Privacy in IoT: Expectations, Causes of Concerns, and Reasons for Concern Mitigation

Naushin Nower
Institute of Information Technology
University of Dhaka

ABSTRACT
Internet of things (IoT) connects billions of devices, people and services, and exchanges data among them. Moreover, IoT has scalability (in terms of the number of devices and sensors), proximity, ubiquity (mass development) and connectedness property which easily violates an individual's privacy by collecting and using personal data. Thus, there is an urgent need for a privacy-preserving tool to ensure an individual's privacy requirements with transparency and control. To develop these tools it is important to understand people's privacy expectations, implications and requirements of IoT to understand how people feel about their privacy requirements. In this paper, a rigorous analysis is performed on existing different surveys and interviews to find out individual's privacy expectations from IoT sensors, privacy concerns and reasons for privacy concerns mitigations. The finding suggests that although privacy preferences are diverse and context-dependent, still some general factors that affect all.

General Terms
Privacy in IoT

Keywords
Internet of things, Privacy, Concerns

1. INTRODUCTION
In recent years, the Internet of Things (IoT) has drawn significant research attention because of technological advancements and rapid convergence of wireless communication, digital electronics, and micro-electromechanical systems technologies. According to the Cisco report, the number of devices connected to the Internet has exceeded the number of human beings in the world. Reports show that the number of Internet-connected devices is expected to increase twofold from 22.9 billion in 2016 to 50 billion by 2020 [1]. IoT connects huge devices which include PCs, smartphones, tablets, cameras, microphones, Wi-Fi enabled sensors, wearable devices, tracking technology, household appliances, and network access into non-computer products. These connected devices produce a massive amount of data and much of these data is personal data. As a result, these huge amounts of data raise issues of control, consent, and transparency, and increasingly erode the boundary between the private and public life [2]. Moreover, IoT is characterized by the number of factors: scale (number of devices and sensors), proximity, ubiquity (mass development) connectedness, etc. These factors make it easier for the individual to identified, tracked, profiled and influenced and thus huge impact on an individual's privacy. As a result, people are more concerned about using different applications of IoT[3].

IoT intensifies existing privacy challenges and creates new ones. IoT sensors have increased sensor scale and proximity which creates potential concerns for continuous monitoring of people’s activities, behaviors, speech, health, and emotions. The ubiquitous presence of IoT sensors makes people identifiable in public and private spaces[2-3]. Moreover, IoT sensors are connected, as a result, it allows IoT companies to penetrate the privacy walls of the home and provide access to private data to third parties. Thus, the combination of the above trends – sensor scale and proximity, continuous monitoring, increased identifiability which breaches the walls of the privacy – points to a potential decrease in people’s ability to find private places of reserve and solitude generally.

Privacy concerns in IoT are a burning issue nowadays because IoT associates more gadgets together in the area of smart home, smart meter, healthcare, smart retail, wearable and many more. For example, the smart home provides huge potential in saving time, increases personal productivity and also provides a level of convenience [4]. But at the same time, it brings a lot of privacy concerns for the users. These concerns start with the digital assistant products that continuously listen to the activation words and people’s conversation at home and send these to the corporate servers. The same risk is associated with smart healthcare applications, where several devices (infusion pumps, heart monitoring implants, wearable, Fitbit) continuously monitor patients’ daily activity, tracks location and finally sends data to the cloud for further analysis. As a result, people are concerned about the privacy of personal data in the smart home, healthcare, and other IoT related applications. Risk of patients’ exposure to private life, data eavesdropping, ownership of data and location privacy is the most prominent concerns of using healthcare technologies.

As the number of connected devices increases in IoT, chances of personal data collection is also intensified. Thus people’s concerns about the privacy of their data are much more amplified. Thus, there is a need for a tool that smoothes the user's concerns by ensuring transparency and user control so that the individual's privacy requirements are met. To do that it is very important to understand people's privacy feelings and implications about IoT applications and need to recognize the situation where they want to control their privacy. Hence, it is normal that by minimizing the privacy risks people are more willingly accept the advantages of IoT. As a result, developers and researchers should understand how different factors influence and impact people's privacy perceptions in an IoT environment. This will enable them to design a privacy-preserving IoT system and services. To address this problem, a rigorous analysis of different surveys and interviews from existing work are conducted and provided meaningful insights that must be ensured before setting up an IoT environment. The contribution of this paper is as follows: 1) Identifying the user's expectations and preferences from the
IoT sensors, 2) Causes of privacy concerns, and 3) Situations and reasons when concerns are mitigated.

