2013 IEEE 13th International Conference on Data Mining Workshops

Differentially Private Anomaly Detection with a Case Study on Epidemic Outbreak Detection

Liyue Fan, Li Xiong
Dept. Mathematics & Computer Science
Emory University
Atlanta, Georgia 30322
Email: (lfan3,lxiong)@mathcs.emory.edu

***Abstract*—Anomaly detection is an important problem that has been studied in a variety of application domains, ranging from syndrome surveillance for epidemic outbreaks to intrusion detection in computer networks. The data collected from individual users contain sensitive information, such as health records and network usage data, and thus need to be transformed prior to the release for privacy preservation. In this paper, we propose a novel framework for anomaly detection with differential privacy. Real-time private user data can be aggregated and perturbed to guarantee privacy, while the posterior estimate is released continuously for anomaly detection tasks. Our framework is not limited to any specific application domains. We illustrate the sensitivity analysis and evaluate our framework in the context of syndrome surveillance. Empirical results with simulated data sets confirm the effectiveness of our solution while providing provable privacy guarantee.**

***Keywords*-Anomaly Detection, Differential Privacy, Time Series**

I. Introduction

Anomaly detection is an important problem that has been studied in a variety of application domains. Private user data can be collected and analyzed in real-time for rare events and unusual patterns. Consider the following examples:

- **Syndrome Surveillance** Information of individual patients is collected at an Emergency Department. The aggregated statistics, such as daily counts of Influenza cases, can be shared with third party researchers to detect epidemic outbreaks as early as possible.
- **Network Monitoring** An Internet service provider gathers data from individual users about their network activities. The aggregated data, for instance, the number of new connections during each time interval, can be monitored and analyzed by third parties to detect potential intrusions and attacks.

As in the above examples, individual user data are accessible to and can be collected by a trusted server. The collected data may be then aggregated and continuously shared with other un-trusted entities for anomaly detection purposes. The trusted server, i.e. data publisher, is assumed to be bound by contractual obligations to protect the user's interests, therefore it must ensure that releasing the data does not compromise the privacy of any individual who contributed data. The goal of our work is to enable the publisher to share useful aggregate statistics over individual users continuously for anomaly detection while guaranteeing their privacy.

The current state-of-the-art paradigm for privacy-preserving data publishing is *differential privacy* [1], which requires that the aggregate statistics reported by a data publisher be perturbed by a randomized algorithm $\mathcal{A}$, so that the output of $\mathcal{A}$ remains roughly the same even if any single tuple in the input data is arbitrarily modified. Differential privacy can be achieved by the Laplace mechanism [2], which adds a Laplace perturbation noise to each aggregate.

The challenge of releasing continual aggregates with differential privacy guarantee is high perturbation error. In the example anomaly detection tasks, private data values are aggregated and monitored for a long period of time, e.g. T time stamps. A straightforward application of Laplace mechanism by adding a perturbation noise at every time stamp can lead to a high overall perturbation error, i.e. $\Theta(T)$, leaving the released data useless especially when T is large.

Few works [3]–[6] studied the problem of sharing continual aggregate statistics or aggregate time series with differential privacy. Rastogi and Nath [5] proposed an algorithm to share time series data based on the Discrete Fourier Transform. Since the entire time-series is required to perform the Fourier transformation, this method is not applicable to real-time anomaly detection tasks. Chan et al. [3] and Dwork et al. [4] studied separately the differentially private continual counter over a binary stream and achieved a bounded error at each time stamp. However, both works adopt an event-level privacy model, with the perturbation mechanism designed to protect the presence of an individual event rather than the presence or privacy of a user. We recently proposed FAST [6], a framework with filtering and adaptive sampling for monitoring aggregate time series with user-level privacy guarantee. In this paper, we adapt FAST framework and filtering techniques for anomaly detection purposes.

978-0-7695-5109-8/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDMW.2013.129

A. Contributions

To our best knowledge, we take a first step towards a solution for differentially private anomaly detection. We consider the problem of releasing private, aggregate user statistics continuously in real-time for anomaly detection tasks. Note that certain anomaly detection methods that require individual user attributes, such as outlier detection, are beyond the scope of this work. Discussions about different types of anomaly detection algorithms are provided towards the end of this paper.

