Evaluating Practical Privacy Attacks for Building Data Anonymized by Standard Methods

Schwee, Jens Hjort; Sangogboye, Fisayo Caleb; Kjærgaard, Mikkel Baun

Publication date:
2019

Document version
Final published version

Citation for published version (APA):
Evaluating Practical Privacy Attacks for Building Data Anonymized by Standard Methods

Jens Hjort Schwee  
jehs@mmmi.sdu.dk  
University of Southern Denmark  
Odense, Denmark

Fisayo Caleb Sangogboye  
fsan@mmmi.sdu.dk  
University of Southern Denmark  
Odense, Denmark

Mikkel Baun Kjærgaard  
mbkj@mmmi.sdu.dk  
University of Southern Denmark  
Odense, Denmark

ABSTRACT
The availability of inexpensive IoT sensors enables the collection of a wide range of data about buildings and their use by occupants. The data originating from these sensors are in many cases privacy-invasive. Therefore, methods for improving privacy protection for such data is needed. In this paper, we explore if suppression, \( k \)-anonymity or a combination of them, provides sufficient privacy protection on a published dataset. The results show that the specific dataset only was partly protected in the case of the combination of the two privacy protection methods. The attack vectors used to break the protection includes the sensor deployments, and the physically limitations of the sensors. This indicates, the need to consider how to protect the metadata of sensor deployments as well as the sensor streams. This additional metadata protection can serve as requirements, for anonymization methods for time-series data. However, even with good privacy protection, the building dataset might still be vulnerable to attacks since building data contains repeatable patterns which are susceptible to a range of attacks.

CCS CONCEPTS
• Security and privacy → Data anonymization and sanitization; Pseudonymity, anonymity and untraceability; Privacy protection.

KEYWORDS
Data privacy, Pseudonymity, Data anonymization and sanitization, \( k \)-anonymity, Privacy-protecting data publishing, Linkage attacks.

ACM Reference Format:

1 INTRODUCTION
The recent advances and proliferation of pervasive sensors are facilitating the gathering of rich contextual monitoring in built environments. This monitoring enables the optimization of building use, by collecting information about how the occupants are using it, including presence and actions. Sharing such data could reveal private breach about the occupants, which could be used by adversaries for unintended purpose.

The possibility of such privacy information has resulted in several governmental legislation and research activities in the area of privacy-protecting data publishing (PPDP). Governmental policies such as [5] mandate data publishers to ensure full disclosure of the intent of data collection, the collection process, and the protection of the sensitive attributes. PPDP methods for protecting sensitive attributes focus on the concept of preserving data quality while ensuring privacy protection [12]. As an initial step, data publishers utilizes PPDP methods for suppressing the unique identifiers or attributes of data owners in the published dataset. This initial suppression process serves as a state-of-practice for publishing building-related datasets [3]. While this suppression can guarantee high usability of the dataset, it is vulnerable to privacy breach by adversaries with additionally knowledge. Jia et al. [7] demonstrated that with prior knowledge of a data owners pattern, records could easily be re-identified in a published dataset.

In this paper, we evaluate the privacy protection provided using suppression, \( k \)-anonymity and the combination of the two, on a public buildings occupant dataset [3]. Also, we explore several of the vulnerabilities, using record linkage and probabilistic attacks. The results obtained from this work, demonstrates that more knowledge is needed on what attacks are possible on building data, because most building data has temporal repeatable patterns and physical laws limits possible sensor values. Lastly with a case study, we provide insights that provides a new set of requirements for designing and developing data anonymization algorithms for time-series.

2 RELATED WORK
Fung et al. [6], have produced a categorization of sensitive attributes in a dataset which includes: Explicit identifier (EID), quasi-identifier (QID), sensitive attributes (SA), and non-sensitive attributes (NSA). When publishing data, only the NSA part of the data can be shared in its original format. The remaining needs to be anonymized or even removed.

Researchers have defined a number of attacks which can be performed on data streams. In record linkage, the attacker is trying to identify a small number of individuals by using the sanitized attributes, typically the QID’s, and joining them with a related table. \( k \)-anonymity [13] is a privacy model which is satisfied if at least \( k - 1 \) other records have the same set of QID’s in the output. Other privacy models have been developed such as \( l \)-diversity [8], \( \delta \)-presence [9], \( \epsilon \)-differential privacy [4]. These are popular privacy models for protecting against attribute linkage, table linkage and probabilistic attacks, respectively. While these privacy models are conceptually attack-prove to these forms of attacks, it is uncertain, what parameter specification for each privacy model can suffice for protecting the privacy of occupancy-related dataset in buildings.