The rest of the paper is organized as follows. At first, the related work about privacy-preserving architecture and people's privacy feelings are discussed. Then, analyze different privacy-related study are analyzed and the reasons for people's privacy concerns, causes of concerns and how to mitigate privacy concerns are investigated. From the analysis, it can be concluded that before setting up an IoT environment, it is needed to address and solve the people's privacy issues properly.

2. RELATED WORK
The privacy implications of IoT devices are of significant interest to researchers and many works have been already done. Some of the related works focus on designing the privacy-preserving architecture and others concentrate on the privacy concerns in different IoT applications.

2.1 Privacy-Preserving Architecture
In the paper [5], authors have proposed a dynamic privacy analyzer (DPA) to protect user privacy from smart meter data. The proposed DPA receives smart meter data from the smart meter gateway at home and performs anomaly detection. It has two components: a) extraction of privacy requirement through anomaly detection and b) privacy quantification and preservation. The first component performs anomaly detection on smart meter data and based on the anomaly detection, it sends alert to the user's smart-phone. However, privacy requirements are diverse and depend on the individual. Thus to extract privacy requirements, there is an urgent need to analyze user's privacy from their verbal concerns, their privacy requirements, and feelings, etc rather than anomaly detection by machine.

The authors of the paper [6] have proposed a cloud-based privacy-invasive architecture. In the proposed framework, the user does not need to upload her raw data to the cloud nor hide the data from using any cryptographic methods. Instead, its service provider's responsibilities to preserve the necessary information and to discard irrelevant information. However, their work based on the assumption that the feature extractor module on the client-side is responsible to preserve user's privacy. However, privacy depends on the individual, culture and even on the country thus, feature extractor cannot be guaranteed to preserve an individual's privacy.

Besides these, in the paper [7], attribute-based encryption (ABE) technique is used to address the privacy and confidentiality of the data shared in blockchain-based IoT ecosystems. The authors have used ABE, where single encryption provides both confidentiality and access control and has been identified as a potential technology for data sharing in decentralized networks. However, none of these papers consider people's true feelings and concerns when preserving privacy.

2.2 Privacy Concerns
In the existing literature, many researchers focus on people's concerns, factors, individual's impressions on privacy. Studies have investigated various factors that can impact privacy on big data [8], and suggests some new methods of data collection in the IoT environments that have led to new privacy challenges and introduced new factors. Some of these challenges include obtaining consent for data collection, allowing users to control and choose the data they share, while at the same time ensuring the use of collected data is limited to the stated purpose [8].

The authors of the paper [9] concentrates on the contextual factors that affect users’ privacy perceptions of IoT environments. To understand the privacy perceptions, a public online survey (N=236) is deployed and interviews (N=41) are conducted to explore factors that could have an influence. However, they focus only on the concerns of the users and their analysis is based on an online survey and interview rather than experienced users. The paper [10] focuses on the people's privacy related to wearable devices. The authors identify potential privacy concerns about wearable's and finally conclude that users have different levels of privacy depending on types of the wearable they use.

The paper [11] discusses different influential factors about smart home user's perceptions of privacy. The authors have conducted several semi-structured interviews with smart home owners and investigated the perceptions of smart home privacy and risks. From their interview, it is observed that owners of smart homes value convenience and connectedness. The user wants to share their data whenever they perceived potential benefits. Finally, the paper concludes that users are skeptical of privacy risks from non-audio/visual devices.

3. PRIVACY EXPECTATIONS CAUSES AND REASONS FOR MITIGATIONS
In this section, the people's general privacy exceptions from IoT sensors, causes of privacy concerns and reasons for concerns mitigation are investigated. Most of the literature discusses privacy expectations from specific applications such as smart homes, smart meter, wearable devices, etc. However, some factors influence privacy in all types of applications which is the area of our concentration.