Framework: We propose a novel framework for detecting anomalies from continual aggregate user statistics with differential privacy guarantee. Given an anomaly detection task, sensitivity analysis can be conducted to evaluate the privacy risk in that specific application domain. Laplace perturbation mechanism is adopted to achieve the desired level of differential privacy guarantee. FAST filtering algorithm is utilized to generate posterior estimates in order to improve the accuracy of released aggregates.

Case Study: Despite of the actual application domain, our framework can enable any anomaly detection algorithms based on continual aggregate statistics. We demonstrate the usability of our solution with a case study on epidemic outbreak detection. To reduce the overall perturbation error, we derive a practical estimate of the global sensitivity for releasing daily Influenza counts in 10 year's period with public survey statistics. We shown the practical sensitivity, Δf, can be much smaller than the total number of time stamps T, i.e. $\Delta f \ll T$. With the practical sensitivity Δf, we establish the perturbation model as well as the filtering models in our framework for Influenza outbreak detection correspondingly.

Evaluation: We empirically study the utility of the private, released aggregates provided by our solution in the context of Influenza outbreak detection. We perform our experiments with six simulated data sets that are generated for outbreak detection evaluations. Three extensively used epidemic outbreak detection algorithms, i.e. C1, C2, and C3 [7], are tested on both the original data series and the private, released data series. The results confirm that the released aggregates by our solution retain high utility for anomaly detection while providing a strong privacy guarantee.

The organization of the paper is as follows: we review the recent works related to anomaly detection and differential privacy in Section II; Section III introduces the formal differential privacy definition; in Section IV, we present our proposed framework as well as technical details of each component and demonstrate our solution through a case study on epidemic outbreak detection; Section V includes a set of empirical studies on the utility of private data released by our framework; in Section VI we conclude this paper and provide discussions on potential future work.

II. Related Works

Here we briefly review recent, relevant works on anomaly detection, epidemic outbreak detection for public health interest, and differential privacy.

A. Anomaly Detection.

At an abstract level, an anomaly is defined as a pattern that does not conform to expected normal behavior [8]. Anomaly detection has been researched within diverse research areas and application domains, such as epidemic outbreak detection for syndrome surveillance [9] and intrusion detection for cyber-security [10]. Both tasks are conducted on continual aggregate statistics, i.e. aggregate time series. The aberration detection algorithms studied by Jackson et al. [9] aim to find disease outbreaks from daily counts of diagnosed cases. The work of Caberera et al. [10] addressed network intrusion detection by establishing statistical models for the number of incoming connections within a given time interval.

Specifically for outbreak detection, there are a multitude of algorithms that have been reported and applied to a variety of disease studies. Most algorithms compare the current signal, i.e. current *count* of diagnosed cases, with a baseline period, i.e. previously released $counts$, in order to determine whether there is an outbreak or not. Three control-chart based algorithms proposed by Hutwagner et al. [7], commonly referred to as EARS [11] C1, C2 and C3, require little baseline data and have been found to provide early detection of outbreaks. The Negative Binomial Cusum (NBC) method, originally proposed in Hawkins and Olwell [12], is reported to reduce the number of false positives generated by other cusum methods. Unlike the cusum methods, the Historical Limit Method (HLM) [13] incorporates historical data and accounts for seasonality by design. Gatton et al. [14] proposed to model the number disease cases in a time period as a Poisson process and their method considers years of historical data as baseline.

B. Differential Privacy.

Our work is designed to provide differential privacy guarantee. Dwork et al. [2] first established the guideline to guarantee differential privacy for individual aggregate queries by calibrating the Laplacian noise to the global sensitivity of each query. They showed that any function with low sensitivity can be computed privately while possibly preserving high accuracy. However, the sequential composition property of differential privacy, studied in [15], imposes great utility challenge to applying Laplace mechanism to long time series, which is crucial to anomaly detection tasks.