Jia et al. [7, 11] proposed PAD, an implementation of \( k \)-anonymity, that sanitizes building datasets and protects against record linkage. The unique feature of PAD, unlike traditional \( k \)-anonymous privacy models, is that it incorporates the interests of a data analyst using a published dataset in the publication process. Incorporating
A significant part of the work in the field have been focused on the protection of sensitive databases with tabular data and typically with a defined set of queries. In this paper, we consider time-series streams without the queries limitation. We consider a subset of the practical attacks possible for time-series data, to explore a new set of privacy requirements for this type of data.

3 PRACTICAL PRIVACY ATTACKS

For the study, we have considered a sanitized public building dataset [3]. This dataset includes high fidelity streams from three types of sensor modalities covering four rooms in a public building. The published data have been collected in a university teaching building. Arendt et al. [3] indicates that the data have been obscured by suppressing the actual room name and by changing the order of the days, remove the mapping to a specific day. The metadata states that the rooms are two classrooms and two study zones with seating-capacities of 84 and 32 respectively. The data has been obtained between in the period 2017-03-22 to 2017-04-05. The dataset has three modalities of sensor streams include CO\textsubscript{2} (in ppm), occupants counts (in counts) and the damper position (in percent) of the VAV for the four rooms. Each reading has the flowing metadata: (1) **DayId**, a number of the day, randomly generated identifier. (2) **Time**, the time of day, in local time. (3) **Weekday**, number of the weekday. (4) **Holiday**, number indicate if it is a national holiday.

Given this data, our goal is to carry out privacy attacks to re-identify the following information: (1) The original sequence of the days. (2) Information about the activities in the rooms. (3) Identifying the actual names of the rooms.

Next, we will present the different kinds of attacks for identifying the above information using the published dataset, and other publicly accessible information from several sources about the building.

3.1 Identify the Correct Sequence

In the published dataset, the original sequence of the daily patterns of streams has been re-ordered, for protecting any inherent patterns that may reveal private information of people or activities in the four rooms. Given that the first and last days of the released dataset are 22nd March and 4th of April 2017, which both are wednesdays, this information can be used to find the original structure of the days. Furthermore, an attacker knows that the streams are captured in a sequence. Given this, an attacker can specify a threshold for each sensor stream and examine the first and last readings of each day in the streams to detect the readings of possible prior and posterior days. The catch, however, is that, this threshold parameter may need to be tuned for each room. Secondly, a probabilistic approach could be adopted where the likelihood of prior and posterior days are computed to detect the most probable sequence of days.

First we will evaluate if we can identify the order of the days, using the threshold-based approach, with fixed thresholds between 0-10. These thresholds could not identify any unique sequence for any of the sensor streams, when using all of the rooms combined. Subsequently, considering rooms and streams individually. Shows that the privacy attack was not successful for the VAV or the occupancy count streams. However, for the CO\textsubscript{2} stream, room 2 has a unique pattern using a threshold between 0-4 and room 3 have for a threshold of 6, all of these provide the same order of the days, hence the attack was successful.

3.2 Finding the Identity of the Rooms

Given the success of the previous attack, we utilize the identified order of the data to attack the metadata about the room locations. In [3], the authors state that the classrooms poses the most privacy risk since they can be used for evaluating the performance of the lecturer. Therefore, we will focus our attack on these rooms' identity, which are room 1 and 4. The deployment of the sensors in the building can be found in [10]. This highlights that there are only four rooms in the building which have the published sensor setup, two classrooms, and two study zones. Combining this information with the university map services [1], we can identify the two classrooms as U180 and U182.

The historical and future activities for all of the rooms are available from TimeEdit [2]. Exploring the schedules for the classrooms in the data period, it can be observed that on Tuesday room U182 have an activity from 16:00-20:00, where U180 does not, creating two options for attacking the occupant patterns, on the 28 of Match and the 4 of April. Considering the activity on Tuesdays the 28th, in room 1, there was occupants in the period where in room 4 there was not. Hence the room 1 is U182 and room 4 is U180.

