# Privacy Interpretation of Behavioural-based Anomaly Detection Approaches

Muhammad Imran Khan[†], Simon N. Foley[‡], and Barry O'Sullivan[†]

[†]Insight Centre for Data Analytics, School of Computer Science and Information Technology, University College Cork, Ireland.
[‡]Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway.

**Abstract**

This paper proposes the notion of '*Privacy-Anomaly Detection*' and considers the question of whether behavioural-based anomaly detection approaches can have a privacy semantic interpretation and whether the detected anomalies can be related to the conventional (formal) definitions of privacy semantics such as $k$-anonymity. The idea is to learn user's past querying behaviour in terms of privacy and then identifying deviations from past behaviour in order to detect privacy violations. Privacy attacks, violations of formal privacy definition, based on a sequence of SQL queries (query correlations) are also considered in the paper and it is shown that interactive querying settings are vulnerable to privacy attacks based on query sequences. Investigation on whether these types of privacy attacks can potentially manifest themselves as anomalies, specifically as privacy-anomalies was carried out. It is shown that in this paper that behavioural-based anomaly detection approaches have the potential to detect privacy attacks based on query sequences (violation of formal privacy definition) as privacy-anomalies.

## 1 Introduction

The recent past has witnessed an exponential increase in the volume of data being collected by organizations. This has been enabled by the aggressive development of computing technologies. Data is fuelling most of the revolutionary technologies. Technological advances and widely available data have nurtured the area of data analytics.

Data analytics aims to discover meaningful insights from the data that may lead to improved decision-making. Data analytics offers a broad spectrum of benefits, for example, it enables a contemporary organization to anticipate business opportunities as well as enable the delivery of relevant products to its customers. In a nutshell, analytics over a large volume of data has the potential to impact businesses and our society.

Data comes from multiple sources and may include sensitive personal data. On the one hand, one cannot deny the importance and value of data, while on the other hand, the usage, storage, and access to this data can raise privacy concerns. The recently enacted EU's General Data Protection Regulation (GDPR) [1] makes it more challenging to use personal data for analytics. However, once the data is anonymized, it is considered to no longer be personal data [1, 2]. Achieving anonymization is non-trivial while preserving the utility of data. Increasing the level of anonymization protects data but reduces utility. Thus organizations must trade-off the need for more in-depth analytics against the privacy of individual's data. In essence, it is a long-standing open problem to get high-quality analytics by querying the databases consisting of information about individuals while preserving the privacy of those individuals.

Numerous incidents have been reported where privacy was compromised due to poor anonymization of released data, for example, the famous case of Netflix [3], AOL [4], de-anonymization of NYC taxi data [5], and the famous case of the Massachusetts Governor [6]. In [6], it was shown that by linking on shared attributes (zipcode, birth date, and gender) in two datasets, Massachusetts Group Insurance Commission's released data (considered anonymous) and voter rolls, records belonging to the Massachusetts Governor were identified. Researchers have devised formal privacy definitions[1] [7, 8, 9] when these definitions are followed then the anonymized data manifests some formal guarantees. There are several privacy definitions to anonymize data, including, $k$-anonymity, $l$-diversity, $t$-closeness, and differential privacy. The majority of the syntactic privacy definitions were designed for a one-time release of data. In contrast to these definitions, differential privacy is for interactively querying a database. However, differential privacy has practical limitations as well; for instance, differential privacy allows only a limited number of queries to be answered. Allowing an unlimited number of queries results in higher noise; thus, the ability to observe correlations between attributes are lost, which is not desirable for richer analytics [10]. Approaches to detect privacy violations, while allowing an unlimited number of queries while having richer analytical utility are desirable.

Existing work in literature on detecting malicious access (security attacks) to Database

[1]Privacy definitions, in literature, are otherwise known as privacy models, privacy criteria, privacy metrics, privacy constraints as well as privacy principles.

Management System (DBMS) have shown the effectiveness of behaviour-based anomaly detection system to detect these security attacks. This work looks at to what extent these kinds of techniques can be used to detect privacy attacks. This work considers the research question that whether one can provide a privacy semantics for behavioural-based approaches or relate the notion of privacy-anomaly detection to the conventional definitions of privacy semantics? In order to answer the above-mentioned questions, the notion of *'Privacy-Anomaly Detection'*(*PAD*) is introduced in this paper. PAD learns privacy criteria from past interactions (audit logs) and uses this criteria to check whether the current behaviour is different from past behaviour with respect to privacy. The PAD architecture falls within an interactive query system setting for microdata release.

We describe a naïve instantiation of PAD using $k$-anonymity privacy criteria which we refer to as ($k$-anonymity)-PAD. A study is carried out to investigate whether a security-anomaly detection system, in particular, the n-gram approach presented in [11], can detect these ($k$-anonymity)-PAD privacy-anomalies.