Among the few works addressing the problem of sharing time series data with differential privacy, Rastogi and Nath [5] proposed an off-line algorithm based on the Discrete Fourier Transform (DFT): it first performs Discrete Fourier Transform on an input time series of $counts$ and retains/perturbs the first l DFT coefficients to guarantee

differential privacy. The released series can then be derived from Inverse Discrete Fourier Transform (IDFT) with the perturbed coefficients. This algorithm has shown to preserve trend information in the released series. However, it is not compatible for real-time anomaly detection. The recent works [3] [4] on continuous data streams defined the *event-level* privacy to protect an event, i.e. one user's presence at a particular time point, rather than the presence of that user. For example, if one user contributes to the aggregation at time $k-1$, k, and $k+1$, the event-level privacy protects the user's presence at only one of the three time points, resulting the rest two open to attack. We recently proposed the FAST [6] framework for sharing real-time aggregates while providing user-level privacy guarantee. It samples long time-series to reduce the impact of perturbation noise generated by the differential privacy mechanism, and simultaneously uses filtering to improve the accuracy of released aggregate at every time stamp. In this paper, we adapt FAST framework and filtering techniques to release accurate aggregates in real-time for anomaly detection applications.

III. Privacy Definition

The privacy definition adopted in our framework is α-*differential privacy* [1]. A mechanism is differentially private if its outcome is not significantly affected by the removal or addition of a single user. An adversary thus learns approximately the same information about any individual user, irrespective of his/her presence or absence in the original database.

Definition 1 (α-Differential Privacy [1]): A non-interactive privacy mechanism $\mathcal{A}$ gives α-differential privacy if for any dataset D_1 and D_2 differing on at most one record, and for any possible anonymized dataset $\widetilde{D} \in Range(\mathcal{A})$,

$$Pr[\mathcal{A}(D_1) = \widetilde{D}] \leq e^{\alpha} \times Pr[\mathcal{A}(D_2) = \widetilde{D}] \qquad (1)$$

where the probability is taken over the randomness of $\mathcal{A}$. The privacy parameter α, also called the *privacy budget* [15], specifies the degree of privacy offered. Intuitively, a lower value of α implies stronger privacy guarantee and a larger perturbation noise, and a higher value of α implies a weaker guarantee while possibly achieving higher accuracy. Two databases D_1 and D_2 that differ on at most one record are called *neighboring databases*.

Dwork et al. [2] show that given function $f : D \rightarrow \mathbb{R}^d$, α-differential privacy can be achieved as follows:

$$\tilde{f}(D) = f(D) + (\nu_1, ..., \nu_d) \; . \qquad (2)$$

ν_k are independent draws from a Laplace distribution $Lap(\frac{\Delta f}{\alpha})$ with 0 mean and the following probability density function:

$$p(x) = \frac{1}{2\frac{\Delta f}{\alpha}} e^{-|x|/\frac{\Delta f}{\alpha}} \; . \qquad (3)$$

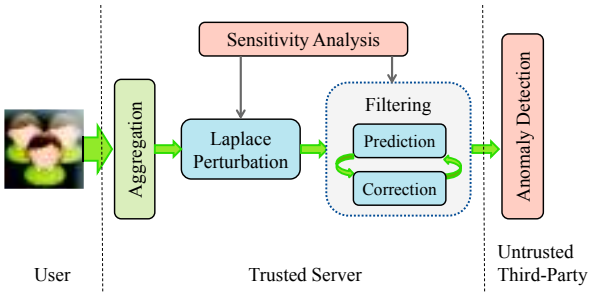


Figure 1. Proposed Framework

Δf represents the *global sensitivity* [2] of the function f. Given function $f : D \rightarrow \mathbb{R}^d$, the global sensitivity of f is defined as:

$$\Delta f = \max_{D,D'} ||f(D) - f(D')||_1 \qquad (4)$$

where the maximum is taken over all pairs of neighboring databases.

Example 3.1 (count Sensitivity): For a *count* query, adding or removing an individual user from database D would change the output by at most 1, i.e. $\Delta count = 1$.

In the example anomaly detection tasks, it is required to output a series of *count* queries, i.e. $f(D) = \{x_1, ..., x_T\}$, where x_k represents the number of events/occurrences during the kth time interval. The global sensitivity of f can be bounded by $\Delta f \leq T$, since adding or removing an individual would change x_k by 1 for each k, $k \in \{1, ..., T\}$.

The work [4] proposed a differentially private continual counter with the notion of *event*-level privacy, where the neighboring databases differ at u_i, a user u's contribution at time stamp i. In our study, we provide a stronger privacy guarantee, *user*-level privacy, where the neighboring databases differ at the user u, i.e. u's contribution at all time stamps, thus protecting sensitive information about user u at any time.