3.3 Record Linkages with Public Data

In this section, we will investigate if the dataset combined with the external data, can identify some activities in each of the rooms. Firstly, identifying the start and end of each scheduled activity. Furthermore, use the occupant patterns to identify breaks throughout the activities. We estimate the start of activities by finding the maximum number of participants for a particular activity. Furthermore, we hypothesize that the instance where there exist at least 50% of this number, the event should have started. Conversely, we hypothesize the end as when the number of participants for the activity goes below 50%. Breaks are specified as a period longer than 5 minutes where the occupant count is lower then 50% and then goes back above 50%.

Figure 1a shows the occupancy count obtained from room 1 on March 27. On this day the room had three lectures, where the first lecture occurred from 8:00-10:00, the second from 10:00-12:00 and the last from 12:00-16:00. Visually it can be observed that some of the occupants arrived late for the first lecture of the day, however, this is not captured using the 50% occupants methods, which estimates the start at 8:00. During this activity, there is a drop in the count of occupants in the middle of the lecture, this could be a break in the lecture, but is not found by the break detection. The activity starting at 10:00, has a less stable occupants pattern than the other lectures. Indicating by more variance in the number of attendance. Furthermore, it can be observed that the lecture ended earlier. The 50% occupant classifiers detected the start of the class to be 10:00, a break in the class from 11:36 to 11:46, and the end time to be 12:00. The last scheduled activity is a four-hour lecture, visually it can be observed that some occupants arrived late.
break was given and that the lecture ended before time. The 50% method finds the lecture to have started on time, a break from 13:24 to 13:31, and estimates the class to have ended at 15:30.

4 PROTECTING THE DATASET

In this section, we will protect the original unsanitized streams and the resulting sanitized streams presented in [3]. Subsequently, we apply k-anonymity for protecting both datasets. This provides a k-anonymized dataset for the original unsanitized streams and a combination of suppressed and k-anonymized dataset for the sanitized streams. We are going to explore which level of k is necessary for protecting the data against the presented attacks. For both of the experiments, we are going to use k-anonymity implementation presented in PAD [7] and vary the k values from 2 to 7. Because PAD presents a customizable way for microaggregating a sanitized dataset, we have adopted the generic microaggregation method for sanitizing the daily profiles of each of the sensor streams.

Privacy Attack on an only k-anonymized Dataset Using the original k-anonymized streams, with date and time-stamps, we will try to identifying the identity of the classrooms. To carryout the attack, we are going to use the periods where only one of the rooms had scheduled activities, namely the Tuesdays. With all of the configurations of k’s, we were able to identify the rooms. For room 4, there was approximately 10% of the occupants in room 1, indicating that room 1 is room U182. See Table 1, for an overview of the successful attacks for these datasets.

Figures 1b and 1c highlights the k-anonymized occupant streams for a particular day in room 1 with k values of 2 and 7 respectively. For the k-anonymized occupant streams with, the k of 2, we used the same attack method to detects the start, end, and breaks of an activity. The first activity is estimated to start at 8:02 and end at 10:00. The second activity is estimated to start at 10:00, end at 12:00, and to have a break from 11:19 to 11:59. This break is substantially longer than the one in the original data. From Figures 1a and 1b, it can be observed that the drop of occupants is less steep in the k-anonymized version and that the minimum of occupants is low. The long break is estimated since the maximum amount of occupants in the period is 45. Therefore, is the threshold set to 22, which is higher than the presumably amount at 30. Hence, the longer break estimation. The final activity of the day is estimated to start at 12:01, and end at 13:51. The end time is estimated to be much earlier than using the original, due to the maximum amount of occupants in the room is found to be 58 and for a while just after 13:51, there is 20-28 occupants in the room. This indicates that the threshold needs to be optimized for the sanitized data. Furthermore, the method does not find the break doing the lecture, due to the break in this version, being only 4 minutes long. Considering the k-anonymized data with a k of 7, all of the estimations using the 50% threshold are the same as the k of 2 version. However, looking into the time-series plot of Figure 1c, it is much harder to visually see a break in the lectures, especially doing the last lecture of the day. This is expected since the data contains a more general trend then the k of 2 version.

For considering the usefulness of the occupant data, we have calculated the normalized root-mean-square-error (NRMSE) using the max setting capacity of the individual rooms, as the max in a min-max normalization before calculating the RMSE, the results can be found in Table 1. The results shows that the higher level of k, the higher NRMSE. This indicates that the higher level of k the less useful is the sanitized occupant data.

