IoT-Detective: Analyzing IoT Data Under Differential Privacy

Sameera Ghayyur*, Yan Chen**, Roberto Yus*, Ashwin Machanavajjhala**, Michael Hay†,

Gerome Miklau††, Sharad Mehrotra*

* Department of Computer Science, UC Irvine, {sghayyur, ryuspeir}@uci.edu, sharad@ics.uci.edu

** Department of Computer Science, Duke University, {yanchen, ashwin}@cs.duke.edu

† Department of Computer Science, Colgate University, mhay@colgate.edu

†† College of Information and Computer Sciences, UMass Amherst, miklau@cs.umass.edu

ABSTRACT

The success of emerging IoT applications depends on integrating privacy protections into the IoT infrastructures to guard against privacy risks posed by sensor-based continuous monitoring of individuals and their activities. This demonstration adapts a recently-proposed system, PeGaSus [2], which releases streaming data under the formal guarantee of differential privacy, with a state-of-the-art IoT testbed (TIPPERS [9]) located at UC Irvine. PeGaSus protects individuals' data by introducing distortion into the output stream. While PeGaSuS has been shown to offer lower numerical error compared to competing methods, assessing the usefulness of the output is application dependent.

The goal of the demonstration is to assess the usefulness of private streaming data in a real-world IoT application setting. The demo consists of a game, IoT-Detective, in which participants carry out visual data analysis tasks on private data streams, earning points when they achieve results similar to those on the true data stream. The demo will educate participants about the impact of privacy mechanisms on IoT data while at the same time generate insights into privacy-utility trade-offs in IoT applications.

CCS CONCEPTS

• **Security and privacy → Usability in security and privacy**;

KEYWORDS

Differential Privacy; PeGaSus; Tippers; IOT

ACM Reference Format:

Sameera Ghayyur*, Yan Chen**, Roberto Yus*, Ashwin Machanavajjhala**, Michael Hay†, Gerome Miklau††, Sharad Mehrotra* . 2018. IoT-Detective: Analyzing IoT Data Under Differential Privacy. In *SIGMOD'18: 2018 International Conference on Management of Data, June 10–15, 2018, Houston, TX, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3183713.3193571

SIGMOD'18, June 10–15, 2018, Houston, TX, USA

ACM ISBN 978-1-4503-4703-7/18/06…$15.00

https://doi.org/10.1145/3183713.3193571

1 INTRODUCTION

Emerging Internet of Things (IoT) technologies [5, 8] promise to revolutionize domains like health, transportation, smart buildings, smart infrastructure, and emergency response. IoT has the potential to connect a large number of commodity devices (e.g., sensors, actuators, controllers) into an integrated network that can empower systems with new capabilities and bring transformational improvements to existing systems [5, 7]. In IoT systems, sensors are used for fine-grained monitoring of the evolving state of the infrastructure and the environment. Our interest is in user-centric IoT spaces (as per the IEEE P2413 standard [6]) wherein sensors of diverse types (e.g., cameras, cell phones, WiFi access points (APs), beacons) are used to create awareness about subjects/end-users, their interactions with one another, and with the space.

While fine-grained continuous monitoring offers numerous benefits, it raises several privacy concerns [1, 4, 10]. To appreciate such concerns, consider smart buildings, such as smart office spaces and/or smart retail spaces, that track individuals' location and activity to provide customized experience based on user's context. Such services could include customized HVAC control based on user's preference, help locating nearby resources, and/or customized coupons/incentives in a retail setting. Fine-grained monitoring, besides enabling customized services, also raises significant concerns about the data collector being able to use the data captured to infer properties such as religious beliefs, gender, personal habits of individuals (e.g., smoker/non-smoker), among others, which individuals may not be comfortable sharing without explicit consent. Our own experience in developing TIPPERS [9][1] shows that low-level sensor data captured by WiFi APs, motion/light sensors can allow for inferences about individuals, their locations, and their work habits.

We assume that the IoT infrastructure is trusted but that privacy violations for monitored individuals may result from the release of collected data through the many applications envisioned for IoT. The privacy literature has shown that serious disclosures can result even when data is anonymized or the released data consists of aggregate statistics about groups of individuals. We use differential privacy [3], with appropriate privacy parameters, to offer protection to individuals whenever data is released beyond the trust boundary of the IoT system.