IV. Differentially Private Anomaly Detection

In this section, we introduce a framework for anomaly detection based on continual aggregate statistics with differential privacy guarantee. We provide an overview of the proposed framework and technical details of each component through a case study of epidemic outbreak detection for syndrome surveillance.

A. Framework Overview

An overview of our proposed framework for privacy-preserving anomaly detection is shown in Figure 1. The input to the framework is a time series of aggregates that are collected from individual users by a server. Since the server is trustworthy, such as an Emergency Department or Internet Service Provider, we assume no privacy risk at the aggregation step. At every time stamp, the raw aggregate,

e.g. the number of Influenza cases or the number of incoming connections, is perturbed by the *Laplace Perturbation* module to enforce the pre-defined level of privacy guarantee. The perturbed aggregate is then received by the *Filtering* module to produce a posterior estimate. There are two internal procedures, i.e. *prediction* and *correction*, that are performed recursively at every time stamp. The posterior estimate, which is less noisy than the purely perturbed aggregate, can then be used by an anomaly detection algorithm.

On the other hand, given a real-time anomaly detection task, such as epidemic outbreak detection or network intrusion detection, a couple task-specific aspects need to be considered to set up the privacy preserving framework. Firstly, what aggregate is required at each time stamp, e.g. the $average$ value of a certain attribute or a $count$ of records that satisfy certain predicate? Secondly, what is the privacy principal in the application domain, an $event$ or a $user$? With these questions settled, sensitivity analysis can be performed to understand the privacy risk of releasing continual aggregates given the anomaly detection task. We show a practical estimate of the global sensitivity can be derived with domain specific knowledge/statistics. The Laplace perturbation model as well as the internal models for filtering can be established correspondingly with respect to the global sensitivity. We will introduce a case study on epidemic outbreak detection and discuss each component in details in the following subsections.

B. Epidemic Outbreak Detection

The early detection of disease outbreaks has long been a concern of public health because of the potential to reduce morbidity and mortality. Patient data collected from Emergency Department/urgent-care and ambulatory-care can be aggregated, e.g. daily case counts, and shared continuously as the input to regional or national syndromic surveillance systems, such as CDC BioSense [16], EARS [11], and RODS [17]. Below we provide technical details of each component with an application to real-time epidemic outbreak detection. Note that any anomaly detection task on continual user statistics can be enabled by our proposed framework.

Most outbreak detection methods compare the current case count with a baseline period to determine whether there is an outbreak or not. We choose three extensively used, EARS [11] algorithms C1,C2, and C3 [7] for outbreak detection task. EARS algorithms were developed based on a one-sided positive CUSUM (cumulative sum) calculation. They were named according to their degree of detection sensitivity, with C1 being the least sensitive and C3 the most sensitive. The advantage of EARS algorithms is that they require limited baseline data, e.g. case counts from the previous week, compared to other methods that require that of previous years.

For C1 and C2, the test statistics on day k was calculated as

$$C_i(k) = \frac{r_k - \mu_k}{\sigma_k}, i \in \{1, 2\} \tag{5}$$

where r_k is the disease case count on day k, and μ_k and σ_k are the mean and standard deviation of the counts from the baseline period. For C1, the baseline data is $\{r_{k-7}, ..., r_{k-1}\}$, i.e. one week's data immediately prior the current day. For C2, the baseline is $\{r_{k-9}, ..., r_{k-3}\}$, i.e. one week's data with a two-day gap from the current day. In EARS [11], an *outbreak* is detected when $C_i(k) > 3, i \in \{1, 2\}$.

C3 algorithm uses the C2 statistics from day k and the previous two days. The test statistic for C3 is calculated as

$$C_3(k) = \sum_{j=k}^{k-2} \max(0, C_2(j) - 1). \tag{6}$$

An *outbreak* is detected when $C_3(k) > 2$.

C. Sensitivity Analysis.

For outbreak detection tasks, the daily case count of a certain illness is usually the data of interest. In most existing syndrome surveillance systems, this data is collected daily from the Emergency Department for early detection. Below we analyze the global sensitivity for monitoring the daily count of Influenza cases over T days, where T is a pre-defined project time line, e.g. the number of days over 10 years. Note that for other anomaly detection applications, the sensitivity analysis can be conducted in a similar way with domain-specific statistics/knowledge.