In this work, we also show that PAD-based interactive mechanisms are vulnerable to privacy attacks based on SQL query sequences. We further investigate: whether these types of privacy attacks based can potentially manifest themselves as anomalies and whether one can interpret a security-anomaly detection system in such a way that it can detect privacy attacks as privacy-anomalies. We present the result that privacy attacks (like inferences) can be detected by applying security-anomaly detection system over the logs of interactive querying mechanisms on the basis of a PAD interpretation.

The rest of this paper is organized as follows. Section 2 presents a design of a privacy-anomaly detection system and an instantiation based on $k$-anonymity. Section 3 investigates whether there is a correlation between privacy and security anomalies. Section 4 considers a privacy attack based on query sequence on PAD. Section 5 presents an application of security-anomaly detection system to detect (unknown) privacy attacks as privacy-anomalies. Section 6 concludes this paper.

## 2 Privacy-Anomaly Detection (PAD) System

This section introduce the notion of privacy-anomaly detection and present a naïve instantiation of it based on $k$-anonymity. We argue that this naïve instantiation constitutes the basis for a more advanced form of a privacy-anomaly detection system, analogous with $k$-anonymity which constitutes the basis for more sophisticated formal privacy definitions. The reasons are as follows. Firstly, this is an exploratory study to con-

| age | zipcode | county | gender | salary |
|---|---|---|---|---|
| >55 | 989234 | Cork | Male | 60K |
| >55 | 989234 | Cork | Male | 92K |
| >55 | 989234 | Cork | Male | 77K |
| >45 | 839523 | Cork | Male | 50K |
| >35 | 839777 | Dublin | Male | 60K |
| >35 | 839777 | Dublin | Male | 63K |
| >35 | 839777 | Dublin | Male | 85K |
| >35 | 839777 | Dublin | Male | 70K |
| >35 | 839777 | Dublin | Male | 60K |
| >50 | 839567 | Cork | Female | 72K |
| >50 | 839567 | Dublin | Female | 62K |
| >50 | 839567 | Cork | Female | 92K |
| >50 | 839567 | Dublin | Female | 77K |
| >50 | 839567 | Cork | Female | 68K |

Table 1: A fragment of relation temp_table.

sider the question whether one can provide a privacy semantics for behavioural-based anomaly detection approaches or relate the notion of privacy-anomaly detection to the conventional definitions of privacy semantics? Therefore, using a well-understood privacy model like $k$-anonymity enables better understanding of the subject being explored and helps to avoid underlying complexities associated with other more complex privacy definitions. Secondly, $k$-anonymity served as a foundation of many subsequent formal privacy definitions, which is a good indicator of the applicability of this study onto other privacy definitions.

## 2.1 A $k$-Anonymity based Privacy-profile

In the proposed model $k$-anonymity is used to specify a privacy limit $[\![k\ ,\ q]\!]$, whereby $k$ individual must share the same quasi identifier $q$ values in the result of a query. Intuitively, this means for that particular response, for a sufficient value of $k$, an adversary can only narrow down to $k$ individuals. In the case where an adversary has a secondary dataset with overlapping quasi-identifier values, then the query response can be linked to $k$ different individuals, therefore minimizing the risk of re-identification. In the model the privacy-profile is defined as a set of privacy limits. In terms of privacy, each privacy limit means that in a particular instance of a query response an adversary won't be able to distinguish an individual's quasi-identifier values from $k$ individuals for the set of quasi-identifiers that appeared in the query response.

| age | zipcode | salary |
|---|---|---|
| >55 | 989234 | 60K |
| >55 | 989234 | 92K |
| >55 | 989234 | 77K |

Table 2: A relation $\mathcal{T}_{R1}$ resulting from the query `SELECT age, zipcode FROM temp_table WHERE gender = 'Male';`.

Consider a relation `temp_table`, as shown in Table 1, having several attributes including a sensitive attribute `salary`, and quasi-identifiers `age`, `gender`, `zipcode`, and `county`. For ease of exposition we assume the values for attribute `age` are aggregated into age ranges, for instance, all the values for attribute `age` above 55 are represented as >55. Given a mined privacy limit $[\![3, \{\texttt{age}, \texttt{zipcode}\}]\!]$, in privacy-profile, then the response to the analyst query `SELECT age, zipcode FROM temp_table WHERE gender = 'Male' AND county = 'Cork' AND age > 55;` as shown in Table 2 is not anomalous since the value of $k$ for the the quasi-identifiers {`age`, `zipcode`} in the response is greater than 3.

### 2.1.1 Mining $k$-anonymity based Profiles for PAD

The privacy-anomaly detection consists of two phases, similar to traditional anomaly detection approaches, that are, learning phase and a detection phase. The instances of the privacy model are mined from audit logs in order to generate privacy-profiles. We refer to a privacy-profile that is mined from past logs in the learning phase as a normative privacy-profile. The idea is to learn the $k$ values for sets of quasi-identifier(s) by mining past audit logs and interpret those mined 'privacy limits' as 'normal'.