Our goal in this demo is to explore privacy-utility trade offs offered by methods such as PeGaSus [2] (which provides differential privacy guarantees over streaming data) in supporting real-world

[1] TIPPERS is a smart building infrastructure being built at UC Irvine.

applications for everyday use in a real IoT testbed. TIPPERS uses diverse sensor data to generate a dynamic state of the building and its occupants—in particular, sensors such as WiFI APs, video cameras, and bluetooth beacons are used to determine the location of individuals in the 6-story Computer Science building as a function of time. Such location data is used, in turn, to create a variety of applications (described briefly later) used by building occupants and visitors. PeGaSus, and other privacy mechanisms which introduce noise into released aggregates, are currently being integrated into TIPPERS to offer rigorous privacy protections.

In the proposed demonstration, we focus on the *Building Analytics* application built into the TIPPERS testbed. The application offers end-users an ability to monitor occupancy levels at various granularities (e.g., room, floor, region) and types (e.g., faculty offices, student spaces, conference rooms, meeting rooms, lounge spaces). Historical data can be analyzed at various temporal granularities (minutes, hours, days).

Motivated by the tasks for which the *Building Analytics* application is typically used, we have created a game, IoT-Detective, in which a player is asked to perform one (or more) interactive analytics tasks using a visual analytics tool based on private streams. These include identifying high-occupancy regions, finding unresponsive sensors, or counting the number of times occupancy exceeded a threshold. As part of the game, users are offered differentially private views of the data and are rewarded for both their accuracy and timeliness in finishing the task. In addition to the game, the demonstration will also include a tour of the underlying technology used in TIPPERS to determine location data based on diverse sensors as well as the PeGaSus algorithm that privately answers continuous queries over real-time data streams.

2 PRELIMINARIES

We briefly describe our IoT testbed, called TIPPERS, and the differentially private engine for releasing streaming data, PeGaSus.

2.1 TIPPERS

TIPPERS (**T**estbed for **I**oT-based **P**rivacy-preserving **PER**vasive **S**paces) is an experimental 6-story smart building testbed designed to study the numerous privacy challenges that result from fine-grained monitoring of building occupants and visitors using a diverse set of sensors [9]. To date, TIPPERS has installed 40 cameras, 64 WiFi APs, several hundred bluetooth beacons covering all major regions in the building, over a hundred smart plug meters to monitor energy consumption of connected devices, over six thousand HVAC sensors measuring airflow and ventilation as well as temperature at different parts of the building, and a large number of light and motion sensors. Data from these sensors flows through the TIPPERS system that fuses the underlying sensor data to produce mainly two higher-level data streams – PRESENCE, which monitors location of all individuals who are inside the building as a function of time, and ENERGY, which monitors energy usage at different spatial resolutions. The information managed by the TIPPERS database system is used to build a variety of applications from real-time awareness of resources, people, and events, to mechanisms to perform analytics on historical data.

The focus of the demo is on the *Building Analytics App*, shown in Figure 1. This app provides analytics about data gathered from multiple sensors in the building (e.g., occupancy, temperature, and energy consumption). The user can view occupancy data for different time intervals and space granularities. The application is designed to gain an understanding of how the building is used as a function of time in order to better plan spaces and events, as well as to better control HVAC systems in order to be more energy efficient. For instance, patterns of building usage by occupants for different regions of the building could lead to customized HVAC settings that save energy without inconveniencing occupants. Likewise, occupancy data can also be used to determine if there are regions in the building that are under/over utilized and such information can lead to plans for better space management (e.g., understanding class rooms that are overflowing or underflowing or determining which lounge spaces are popular). The tasks we choose for our experimental game described as part of this demo are motivated by such real world needs of building analysts. For the context of this demo, the main focus is on occupancy data which is derived from PRESENCE data stream. The PRESENCE data stream has continuously been collected now for about two years, resulting in about 300 million location events since January 2016.