3.1 Privacy Expectations
The general discussion about user’s privacy expectations and preferences from IoT sensors data are widely explored in [12-15]. The survey investigates the privacy expectations of 1,007 Amazon Mechanical Turk US workers with 14 different scenarios, which are varied with 8 identified influential factors. The potential factors that extensively have impact on user’s privacy concerns are, the (1) data type, (2) location of data collection, (3) benefits of data collection, (4) the purpose of data collection, (5) the device of data collection, (6) inferred information from the collected data, (7) the retention time and the (8) sharing policy of data with others. Participants were given 14 different scenarios where each scenario includes 1) the type of data, (temperature, location, biometric), 2) device of data collection, 3) location of data collection (home, work, public restroom, 4) how data is used (whether it is shared, or inferred any other information) and 5) retention period. They were asked to provide their comfort level on a five-point Likert scale from Very comfortable, Comfortable, Neither Comfortable Nor Uncomfortable, Uncomfortable and Very Uncomfortable. They were also asked to provide free text answer of the following, whether the use of data collection to be beneficial for them, whether they would allow/deny the data collection in the described scenario, and how often they would like to get the notification about the data collection.

The results show that privacy preferences are diverse but most of them prefer anonymous data collection and short retention time. For example, participants feel comfortable with the temperature sensor (21% people provide very comfortable and 32% provide comfortable) or presence sensor (17% very comfortable and 21% comfortable) but show extreme
discomfort with video (36% strongly uncomfortable and 25% uncomfortable), biometric (45% strongly uncomfortable and 32% uncomfortable) or iris scan sensor (50% strongly uncomfortable and 30% uncomfortable), which can able to identify a person uniquely. Participants prefer public place as a location for their data collection rather than private places. They strongly oppose to collect any data in the home (44% strongly uncomfortable and 20% uncomfortable), public restroom (32% strongly uncomfortable and 29% uncomfortable) but workplace (17% strongly uncomfortable, 25% comfortable and 11% very comfortable) is considered as acceptable to them. Participants prefer short retention time based on use cases, for example they prefer immediate deletion or keep for a week. 33% participants reported forever retention time as a strongly uncomfortable condition. The results also show that 42% participants are uncomfortable as the collected data may be used to infer unwanted information.

From the survey data, authors perform statistical analysis to determine which factors are influential for making the decisions. Authors have constructed five generalized linear mixed models (GLMM) based on five dependent variables: comfort level, allow or deny decisions for the data collection, desire to be notified of data collection every time, desire to be notified once in a while, and desire to be notified only the first time. From this model, the influential factors for each dependent variable are determined based on the Bayesian Information Criterion (BIC) value. For example, comfort level highly influences by the data type (BIC value: 14633) and less influences by the retention time (BIC value: 18103). Allow/deny decision for data collection highly depends on data type (BIC: 15232) and location (BIC: 15297) and less influences by the shared factor (BIC: 18707). The data type is a prominent factor for participant’s desire to be notified every time also. In addition, user perceived benefit is the second important factor for the desire of every time notifications and the shared factor is the least one. The result is same for the desired to be notified once in a while also. And user perceived benefit is the most effective factor for desiring only first-time notification and the least effective factor is the data type and location.

It is observed from the free text response that, the user would like to protect their data using classical privacy and data protection rules (e.g. Fair Information Practices) and want to notify when the data is being collected. It was mentioned by 41% participants that being informed would make them easy to accept. Also, people want to know the purpose and benefits of data collection to make them comfortable.

These factors are influential for location tracking also and similar expectation and concern continues in participant’s mind. As an example, participants of the experiment [13] (the human proximity detection) [14] (space management and human interaction detection) and [15] (communication tracking) showed positive reactions because all the collected data were anonymous and they were well informed about the data collection. In addition, they also had a clear understanding of data collection process and use of the data.