Given an audit log $L^*$, consisting of query responses, $Pri(L^*)$ gives a privacy-profile consisting of privacy limits mined from log $L^*$, where $q \in QI$ represent a set of quasi-identifier. A normative privacy-profile is generated from an anomaly-free past log $L^*_{norm}$ and is denoted by $Pri(L^*_{norm}) = \{ [\![k_1, q_1]\!], [\![k_2, q_2]\!], \ldots, [\![k_m, q_m]\!] \}$. For example, consider the relation $\mathcal{T}_{R2}$ shown in Table 3, the mined value of $k$ for the set of quasi-identifiers {`age, zipcode, county`} is 4, that is, $[\![4, \{age, zipcode, county\}]\!] \in Pri(L^*_{norm})$. In essence we are constructing privacy limit $(L^*, q)$ which returns $k$ as a limit to the privacy in the table for a given $q$. The normative privacy-profile is effectively a set of these privacy limits mined against the logs for a given set of quasi-identifiers. Intuitively, the tuples in the normative privacy-profile shows to what extent one narrows down to individuals records in normative settings.

| age | zipcode | county | salary |
|---|---|---|---|
| >55 | 839523 | Cork | 60K |
| >55 | 839523 | Cork | 92K |
| >55 | 839523 | Cork | 77K |
| >45 | 839523 | Cork | 50K |
| >35 | 839777 | Dublin | 60K |
| >35 | 839777 | Dublin | 63K |
| >35 | 839777 | Dublin | 85K |
| >35 | 839777 | Dublin | 70K |
| >35 | 839777 | Dublin | 60K |

Table 3: A relation $\mathcal{T}_{R2}$ resulting from the query `SELECT age, zipcode, county FROM temp_table WHERE gender = 'male';`.

### 2.1.2 Detecting Privacy-anomalies

The detection phase, in terms of privacy, checks if an adversary is able to narrow down to fewer than $k$ individuals for a given set of quasi-identifiers in the normative profile. In the instance, where the adversary is able to narrow down to fewer than specified $k$ individuals for a given set of quasi-identifier then this instance is labelled as a privacy-anomaly and poses higher risk of re-identification relative to normal. During the detection phase, the run-time profile $Pri(L^*_{run})$ constructed given a run-time log $L^*_{run}$. $Pri(L^*_{run})$ is the constructed run-time profile. Given privacy limits $[\![k_i, q_i]\!]$ and $[\![k_j, q_j]\!]$ then $[\![k_i, q_i]\!]$ subsumes $[\![k_j, q_j]\!]$ (denoted $[\![k_i, q_i]\!] \preceq [\![k_j, q_j]\!]$) if imposing privacy limit $[\![k_j, q_j]\!]$ instead of $[\![k_i, q_i]\!]$ leads to no additional loss of privacy. Formally,

$$[\![k_i, q_i]\!] \preceq [\![k_j, q_j]\!] \equiv q_i \subseteq q_j \wedge k_j \geq k_i$$

In the case where $[\![k_i, q_i]\!] \in Pri(L^*_{norm})$ and $[\![k_j, q_j]\!] \in Pri(L^*_{run})$ then $[\![k_i, q_i]\!] \preceq [\![k_j, q_j]\!]$ means that $[\![k_j, q_j]\!]$ can be safely replaced by $[\![k_i, q_i]\!]$ without any loss of privacy. If a privacy limit subsumes another intuitively it means if the subsumed privacy limit is replaced by the one that subsumes it then there is no loss of privacy.

Consider the response of a query at run-time shown in Table 4, and that there exists a privacy limit $[\![3, \{age, zipcode\}]\!]$ in $Pri(L^*_{norm})$. The mined value $k$ of the set of quasi-identifier $\{age, zipcode\}$ is greater than 3 therefore this privacy limit $[\![5, \{age, zipcode\}]\!]$ in $Pri(L^*_{run})$ is considered to be subsumed by the privacy limit $[\![3, \{age, zipcode\}]\!]$ in $Pri(L^*_{norm})$. In terms of privacy, it means given that this instance of query response an adversary can narrow down so many individuals as one normally is able to for a given set of quasi-identifiers.

| age | zipcode | salary |
|---|---|---|
| >50 | 839567 | 72K |
| >50 | 839567 | 62K |
| >50 | 839567 | 92K |
| >50 | 839567 | 77K |
| >50 | 839567 | 68K |

Table 4: A relation $\mathcal{T}_{R3}$ resulting from the query `SELECT age, zipcode FROM temp_table WHERE gender = 'female';`.

