29 of the EU Data Protection Directive which was further analyzed and clarified by a Working Party in 2013. This published work, in summary, states that purpose specification comes down to avoiding vague terms such as 'improving user experience'. According to the party, a case by case analysis is required to test the future use compatibility of data with the original purpose. Several examples are provided for both blocks which thoroughly explain the guideline in practice. A more specific approach to purpose is defined in (Zoonen, 2016), two general end targets can be considered for an application: service or surveillance. The former usually raises fewer privacy concerns and it focuses on providing collective benefits. The latter, however, can be more intrusive and focuses on small groups of people with the potential consequences for individuals. The effort is put here to determine the privacy-purpose trade-off of the application in the end. The weight of 2 is given to data purpose, with impacts of 1 for service and 2 for surveillance.

Data type

Although a piece of data can be private from the point of view of one person and the opposite from another's, the application's data type should be put in categories initially and then further labeled with more detail. These categories include location-based, body and mind, social life, behaviors and actions, and finally media. Each of these data types pose different privacy risks and higher scrutiny is needed if more than one type is collected by the application (Eckhoff & Wagner, 2017). Data type weight is 3 and its components have different potential risks. Location-based data has the impact of 3, body and mind of 2 and media and behaviors and actions of 1. If more than one type of data is collected, the impacts are added together.

Data ownership

It is important to shed light on data ownership. It should be determined if and to what extent users have control over their collected information by an application. The case of full control is also known as "right to be forgotten" which means a user can ask for erasure of their personal data from an organization under certain circumstances. Typically, these circumstances are in favor of the user when the organization is violating a law or the user's consent or it is using the data only for direct marketing purposes, and they are in favor of the organization when it is processing the data for a public and collective beneficial goal (Union, 2018). The amount of control that users have over their data, determines how fast and easy they can act in the event of change of mind or perspective or even a privacy violation to limit or totally delete their collected information from an application. Data ownership gets a weight of 1. If some level of restricting is available for the user, the impact is 1 and if on the other end of spectrum, the application claims full ownership of data, the impact is 2.

Architecture

The application's digital structure and also its physical infrastructure enabling it to run are the center of discussion here. The way it is developed and put together as a final product is based on an architecture. This can partly be outsourced to third party companies as well, which can provide a service at any level or different aspects. Various ways of protection should be applied to different parts of the final products architecture as well.

Third-party involvement could potentially be beneficial or harmful. Additional players in the field of data collection, storage and process would mean additional opportunities for outside attackers to violate privacy of the users. On the flip side, a third-party organization which is trusted by users and reliable based on rules and regulations can lessen privacy risks by providing for instance a secure cloud storage place (Kim & Kim, 2011). This case for example, adds additional layers of protection and data separation (Eckhoff & Wagner, 2017)

One of the most direct ways to protect privacy of users is using proper cyber-security measures. (Johnsen, 2017) provides a good literature review on software ecosystem vulnerabilities and mitigation solutions for improving safety, security and resilience. It is worth mentioning the distinction between safety and security in the context here. (Firesmith, 2003) gives clarifying definitions as "the degree to which malicious harm is prevented, reduced and properly reacted to" for security and "the degree to which accidental harm is prevented, reduced and properly reacted to" for safety. The implemented measures are meant to protect the application and infrastructure from threats to both safety and security. These methods vary greatly depending on the application's nature, architecture, processing power of the devices used, etc. Ideally, the knowledge about the measures is available to everyone which makes risk assessment analysis easier. Otherwise, this inaccessibility itself can tarnish the trust users are putting in an organization. Data encryption, anonymization, separation, onion routing, etc. are few examples of technical methods to secure privacy (Eckhoff & Wagner, 2017).

The result is a myriad of application architecture and structures, various parties involved and numerous technical differences. Based on the nature of technologies involved and the final ecosystem under evaluation, different cybersecurity tests such as penetration, DDoS, robustness, etc. must be performed according to rules and regulations (Johnsen, 2017). The weight of 4 is assigned to architecture. Approval percentage of different tests on an ecosystem makes up its risk impact. Performance approval of 80- 100, 50-79 and under 50 result in impacts of 1 to 3 respectively.

The summary of proposed risk grading system and the components described above can be seen in the following table. Applications graded between 11-22 are categorized as low-risk, between 23-33 as mid-risk, and finally 34-43 as high-risk.

Table 1. Risk Profile Summary

	Weight	Impact				
Data collector	1	Government	Private company		Public	Personal
		1	2	*	2	1
Data purpose	2	Surveillance	Service			
		2	1			
Data type	3	Location-based	Body and mind	Media	Behavior and action	
		3	2	1	1	
Data ownership	1	Semi-control	No-control			
		1	2			
Architecture	4	High-approval	Mid-approval	Low-approval		
		1	2	3		

Results and discussion

We started the research with an open mind and tried to test and go through papers with different focuses on aspects of privacy issues, smart cities, risk assessment, etc. As time went by, we formed a better idea of our outline and what we were aiming to achieve. Once we had a clearer vision, we each took the responsibility of our assigned task and narrowed our scope of research and focus. Through a comprehensive review of academic papers, website articles, laws and regulations, etc. we defined our scope and developed a framework; To better illustrate what we mean by built environment and the tool necessary to navigate us when we are dealing with privacy problems in this environment.

The hope is that the scope is thorough and the framework is general enough to encompass endless possible combinations of integrated technologies operating in built environment in smart cities; Not so small to ignore other prominent aspects and not so vague and ordinal to be similar to existing guidelines.

Reference

Eckhoff, D., & Wagner, I. (2017). Privacy in the Smart City—Applications, Technologies, Challenges, and Solutions. *IEEE Communications Surveys & Tutorials, 20*(1), 489-516.

Firesmith, D. (2003, December). *Common Concepts Underlying Safety, Security, and Survivability Engineering.* Retrieved from Carnegie Mellon University: https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=6553

Foundation, O. (2016, April 8). *https://owasp.org/www-project-top-10-privacy-risks/.* Retrieved from https://owasp.org/www-project-top-10-privacy-risks/

Hornberger, J. G. (2014, March 4). *THE FUTURE of FREEDOM FOUNDATION*. Retrieved from https://www.fff.org/2014/03/04/private-vs-government-data-collection/

Johnsen, S. O. (2017, April 6). *Risks, Safety and Security in the Ecosystem of Smart Cities.* Retrieved from IntechOpen: https://www.intechopen.com/books/risk-assessment/risks-safety-and-security-in-the-ecosystem-of-smart-cities

Kim, K., & Kim, J. (2011). Third-party Privacy Certification as an Online Advertising Strategy: An Investigation of the Factors Affecting the Relationship between Third-party Certification and Initial Trust. *Journal of Interactive Marketing, 25*(3), 145-158.

Park, Y. J. (2011). Digital Literacy and Privacy Behavior Online. *Communication Research*, 215-236.

Union, E. (2018, May 25). *GDPR*. Retrieved from https://gdpr.eu/right-to-be-forgotten/

Wagner, I., & Boiten, E. (2018). Privacy Risk Assessment: From Art to Science, by Metrics. *Data Privacy Management, Cryptocurrencies and Blockchain Technology.* Barcelona.

Zoonen, L. v. (2016). Privacy concerns in smart cities. *Government Information Quarterly, 33*(3), 472-480.

This document has been submitted by the student as a course project report, for evaluation.
It is NOT peer reviewed, and may NOT be cited as a scientific reference!