4. CAUSES OF CONCERNS

People feel troublesome when their every step is noticed and the goal of data collection is not clear to them. This expression of location tracking is exemplified in [16]. Fisher and Monahan [16] investigate the social context of an RFID based tracking and monitoring system deployed in a hospital. Their work incorporates seven hospitals, technology company in Southwestern United States, a industry conference. They conducted a series of site visits and interviews at three hospitals, industry and also conducted phone interview at four flagship RFID hospitals. Their interview includes total 60 people including 12 hospital administrators, 8 physicians, 8 nurses, 17 technical hospital staff members, and 15 technology industry vendors and consultants in U.S. The interview mainly focused on staff’s involvement in the RFID system, how RFID was used in tracking, and effect on workload after implementing RFID in hospitals. The participants were also asked to provide their advice and policy to improve the RFID system. The interview revealed tracking as an extra burden because it seems “big brother” was watching the staffs. Specially, this concern was prominent among the nurses and in some cases, it was so troublesome for them to carry and check the RFID system operational, since carrying RFID was an extra work for them. Tracking hampered the privacy of the nurses since the official breaks were also encountered. Hence to improve the system’s acceptability, it is recommend (1) a customized RFID based system rather than track each step. It is proposed to meet the requirement of the hospital with (2) clear goals and policies and to focus on (3) the staff’s labour concern and privacy.

Besides [16], possible privacy implications of a RFID based workplace are focused in [17]. The authors focus four use cases of RFID systems in a workplace, which are identification, altering (take action based on the obtained information), continuous tracking and authentication and discuss privacy concerns on these use cases. Although wearable RFID is used to enhance productivity and ensure safety, it can disclose the time of a person spending in a workstation (by tracking and identification) or in front of a workstation (by authentication) and can also keep track of arrival and departure time. Continuous collection of data may reveal confidential information and may link to another database (e.g. employee’s medical record provide to get some benefit and payroll record etc.) with personal records, which is a clear violation of privacy. In extreme case, an employer unethically can perform data mining on the collected data to reveal more information (altering). Lack of strong encryption in RFID makes privacy concerns for possible unauthorized access and information leaking. In addition, there is no rule on how to collect and use of data from wearable technologies. Thus, precise policies of the data collection, share and usage are suggested in [16-17] before implementing the RFID in mentioned use cases.

Unauthorized access and lack of data sharing policy also create privacy concerns in healthcare domain. Benjamin [18] concentrates on the usability and limitations of RFID in user authentication, patient tracking, medication, and safety. Although, the usability of RFID is increasing in different sectors but privacy and security issue hampers its expected usability in healthcare domain. This is because unauthorized data detection and unencrypted data transmission within RFID tags are not property addressed. RFID tags stores unencrypted patient’s data (e.g. name, date of birth, medical record etc.) which can be easily accessed. Patient’s data privacy is very important in healthcare sector and HIPAA (Health Insurance Portability and Accountability Act) specifies 18 protect health identifier must be anonymized during sharing. As transmission of the unencrypted data within RFID may cause unauthorized access, which is a distil violation of privacy, thus suggests a standard for RFID data storage and encryption.
Thus the reasons for privacy concerns are 1) identifiable and continuous tracking, 2) unauthorized access of data and reveal of confidential information by linking to another database, 3) lack of strong encryption, 4) lacking of any rule to collect and use data.

5. REASONS FOR CONCERNS MITIGATIONS

However, recent studies observe that privacy apprehensions mitigate due to the transparency, trustability, and reliability of the system increases over time. For example, Mathur et al. [8] developed a system to understand the effectiveness of different metrics (e.g. noise, dress color, air quality, mood, and activity) in a quantified workplace of two European offices of research organizations. The survey and interview highlights the effectiveness of different metrics, employee’s reaction to data collection and visualization policy of the developed system. The survey incorporates 70 employees to understand employee’s experience with developed systems and semi-structured interviews with 20 employees (9 females and 11 males) to explain their involvement in the systems and the privacy concerns they have perceived. Among these 20 participants, 6 were non-European citizens, 14 were European citizens and their age was between 26-43. From the various collected data (noise, dress color, air quality, mood, and activity) 50.5% participants defined none of the data was privacy invasive, whilst 26.15% reported mood data could reveal their privacy, color and activity was defined by 23.08% and 21.5% users respectively as privacy invasive. Noise (9.23%) and air quality (1.54%) was defined least privacy invasive by the users.