# 3 Security-anomaly Detection System Detecting Privacy-anomalies

This section explores whether privacy-anomalies (as identified by the model in Section 2.1.1) are also identified as security-anomalies by a security-anomaly detection system in [11]). The security-anomaly detection system in [11] relies on n-grams to construct profiles of querying behaviours using audit logs of SQL queries. The system in [11] effectively detects malicious accesses by insider to a database management system. We consider a variation of the hospital dataset, a fragment of the dataset is shown in Table 5. Logs were generated for construction of a normative profile and another for the construction of a run-time profile. The training logs (anomaly-free) for the n-gram based approach are denoted by $L^{hosp}_{norm}$, while the anomalous run-time logs for the hospital datasets are denoted by $L^{hosp}_{run}$. The next section studies whether a security-anomaly detection system detects privacy-anomalies identified by the privacy-anomaly detection system for these dataset.

To construct normative and run-time profiles using the n-gram model, selection of an appropriate value of the size of n-gram was desirable for the hospital dataset. To select an appropriate size of an n-gram in this scenario, test logs $L^{hosp}_{test1}$ and $L^{hosp}_{test2}$ were generated in a safe environment (anomaly-free). N-gram profiles were constructed with varying n-gram size, that are, $ngram(L^{hosp}_{test1}, n)$ and $ngram(L^{hosp}_{test2}, n)$, and generated profiles were compared. Figure 1 depicts the number of n-gram mismatches arising when comparing the normal test $ngram(L^{hosp}_{test1}, n)$ and $ngram(L^{hosp}_{test2}, n)$, for different values of n. From the experiments, the n-gram of the size of 4 ($n = 4$) was considered optimal as it resulted in an acceptable number of mismatches.

Once the value of $n$ was decided upon, the normative and run-time profiles were constructed for the experiments. Given the training logs $L^{hosp}_{norm}$ and $L^{hosp}_{run}$ n-gram profiles were constructed such that $ngram(L^{hosp}_{norm}, 4)$ and $ngram(L^{hosp}_{run}, 4)$, and subsequently the

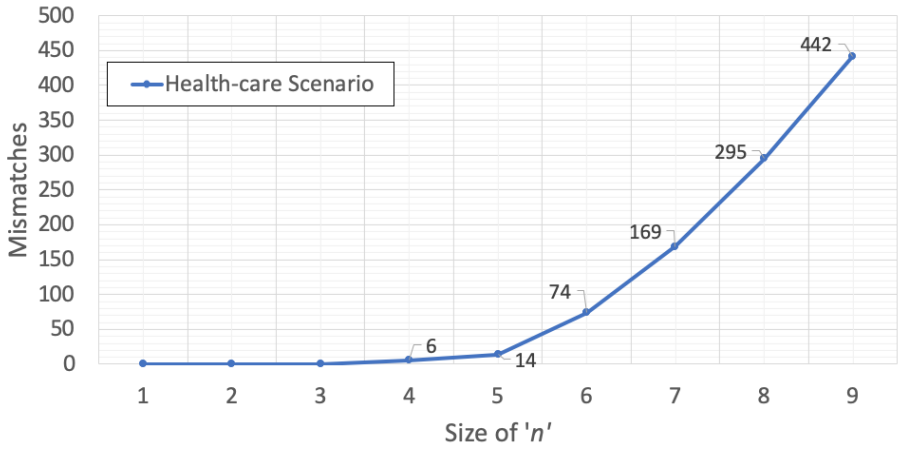


Figure 1: The figure shows the number of mismatches between $ngram(L^{hosp}_{test1}, n)$ and $ngram(L^{hosp}_{test2}, n)$ for different values of $n$.

normative and runtime profiles were compared.

The same queries in logs $L^{hosp}_{norm}$ and $L^{hosp}_{run}$ were executed in the presence of the privacy-anomaly detection system (described in Section 2) resulting in logs of query responses $L^{hosp*}_{norm}$ and $L^{hosp*}_{run}$. Subsequently, a normative privacy-profile $Pri(L^{hosp*}_{norm})$ and a run-time $Pri(L^{hosp*}_{run})$ profiles were constructed and compared.

The attribute `patient_ID` and `e-mail_ID` were considered as a unique identifier, the attribute `diagnosis` was considered as a sensitive attribute while the rest of the attributes including `first_name`, `last_name`, `status`, `dob`, `gender`, `city`, and `marital_status` were considered as quasi-identifiers. For the experimentation, two categories of privacy-anomalies were injected as described in Table 6. Using this anomaly-containing run-time log, from 15 privacy-anomalies 13 were detected by the n-gram based security-anomaly detection system proposed in [11] and the privacy-anomaly detection system proposed in this paper.

## 3.1 Detected Privacy-anomalies

The n-gram based security-anomaly detection system detected all those privacy-anomalies that were generated by injecting one more attribute into the relation. The privacy-anomalies injected by adding one more attribute were identified as privacy-anomalies by both systems. The reason that they were identified was because there were no n-gram that contained a reference to new attribute in its query abstraction.