Privacy Attack on k-anonymized and Suppressed Dataset In the this section, we will explore how to attack the released data which have been anonymized with k-anonymity, with k’s from 2 to 7. We will like to identify the sequence of the days, enabling the record linkages attacks. First, we will attempt identification using the threshold based attack. This method does not provide a unique sequence of days using any of the sensor streams and any of the rooms. This is likely due to the streams within a group is indistinguishable, and therefore there is not a unique pattern using a fixed threshold. Instead we consider a probabilistic approach, estimating the order based on likelihood. Using this attack, we are able to find the first day in the data, which is the Wednesday the 22 of Match. We will estimate the first day to be the Wednesday which has the maximum difference between any Tuesday. Next, we will use the identified day to estimate the next, based on the minimum difference between the candidates for the next day, in all of the rooms using the CO2 streams, since it was found to be the best attack vector in the previous attacks. In all cases but the data which had a k of 7, we are able to find the right first day. However, we are not able to find the right order in any of these anonymized versions of the data. We can identify the next day in 10 of the 14 cases using the anonymized data with k’s of 2-6. In the case of k of 7, can identify seven next days. See Table 1 for summary of the identified values. Furthermore, in all cases of the anonymized data, we are able to identify the two teaching rooms using the same method as previously.

Figure 1: Time-series for the occupancy in room 1 on the day with DayId 13. The red lines are the start and end time of schedule classes in the room for the given day.
Table 1: Summary of which attacks were successful using the various datasets and the usefulness of the occupant data in normalized root-mean-square-error (NRMSE). Check marks are successful attacks and X’s are unsuccessful attacks.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>First Day / Last Day</th>
<th>Amount of Next Days</th>
<th>Occupant Attendance Error (NRMSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Released dataset</td>
<td>✓</td>
<td>✓</td>
<td>Same score as for the released with the same k</td>
</tr>
<tr>
<td>k-k-anonymized</td>
<td>✓</td>
<td>✓</td>
<td>10 0.0544</td>
</tr>
<tr>
<td>with k of 2-7</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Released dataset</td>
<td>✓</td>
<td>✓</td>
<td>10 0.0544</td>
</tr>
<tr>
<td>k-anonymized</td>
<td>✓</td>
<td>✓</td>
<td>10 0.0811</td>
</tr>
<tr>
<td>with k of 3</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Released dataset</td>
<td>✓</td>
<td>✓</td>
<td>10 0.0931</td>
</tr>
<tr>
<td>k-anonymized</td>
<td>✓</td>
<td>✓</td>
<td>10 0.0959</td>
</tr>
<tr>
<td>with k of 4</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Released dataset</td>
<td>✓</td>
<td>✓</td>
<td>10 0.1012</td>
</tr>
<tr>
<td>k-anonymized</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>with k of 6</td>
<td>✓</td>
<td>✓</td>
<td>7 0.1054</td>
</tr>
<tr>
<td>Released dataset</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

The privacy model of $k$-anonymity assumes that each record is an individual. In our evaluation, we are using daily profiles as the input for the method. Therefore, the privacy guarantee of the model is not necessarily complied, as showcased in this paper. However, the evaluation shows that the higher values of $k$, the harder to attack the streams, especially when combined with suppression. Using the model of $l$-diversity might be a good candidate to explore since it protects against attribute linkage attacks, and since an adversary might have prior knowledge about the building. Furthermore, we need to define what to be grouped in the model, and the daily profiles, as explored in this paper, might not be the best choice.

As part of the process of publishing a dataset, a data publisher needs to consider what information is publicly accessible since this data is to be considered for selecting the privacy model. Furthermore, considering what part of the streams can be considered as sensitive and what prediction algorithms are available for the released streams. These steps are a challenge for a data publisher, and therefore it would be good if there was an automatic attack suite to perform a number of attacks on datasets before sharing with externals or publishing. This attack suite should perform attacks which are typical vulnerabilities of datasets, including record linkage and the probabilistic attacks.

ACKNOWLEDGMENTS

This work is part of coordination and participation in IEA EBC Annex 79 supported by EUDP: 64018-0558 and HBDEx supported by the SDU ODE initiative.

REFERENCES