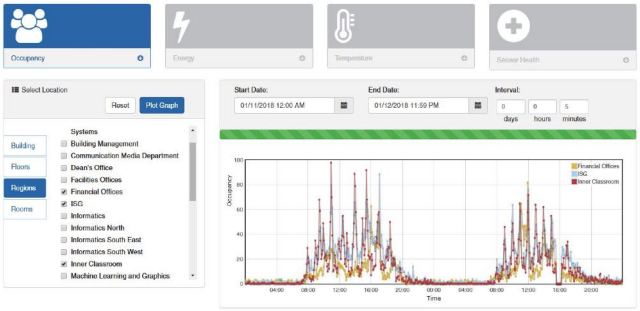

Figure 1: Screenshot of the Building Analytics app.

2.2 PeGaSus

PeGaSus is a novel system for releasing continuous query answers on real time streams under differential privacy [2]. PeGaSus assumes the input has been pre-processed into a stream of tuples (u, s, t) meaning user u was observed in state s at *logical* time t. The logical timestep captures a short window of time (e.g., 5 minutes). States correspond to events of the form "user u connected to a specific WiFi AP." Pre-processing ensures that, at each time t, a user can be in at most m states for some fixed and known m.

PeGaSus supports a variety of continuous queries over the data stream. The most basic query is the *unit counting query*, which corresponds to releasing the number of users in a given target state at each time point. It supports other queries over a single target state such as sliding window sum queries—which report aggregated counts over time windows—and event monitoring queries—which report whether or not a specific temporal event occurred (e.g., the number of connections exceeding a threshold). PeGaSus also supports queries over *multiple* target states (e.g., monitoring individual loads on each access point), and *aggregations* over states (e.g.,

monitoring loads aggregated over all access points on a floor of a building).

PeGaSus ensures event-differential privacy. Informally, this means that modifying the stream by adding or removing (up to m) tuples from a single user u at a single logical time t does not significantly change the output (quantified by privacy loss parameter ϵ). We refer the reader to the full paper [2] for a formal privacy statement, and its implications.

PeGaSus consists of three modules: a *Perturber*, which generates a stream of noisy counts, a *Grouper*, which privately partitions the stream into contiguous regions that have roughly uniform counts and a query specific *Smoother*, which combines the output of the *Perturber* and *Grouper* to generate the final estimate of the query answer at each time step. Only the *Perturber* and *Grouper* access the sensitive stream.

3 DEMO OVERVIEW

The demo consists of two parts. The first part is IoT-Detective, a game where the demo participant plays the role of the building analyst and uses a tool similar to the Building Analytics app to explore the differentially private data and perform various analysis tasks. The objective of the game is to perform analysis as accurately as possible and achieve the highest score across all attendees. The second part of the demo is a brief, behind-the-scenes tour of the underlying technologies (TIPPERS and PeGaSus).

The *target group* of this demonstration is the conference attendees. The players do not need to have any prior knowledge of differential privacy.

3.1 The IoT-Detective Game

The first part of the demo will be presented as a game where a player — in this case, a SIGMOD attendee — is challenged to identify a real world event or pattern using tools provided by TIPPERS on the differentially private data, much like a building manager might in a real-world deployment. The purpose of the game is two-fold: 1) To illustrate the privacy-utility tradeoffs in the differentially private data generated by PeGaSus in a way that engages SIGMOD attendees; 2) User-test this tool for a future study of whether users can use differentially private data for IoT analytics.

To play the game, the demo participant interacts with the IoT-Detective game (see Figure 2), which is very similar to the Building Analytics app, but has some additional game-specific features, such as a timer, leader board, etc. The game is played in rounds and a player can play as many rounds as possible in the allotted time. In each round, the player is given a specific task which requires answering a factual question about types of events during certain time periods (e.g., to identify the most likely time a weekly meeting occurs). The player can then use the app to navigate through the data to identify the relevant (differentially private) data streams and temporal windows and derive an estimate for the answer. The accuracy of the answer is measured in terms of the difference between the player's estimate and the correct answer on the true (non-differentially private) data. Players will be rewarded with points after accurately accomplishing each task. The amount of points will depend on a combination of the accuracy of their estimate, the time taken to complete the task, and the number of tasks they have completed (to incentivize participants to play more than one round). The demo will track player points and maintain a leader board to encourage friendly competition.