One of the detected privacy-anomalies corresponds to the query shown below.

| dob | city | gender | diagnoses | country | ... |
|---|---|---|---|---|---|
| 1981 | Dublin | Male | Flu | Ireland | ... |
| 1981 | Dublin | Male | Flu | Ireland | ... |
| 1981 | Dublin | Male | Diarrhoea | Germany | ... |
| 1920 | Cork | Male | Heart Disease | Ireland | ... |
| 1981 | Galway | Female | Acne | Ireland | ... |
| 1984 | Galway | Male | Flu | Spain | ... |
| 1984 | Galway | Male | Diabetes | Ireland | ... |
| 1984 | Galway | Male | Hypertension | Ireland | ... |
| 1984 | Galway | Male | Leg Fracture | France | ... |
| ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |
| ~~1981~~ | ~~Dublin~~ | ~~Male~~ | ~~Flu~~ | ~~Germany~~ | ... |

Table 5: A fragment of hospital dataset.The strike-through attribute values represents a deleted row.

| Description of privacy-anomalies | Number of anomalies injected |
|---|---|
| Addition of one or more attributes to the base relation shown in Table 5. For instance, a new attribute, like `country`, was inserted in the relation and queries were made to retrieve this attribute values. | 5 |
| Update or Deletion of records from relation shown in Table 5 | 10 |

Table 6: Description of Privacy-anomalies injected.

| dob | city | diagnoses |
|---|---|---|
| 1920 | Cork | Heart Disease |

Table 7: Response to a undetected privacy-anomalous query.

```
SELECT diagnoses, dob, city, country
FROM hospitalDB
WHERE dob = '1981'
AND city = 'Dublin';
```

The normative privacy-profile contains no privacy limit reference to the new (or combination of new) attribute.

## 3.2 Undetected Privacy-anomalies

A privacy-anomaly undetected by the n-gram based approach but detected by the privacy model is:

```
SELECT dob, city, diagnoses
FROM hospitalDB
WHERE dob = '1920'
AND city = 'Cork' ;
```

The query returns a relation with one record as shown in Table 7. It is identified as a privacy-anomaly by the privacy model for the reason being that the specified value of $k$ for the specified set of quasi-identifier meant that an adversary was able to single out an individual. This anomaly is undetected by n-gram based security-anomaly detection approach because there was an n-gram in normative profile contained a reference to this query abstraction.

## 3.3 Identifying Appropriate Privacy Limits

In order to find the optimal values of $k$, in the mining process, in theory, all the combinations of quasi-identifiers need to be considered. This, in essence, is a combinatorial explosion, especially in the case of a large number of quasi-identifiers. Additionally, one may discover either very large or very small values of $k$ in practice for certain combinations of quasi-identifiers. Therefore, in order to discover reasonable values of $k$, one may define a range while mining the values of $k$ such that the values falling within

the range and their corresponding combinations of quasi-identifiers are considered for privacy-profiles.

# 4 Privacy Attacks based on Query Sequences

This section demonstrates a privacy attack whereby an adversary discovers information about an individual while the privacy-anomaly detection system is in place. Consider the relation shown in Table 8.

| Name | | City | Age | Salary (sensitive attribute) |
|---|---|---|---|---|
| Mark | Single | New York | [30 - 40] | 112k |
| James | Single | New York | [30 - 40] | 34k |
| John | Single | New York | [30 - 40] | 56k |
| Henry | Single | New York | [30 - 40] | 78k |
| Imran | Single | New York | [30 - 40] | 91k |
| David | Married | London | [30 - 40] | 112k |
| Alice | Married | London | [30 - 40] | 30k |
| Bob | Married | London | [30 - 40] | 45k |
| Aron | Married | London | [30 - 40] | 115k |
| Harry | Married | London | [30 - 40] | 180k |
| Jordan | Separated | Cork | >40 | 65k |
| Ryan | Separated | Cork | >40 | 100k |
| Bentley | Separated | Cork | >40 | 80k |
| Simon | Married | Rennes | >40 | 150k |

Table 8: Relation updated_table_smp.

| | |
|---|---|
| $Q_1$ | SELECT Salary<br>FROM updated_table_smp<br>WHERE city = 'Rennes'; |
| $Q_2$ | SELECT Salary<br>FROM updated_table_smp; |
| $Q_3$ | SELECT MaritalStatus, Salary, Age<br>FROM updated_table_smp<br>WHERE City= 'New York' AND city ='London' AND city = 'Cork'; |

Table 9: Sequence of queries executed over the relation updated_table_smp shown in Table 8.

Suppose the sequence of queries $Q_1$, $Q_2$ and $Q_3$, shown in Table 9, are executed

over the relation updated_table_smp shown in Table 8.