The participants were asked to input their mood and activity in office tablets and fixed sensors were used to collect other metrics. The participants can visualise the refection of their input in mood lamp and dashboard immediately. Initially, the participants had hesitation about the reliability of the systems, but after one week their attitude changes when they had comprehended idea about the data processing and visualized true reflect of their input in the mood lamp and dashboard. It was also noticeable that the 92% employees prefer to respond on the office tablets rather than a personal smart-phone because input from a personal smart-phone can expose the identity. To preserve privacy they also preferred to input mood and activity when they were alone in front of the tablet. The result from interview showed that 50.5% of the participants had little to no concern, 21.5% were neutral, 24% were somehow concerned and only 3% were very concerned.

The most of the participants were confident because of anonymous data collection and real-time visualization of data without any manipulation. The reason for minor concern is to belong a small group where anonymity can be tracked. As a result, it is suggested to implement k-anonymity, such that the individual activity tracking can be infeasible. Thus, to get acceptance from the employee transparency, anonymous data collection and inclusion (ensure for everyone) are identified as most prominent qualities of a system.

More broad considerations of smart home habitant’s perspective on privacy are depicted in [20]. Zeng et al. [20] conducted semi-structured interviews with fifteen smart home users (4 females +11 males) to explore the need, privacy and security concerns and mental model associated with smart home users to provide a better recommendation for the smart home designer. Among these 15 smart home users, twelve users managed their smart homes, and three live in smart homes administered by others. The users of smart home were asked about the devices and installed apps, their technical understanding of the smart home (drawing or describing), security and privacy concerns, mitigation strategy and situation for multi-users scenario. A large number of internet-connected devices were mentioned by the smart home users e.g. intelligent personal assistant, thermostat, camera, power outlets and switches, motion sensor, hub, door lock, smoke detector, leak detector. Users mainly used smart home for physical safety (including security systems, door locks, and smoke detectors; 9 participants), home automation (automatically adjusting lighting, temperature, or other devices; 13 participants), remote control, and in-home sensing. Participants concerned about multiple vulnerabilities e.g. data at risk in the cloud (1 out of 15), weak passwords (5 out of 15), lack of transport level security (4 out of 15), insecure devices (4 out of 15), malicious devices (3 out of 15), unsecured Wi-Fi network (2 out of 15), devices can be unpaired (1 out of 15). Like vulnerabilities there was no particular concern on threat identified by majority of the habitants, e.g. Continuous audio/video recording (3 out of 15), data collection and mining (1 out of 15), adversarial remote control (4 out of 15), network attack on local devices (3 out of 15), spying by other user in home (3 out of 15). Although continuous audio and video recording could be a potential threat for privacy violation, only 3 out of 15 smart home users were concerned about it. However, it was mentioned that multiple users in a smart home create security and privacy challenges, when the primary user was more knowable and controlling capability than incidental users. The interview indicates that the users were tensed about the physical security of their house rather than privacy. Most of the habitants prefer to consider Amazon Echo or Samsung SmartThings as a trusted company to process data and few of them recommend additional privacy features and prefer more specific permission requirement on their device. In addition, the users were confident about service provider and believed mitigation strategies were enough. Thus, trustability is an influential factor for privacy concern mitigation.

This study incorporates smart home users only who are already using smart home. Thus for comprehensive analysis, non-smart home users are needed to incorporate for better understanding. Why not-smarthome users (i.e. not using smart home, for privacy, money or other reasons).