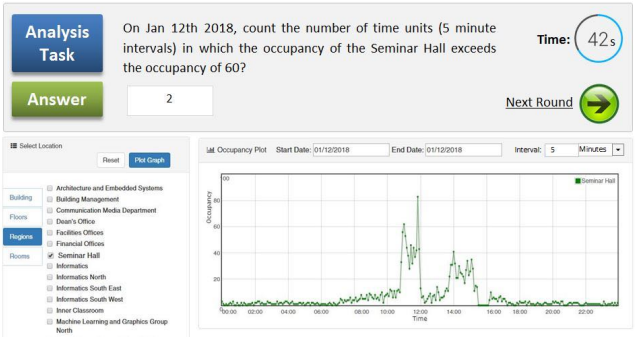

Figure 2: Screenshot of the IoT-Detective game interface.

The accuracy of a player's answer depends on two primary factors. First, it depends on the player's ability to successfully navigate the user interface—thus, the demo is serving as a valuable user test to see if the tool is intuitive and effective for these analytics tasks. Second, it depends on the amount of noise injected into the data stream by PeGaSus. By varying the privacy parameters across users and rounds, we can gather some preliminary data on how much noise is tolerable for varying tasks—thus exploring the practical viability of differential privacy in streaming data settings.

Example Tasks. The Building Analytics Game app will be initialized with a differentially private dataset that reports occupancy information at 5 minute intervals for each room in the building.

An example of a task might be: "On [specific date], count the number of time units (5 minute intervals) in which the occupancy of the [main conference room] exceeds [60]." The parts in brackets can be varied to generate different versions of this task. The motivation for this task is that building managers may wish to detect when a room exceeds its maximum permitted occupancy under fire code regulations, or identify rooms/times in which space is heavily-utilized. Players will be asked to perform a variety of tasks. The following are additional illustrative examples:

- *High occupancy regions.* The rooms can be naturally organized into a fixed set of regions e.g., Facilities Offices, Department of Informatics, etc. This task is to identify which region is the most occupied at night (6pm to 6am) on [a particular day]. Most occupied could mean average number of people are highest during night time. The motivation for this task is better HVAC control at late hours when there are fewer occupants in the building. The accuracy measure can be the difference in rank between the user's choice and the true answer.
- *"Broken" sensors.* We presume here that when a sensor breaks, it no longer senses its environment and continuously reports a constant value, such as zero. Thus, we formulate the task as follows: identify the earliest point in time in which [a particular sensor] starts continuously reporting zero. This is motivated by the practical challenges that building managers face with equipment maintenance. The accuracy measure is distance to

the actual time the sensor breaks (we will artificially modify the dataset to make a sensor appear broken).

- *Occupancy at routine events.* The task is to identify the start time of a regularly occurring event in a particular room e.g. start time of a lecture in a classroom. The motivation is to facilitate better scheduling or detecting events that deviate from a schedule. The accuracy measure is the distance between the player's estimate and the actual start time of the event.

Post-demo empirical evaluation. The demo system will record traces of the games of all participants. In addition to providing immediate feedback to users on their success, we intend to analyze the complete trace to better understand the impact of the privacy mechanism on the usefulness of visually displayed stream data. The trace of game play will allow us to answer questions such as: *For what setting of the privacy loss parameter (ϵ) is task success negligibly impacted? For what settings of ϵ does task success break down? To what extent does task success vary across the user population (e.g. due to differences in skill or attention)? How do the above vary across different tasks? (e.g. are some tasks more tolerant of distortion in the data or of poorly skilled players?) Is there systematic bias in task answers that results from the perturbed data?* Each one of these factors is crucial to a successful integration of PeGaSuS in a real IoT system, and will provide insight into feasible privacy settings and improvements to privacy mechanisms. Although a conference demo is inappropriate as the basis of a formal, controlled user study, the experience of gathering results from the SIGMOD audience will help us design and tune a rigorous study in the future.

3.2 Technology Tour

In the second part of the demo, we will take the participant on a behind-the-scenes tour of the technology underlying sensor data processing in TIPPERS and the differentially private algorithms underlying PeGaSus.