Query $Q_1$ is labelled as an privacy-anomaly (and the query response is suppressed) because the threshold is not satisfied as *k*-anonymity and *DRSQL* is not satisfied by the query $Q_1$. Whereas query $Q_2$ passes the threshold and the response to the query is shown in Table 10 that results in 14 records, that is, the value of attribute Salary, being returned. Query $Q_3$ also passes the privacy criteria and returns 13 records, as shown in Table 11. The adversary, knowing that Simon's record is in the table (as background/external knowledge) and that Simon lives in Rennes, reveals that last remaining entry blocked by the query mechanism is of 'Simon' and the corresponding salary attribute value is 150k.

In particular, the described attack is a form of a differencing attack [12]. Differencing attacks have been seen previously on aggregates. This demonstrates that *k*-anonymity in interactive settings is also susceptible to these differencing attacks. For the purpose of demonstration, the example of a differencing attack is kept simple; however, real world differencing attacks can take more sophisticated forms, where the adversary can make multiple queries to narrow down the aggregate data until the subject's information is not revealed.

| Salary (sensitive attribute) |
|---|
| 112k |
| 34k |
| 56k |
| 78k |
| 91k |
| 112k |
| 30k |
| 45k |
| 115k |
| 65k |
| 100k |
| 80k |
| 150k |

Table 10: Records returned in the response to query $Q_2$.

| MaritalStatus | City | Age | Salary (sensitive attribute) |
|---|---|---|---|
| Single | New York | [30 - 40] | 112k |
| Single | New York | [30 - 40] | 34k |
| Single | New York | [30 - 40] | 56k |
| Single | New York | [30 - 40] | 78k |
| Single | New York | [30 - 40] | 91k |
| Married | London | [30 - 40] | 112k |
| Married | London | [30 - 40] | 30k |
| Married | London | [30 - 40] | 45k |
| Married | London | [30 - 40] | 115k |
| Married | London | [30 - 40] | 180k |
| Separated | Cork | >40 | 65k |
| Separated | Cork | >40 | 100k |
| Separated | Cork | >40 | 80k |

Table 11: Records returned in response of query $Q_3$.

# 5 Applying Security-anomaly Detection to Detect Unknown Privacy Attacks

In general, interactive query mechanisms are susceptible to the attacks described in the previous section, and as a consequence there is little privacy-preserving interactive query mechanisms (specifically for microdata release) in the existing literature. The existing differentially private interactive mechanisms allow a limited number of interactive queries for aggregate data. Restricting the number of queries is a significant barrier for an analyst in achieving the true potential for data analytics. Privacy attacks, similar to the one presented in the previous section, are violations of formal privacy definitions like $k$-anonymity.

Another aspect of these privacy attacks is that the querying pattern to infer information about the subject(s) is unknown, therefore, we refer to them as unknown privacy attacks. Unknown privacy attacks lead to inferring information about the subject(s). An adversary can articulate the queries in different ways to reveal information about a subject(s). In this work, inference implies privacy attacks, that is, the adversary infers information about the subject(s) while a formal privacy definition is in place resulting in a violation of formal privacy definition. Additionally, these privacy attacks are based on query correlation, that is, an individual query is safe in terms of privacy but when considered as sequence they result in the violation of formal privacy definition. This section presents an investigation into whether the inferences can be detected as anomalies.

We present a novel perspective on the detection of privacy attacks by proposing an interpretation of the behavioural-based detection approach. There are a number of behavioural-based anomaly detection detection approaches that can be explored in this context [11, 13, 14, 15]. We investigate the application of n-gram approach, proposed approach in [11], to detect unknown privacy attacks as anomalies in the next section.

## 5.1 Detecting Privacy Attacks as Privacy-Anomalies

A behavioural-based approach to detect inferences as anomalies is described in this section where the n-gram based approach is applied to the audit logs of the ($k$-anonymity)-PAD system. The idea is to model querying behaviours in the presence of a privacy-preserving interactive query mechanism and compare the normative querying behaviour with the run-time querying behaviour to detect deviations.

For the SQL query abstraction, the specialization of the abstraction, as discussed in [11], is employed. The SQL query abstraction that is a tuple representation of an SQL query and consists of query features like relation name, attribute names, the amount of returned data or any statistics on the returned data.

An abstraction of an SQL query $Q_i$ is denoted as $Abs(Q_i)$. The adopted query abstraction technique has also been studied in [16]. The query abstraction technique replaces the constant values in a query $Q_i$ with place-holders (literal 'VAR_VAL'), and is denoted as $Abs(Q_i)$. $Abs(L)$ is defined as the mapping of $Abs(Q_i)$ over the elements $Q_i$ of $L$. The reason to chose this query abstraction is that it gives us a reasonable level of precision in capturing the querying behaviour of user. A more fine grained abstraction would require some symbolic evaluation of the queries which was beyond the scope of this work. Examples of the employed SQL query abstraction technique are shown in Table 12.

The n-gram profiles are generated in the same manner as discussed in [11] that is given a safe audit log of SQL query $L^{pp}_{norm}$ and a run-time log $L^{pp}_{run}$ then the constructed normative profile and run-time profile are $\beta_{norm} = ngram(Abs(L^{pp}_{norm}), n)$ and $\beta_{run} = ngram(Abs(L^{pp}_{run}), n)$. The mismatches are given by $S^{miss}_{\beta_{run}-\beta_{norm}} = \beta_{run}$ - $\beta_{norm}$.