Another reason for privacy concern mitigation is to perceive the benefits of tracking technology. For example, in [15], privacy concerns are decreased, particularly when people are using wearable health tracking technology in workplaces. The authors of [21] focus on people’s attitude towards releasing every step count in a health promotion campaign for three weeks. The author made observations and interviews at a workplace participating in the step-counting campaign organized by a Danish company. Participants used different trackers e.g. pedometer and various smart-phone apps and inputted their steps and other activities (e.g. cycling, golfing, swimming by converting to steps) in the campaign website. The participants formed teams with the goal to walk 1000 steps within 11 working days among 21 campaign days. Total 17 people were participating among 28 employees of a department. The author utilized 12 workdays at the workplace and conducted 9 interviews with both participants and non-participants. It was noticed that participants expect their data used for the step-count purpose only. They did not want to allow linking these data to other databases to reveal more information. Non-participants express a concern in sharing...
every single step-count because of possible misuse of their data. This is because participants need to share their step counts and other daily activity on the website and among the participants. Thus, it is possible to relate private life by sharing step count experience with others. Besides some minor voices, most of the participants observe this sharing as innocent and as an opportunity for interaction and socialization with other colleagues. At the end of the campaign participants willingly accept that the benefits of using wearable health technology are more than concerns compared to they have perceived.

Besides this, the same authors focus on the gaps between the reasons for the health-tracking program and people’s experience in [22]. The intention of the study is to ensure health-tracking benefits for all by understanding people’s expectation. The author conducted interviews with three wellness program administrators and seven employee participants from seven different companies. After that, 581 (45 administrators+536 employees=504 from one (X) company+32 from other (O) companies)) people from fourteen companies were enlisted for the survey from author’s social network. To get more in-depth information, follow-up interviews were conducted with 11 survey respondents with different roles (e.g. employees and administrators, participants and non-participants). The whole study was conducted in North American companies from March to August in 2016. The administrators explained goal, design and implementation process of their health wellness program. The employees were asked to provide their previous experiences about health tracking programs, deciding factors for participation, expectations, their health goals etc. Previous health tracking experiences influence employee’s decision for taking part in the health-tracking program. Participants appreciate that health tracking increases interaction, accountability, and awareness and also perceive company’s care. Majority of participants (X: 83%, O: 78%) want to recommend the health tracking to other colleagues also. Some also prefer more customized heath tracing based on individual health condition. However, the 20% employee from company X and 13% employee from other companies did not participate. Non-participants express time-constrain (X: 51% O: 2 of 32), risk of data misuse, not attractive (18%) program and already being active (X: 28% and 2 of 32) as a reason for not to join in the health-tracking program. Some also have other health goals and prefer different health activities. Few of them (3% of one (X) company and 2 participants from other company(32)) criticise sharing step count as a privacy violation and perceive the risk of data disclosure. Sharing step-count and being monitored makes them uncomfortable in a workplace also. The study concludes multiple reasons for not participating and indicates that the privacy concern is not the primary reason. The authors express this gap as the socio-technological gap. (gap between what the outcome from society and what technology can provide)

Thus, from the analysis it can be concluded that privacy concerns are mitigated when the whole system is 1) transparent and reliable to the users and 2) when the user can perceived benefit.

6. CONCLUSION
Privacy concern is a long-debated issue regarding personal data collection, processing, usage, and transfer. From the above discussion, it is shown that privacy attitude influenced by the following factors: 1) type of data collection (anonymous or uniquely identifiable) 2) location of data collection 3) retention time 4) purpose of data collection 5) benefits of data collection 6) data sharing policy 7) inferred information 8) device of data collection. However, it is shown that concerns can be mitigated if it is properly addressed. From the recent study, it is visible that the adaptation of localization tracking technologies is increasing day by day. This is because; participants can perceive the advantages behind this and become confident about the system reliability. The analysis also exposes that the people react positively when they know the reason for data collection and how an organization uses their data. They want anonymous data collection, notification of data collection and prefer more customized tracking policy rather than a one-size-fits-all tendency. Thus the following have to ensure to mitigate privacy concerns 1) anonymous data collection 2) location should be work or public place with notifications 3) short retention time, immediate deletion or deletion after the purpose served, based on applications 4) the purpose should be informed before tracking 5) who is going to be benefited needs to clarify 6) data sharing policy should be clear and informed to the user 7) whether any information is inferred from the collected data needs to be mentioned 8) need to inform the used device to collect data. In addition, individual’s privacy must be taken into consideration.

7. REFERENCES