TIPPERS. In TIPPERS, the location of individuals is determined dynamically based on several lower level data sources including connectivity to WiFi APs, presence of an individual in a video feed, WiFi fingerprinting, as well as connectivity to different bluetooth beacons[2]. Of the different methods for localizing individuals within the building, using WiFi APs to track client connections is the most useful because (a) it is ubiquitous, since wireless network covers the entire building, and (b) it does not require active participation of, or any software to run on the client machine. One of the key challenges, however, in using such dataset is the relatively coarse granularity (region level) compared to the much finer granularity that can be achieved using beacons and/or cameras. While coarse granularity data suffices for certain applications/analysis tasks (e.g., understanding region level occupancy of the building), for other tasks needing finer granularity (e.g., at the room level) additional mechanisms for localization need to be designed. One of the mechanisms explored in TIPPERS is to postulate the finer granularity localization, viz., room level, as a data cleaning challenge. Additional information such as location of occupants office, calendar entries, data collected over time to observe patterns in the location of individuals, as well as fine granularity location data collected sporadically using other sensors such as beacons placed in some locations, is used within TIPPERS to develop models for fine grained localization using WiFi AP datasets. The demonstration will enable participants to gain insight into the effectiveness of such mechanisms in improving the quality of location data using WiFi APs.

PeGaSus. The demo participant will see a visualization of the various steps in the differentially private stream generation – generation of events (u, s, t) from the TIPPERS data, specifying the privacy object (the unit that an attacker should not learn about), the noisy stream output by the *Perturber*, the contiguous regions of uniformity identified by the *Grouper*, and the final output generated by the *Smoother*. We will also highlight how the choice of the privacy parameter ϵ and the privacy object impact the intermediate and final outputs. We hope this part of the demo will educate participants about the impact of privacy mechanisms on IoT data.

[2]TIPPERS also allows programmers to specify additional "virtual sensors" to transform lower-level sensor data into higher-level observations about presence of individuals.

ACKNOWLEDGMENTS

This material is based on research sponsored by DARPA under agreement number FA8750-16-2-0021 and N66001-15-C-4067 and the NSF under grants 1253327, 1408982, 1409125, 1443014, 1421325, and 1409143. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

REFERENCES

[1] Stephen Kwamena Aikins. 2016. Connectivity of Smart Devices: Addressing the Security Challenges of the Internet of Things. In *Connectivity Frameworks for Smart Devices: The Internet of Things from a Distributed Computing Perspective.*
[2] Yan Chen, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau. 2017. PeGaSus: Data-Adaptive Differentially Private Stream Processing. In *ACM Conf. on Computer and Communications Security (CCS).* 1375–1388.
[3] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *3rd Theory of Cryptography Conference.* 265–284.
[4] Muhammad Umar Farooq, Muhammad Waseem, Anjum Khairi, and Sadia Mazhar. 2015. A critical analysis on the security concerns of Internet of Things (IoT). *Int. Journal of Computer Applications* 111, 7 (2015).
[5] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation of Computer Systems* 7 (2013), 1645–1660.
[6] IEEE. 2015. P2413 IoT standard. https://standards.ieee.org/develop/project/2413.html. (2015).
[7] Sye Loong Keoh, Sandeep S. Kumar, and Hannes Tschofenig. 2014. Securing the Internet of Things: A Standardization Perspective. *IEEE Internet of Things Journal* 1, 3 (2014), 265–275.
[8] Somayya Madakam and Hema Date. 2016. Security Mechanisms for Connectivity of Smart Devices in the Internet of Things. In *Connectivity Frameworks for Smart Devices: The Internet of Things from a Distributed Computing Perspective.*
[9] Sharad Mehrotra, Alfred Kobsa, Nalini Venkatasubramanian, and Siva Raj Rajagopalan. 2016. TIPPERS: A privacy cognizant IoT environment. In *IEEE Int. Conf. on Pervasive Computing and Communication Workshops (PerCom Workshops).* 1–6.
[10] Kai Zhao and Lina Ge. 2013. A survey on the Internet of Things security. In *9th Int. Conf. on Computational Intelligence and Security (CIS).* 663–667.