In order to evaluate the detection of privacy attacks by applying the n-gram based approach to the logs of interactive querying mechanism, a synthetic query generator was designed that had defined a set of SQL query templates. The underlying database was populated with a fragment version of well-known Census (Adult) dataset [17]. Query templates were designed to be executed on the Census dataset. The queries were count queries mimicking a data analytics scenario. For example, the count queries were for the form: how many subjects are Female, how many subjects have a Bachelors

| $Q_i$ | SQL statement | SQL query abstraction $Abs(Q_i)$ |
|---|---|---|
| $Q_1$ | `SELECT city FROM bankDatabase WHERE id = 2` | `SELECT city FROM bankDatabase WHERE id = VAR_VAL` |
| $Q_2$ | `SELECT city FROM bankDatabase WHERE id = 9` | `SELECT city FROM bankDatabase WHERE id = VAR_VAL` |
| $Q_3$ | `SELECT city FROM bankDatabase WHERE id = 3` | `SELECT city FROM bankDatabase WHERE id = VAR_VAL` |
| $Q_4$ | `SELECT city FROM bankDatabase WHERE id = 3 AND Name = "Alice"` | `SELECT city FROM bankDatabase WHERE id = VAR_VAL AND Name = VAR_VAL` |

Table 12: Examples of deployed SQL abstractions.

degree, so on and so forth. For the experimentation, a safe log $L^{pp}_{norm}$ was generated for the construction of normative profile using the synthetic data generator.

In order to construct privacy attacks, five unique records were inserted into the database that leads to inferences, where unique implies that one of the attribute or combination of the attributes existed only once in the entire database. For example, a record was inserted with `occupation` as `post-doc`, that is, in the underlying database there was only one record where the value for `occupation` was `post-doc`. Another record was inserted where the value for `native-country` was set to `Malaysia`, that is, there was only one record where the `native-country` was `Malaysia`. Table 13 shows the make-up of the inserted records to enable privacy attacks.

| # | Description of Unique Record |
|---|---|
| 1 | Attribute `occupation` with value as `post-doc` |
| 2 | Attribute `native-country` with value as `Malaysia` |
| 3 | Attribute `native-country` with value as `Spain` and Attribute `age` as `33` |
| 4 | Attribute `native-country` with value as `Singapore` and Attribute `age` as `32` |
| 5 | Attribute `native-country` with value as `Singapore` and Attribute `occupation` as `Academics` |

Table 13: Inserted unique records in the database to enable privacy attacks.

Queries were made to infer the associated salaries for these five unique records. Table 14 shows the number of queries made to reveal the salary for the records shown

in Table 13. These malicious query sequences were made part of the other logs $L^{pp}_{run}$ for the construction for the run-time profile.

| Privacy Attack # | Inference Sequence Length |
|---|---|
| 1 | 7 |
| 2 | 9 |
| 3 | 17 |
| 4 | 13 |
| 5 | 10 |

Table 14: Length of the query sequences to reveal salaries.

A normative profile $\beta_{norm}$ and a run-time profile $\beta_{run}$ were constructed using $L^{pp}_{norm}$ and $L^{pp}_{run}$, respectively and compared. The size of the n-gram was kept at 4 for the generation of profiles. The sequence of queries made to infer injected unique records was labelled as anomalies the detection phase by the n-gram approach. The Table 15 shows the number of mismatches for each privacy attack.

| Privacy Attack # | $\lvert S^{miss}_{\beta_{run}-\beta_{norm}} \rvert$ |
|---|---|
| 1 | 5 |
| 2 | 7 |
| 3 | 19 |
| 4 | 15 |
| 5 | 13 |

Table 15: Detection of privacy attacks as privacy-anomalies: the table shows the number of mismatches that resulted from each of the 5 privacy attacks with the n-gram of size 4.

The query sequences resulting in inferences were detected as privacy-anomalies, which is a indication of potential effectiveness of n-gram based approach to detect inference and unknown privacy attacks as privacy-anomalies.

# 6 Conclusions

While existing behavioural-based anomaly detection systems considered anomalies arising from anomalous queries of users, this paper explores anomalies characterised in terms of formal definitions of privacy. This work studies privacy semantic notion of behavioural-based anomaly detection systems. The notion of privacy-anomaly detection (PAD) introduced in this work enables one to learn a privacy model from the

past log of interaction with the DBMS and detects deviations as privacy-anomalies. A naïve instantiations of PAD was presented based on $k$-anonymity ($k$-anonymity, DR)-PAD. A study was carried out to examine whether the privacy violations based on a single query detected the privacy-anomaly detection system are also detected by n-gram security-anomaly detection system as anomalies. Results showed a number of single query based privacy violations that had no reference n-gram in normative profiles were labelled as anomalies by n-gram based anomaly detection system. This work also considered privacy attacks that were violations of formal privacy definitions and were based on query correlation where a single query is not privacy-anomalous but a sequence of queries results in a violation of formal privacy definition. Results showed that behavioural-based security anomaly detection system (constructed using n-grams) in [11] detected these privacy attacks as privacy-anomalies. This led to a discovery of another aspect of the n-gram based approach whereby when it is applied on the logs generated by interactive query settings with the presence of formal privacy definition, it has the potential to detect privacy attacks based on query correlation as privacy-anomalies.

Privacy attacks can manifest itself in a variety of unexpected ways, the results suggest that therefore a silver bullet for anonymization may not be a way forward rather utilise defence in depth from privacy perspective. One such privacy control is the proposed privacy anomaly detection system. As a topic of future work, we plan to explore how to compose and compare multiple privacy definitions using multi-criteria decision-making method found in fuzzy logic [18, 19], in particular, known as *triangular-norms* (*t-norms*).

# References


[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L119:1–88, May 2016.

[2] Anonymisation and pseudonymisation. Technical report, Data Protection Commission, Ireland, 2019. Online at: `https://www.dataprotection.ie/en/guidance-landing/anonymisation-and-pseudonymisation`.

[3] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (SP' 08)*, pages 111–125, Los Alamitos, CA, USA, may 2008. IEEE Computer Society.

[4] Michael Barbaro and Tom Zeller Jr. A face is exposed for aol searcher no. 4417749. The New York Times. Online at: http://www.nytimes.com/2006/08/09/technology/09aol.html?mcubz=2.

[5] Alex Hern. New york taxi details can be extracted from anonymised data, researchers say, 2014. The Guardian. Online at: https://www.theguardian.com/technology/2014/jun/27/new-york-taxi-details-anonymised-data-researchers-warn.

[6] Latanya Sweeney. Simple demographics often identify people uniquely. Working paper, 2000. Working paper. Online at: http://dataprivacylab.org/projects/identifiability/.

[7] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[8] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115, April 2007.

[9] C. Clifton and T. Tassa. On syntactic anonymity and differential privacy. In *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, pages 88–93, April 2013.

[10] Jordi Soria-Comas. Improving data utility in differential privacy and k-anonymity. *CoRR*, abs/1307.0966, 2013.

[11] Muhammad Imran Khan and Simon N. Foley. Detecting anomalous behavior in DBMS logs. In Frédéric Cuppens, Nora Cuppens, Jean-Louis Lanet, and Axel Legay, editors, *Risks and Security of Internet and Systems - 11th International Conference, CRiSIS 2016, Roscoff, France, September 5-7, 2016, Revised Selected Papers*, volume 10158 of *Lecture Notes in Computer Science*, pages 147–152. Springer, 2016.

[12] Muneeb Ul Hassan, Mubashir Husain Rehmani, and Jinjun Chen. Differential privacy techniques for cyber physical systems: A survey, 2018. Online at: https://arxiv.org/abs/1812.02282.

[13] Muhammad Imran Khan, Simon N. Foley, and Barry O'Sullivan. *DBMS Log Analytics for Detecting Insider Threats in Contemporary Organizations*. IGI Global", address = , year = 2018, pages = 207-234, chapter = 10.

[14] M. I. Khan, B. O Sullivan, and S. N. Foley. Towards modelling insiders behaviour as rare behaviour to detect malicious rdbms access. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 3094–3099, 2018.

[15] Muhammad Imran Khan, Barry O'Sullivan, and Simon N. Foley. A semantic approach to frequency based anomaly detection of insider access in database management systems. In Nora Cuppens, Frédéric Cuppens, Jean-Louis Lanet, Axel Legay, and Joaquin Garcia-Alfaro, editors, *Risks and Security of Internet and Systems*, pages 18–28, Cham, 2018. Springer International Publishing.

[16] Gokhan Kul, Duc Luong, Ting Xie, Patrick Coonan, Varun Chandola, Oliver Kennedy, and Shambhu Upadhyaya. Ettu: Analyzing query intents in corporate databases. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, pages 463–466, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee.

[17] Dua Dheeru and Efi Karra Taniskidou. UCI machine learning repository, 2017. Online at: `http://archive.ics.uci.edu/ml`.

[18] Francesc Esteva, Lluis Godo, and Carles Noguera. First-order t-norm based fuzzy logics with truth-constants: Distinguished semantics and completeness properties. *Annals of Pure and Applied Logic*, 161(2):185 – 202, 2009. Festschrift on the occasion of Franco Montagna's 60th birthday.

[19] George Metcalfe, Nicola Olivetti, and Dov Gabbay. *Proof Theory for Fuzzy Logics*. Springer Publishing Company, Incorporated, 1st edition, 2008.