Let D be the patient database and $f(D)$ outputs a sequence of counts $\{x_1, ..., x_T\}$, where x_k represents the number of Influenza cases on day k. As discussed in Section III, the global sensitivity of $f(D)$ can be set $\Delta f = T$, which is the length of the surveillance period, roughly 3650 for 10 years' time line. Although theoretically sound, there are two drawbacks to adopting the upper bound of Δf. First of all, it is an impractical estimate and will almost never happen in reality. Recall that the global sensitivity defines the maximal contribution of any individual to the function output. It is very rare that anyone would visit the emergency room and be diagnosed with Influenza for every day in 10 year's period. Secondly, when T is large, the Laplace perturbation error introduced at every time stamp has a variance proportional to T^2, according to Equation (3). The released data with such high perturbation error would be practically useless for detection purposes.

Below we quantify the rareness of $\Delta f = T$ with publicly available survey statistics and provide a practical, smaller valued estimate of Δf. Again, the value of Δf represents the number of times that an individual visits emergency room and is diagnosed with Influenza in 10 years' period. According to the National Hospital Ambulatory Medical

Care Survey: 2010 Emergency Department Summary Tables[1], the number of emergency room visits per 100 persons in 2010 is 42.8 and 3.2% of visits are diagnosed with "Acute upper respiratory infections, excluding pharyngitis", which includes Influenza. We note that young children, under 1 years old, have a higher rate of emergency room visit and other age groups have similar rates. Therefore, we derive the probability of a patient visiting emergency room and being diagnosed with Influenza every year, p, as follows:

$$p \approx 42.8\% \times 3.2\% = 1.4\% \ . \tag{7}$$

Let n denote the number of a patient visiting emergency room and being diagnosed with Influenza in 10 years' period. Given Equation (7), we obtain the following:

$$P(n \leq 2) = \sum_{k=0}^{2} P(n=k) = \sum_{k=0}^{2} \binom{10}{k} p^k (1-p)^{10-k} = 99.9\% \tag{8}$$

which indicates that with 99.9% confidence, any individual patient will contribute to $f(D)$ at most twice in every 10 years.

D. Laplace Perturbation

The Laplace Perturbation component ensures differential privacy by perturbing the input aggregate at every time stamp. Let $f(D)$ outputs a sequence of counts $\{x_1, ..., x_T\}$ on data set D, where x_k represents the number of events of interest during time interval k. Given the global sensitivity Δf and the desired level of privacy α, we can derive a private count sequence $\tilde{f}(D)$ that satisfies α-differential privacy as follows:

$$\tilde{f}(D) = \{z_k = x_k + \nu_k | k = 1, ..., T\} \tag{9}$$

where ν_k are independent draws from $Lap(\frac{\Delta f}{\alpha})$. The default value for Δf is T. When a practical estimate can be derived from domain-specific statistics, we can set Δf with a much smaller value, i.e. $\Delta f = 2$ as in the analysis above for Influenza outbreak detection. This perturbation model in Equation (9) is used in our proposed framework.

A natural concern arises when setting $\Delta f < T$: how about those individuals that are counted more than Δf times in the released data? As the sensitivity analysis above shows, only few patients, i.e. 0.1%, will contribute to the time series for more than two times in 10 years. In reality, those are patients who are more susceptible to Influenza or who have more frequent emergency room visit history, and thus can choose to opt out at the data collection stage without significantly affecting the quality and quantity of collected data. Below we assume that the database consists of the majority patients, i.e. 99.9% patients who contribute to the data series at most 2 times over 10 years, and can be protected by differential privacy.

[1]http://www.cdc.gov/nchs/ahcd/web_tables.htm

E. Filtering

The Filtering component in our framework utilizes time series modeling and estimation algorithms to improve the accuracy of released aggregates. In our context, "filtering" refers to the derivation of posterior estimates based on a sequence of noisy measurements, in hope of removing background noise from signals. In FAST [6], the Kalman filter based estimation algorithm is adopted and is shown to be computationally efficient [18]. We briefly show the state-space model for time series data as well as the filtering algorithms used in our framework.

A state-space model describes the underlying dynamics of a time-series as well as how an observation is derived from the hidden state. For a time series of *count* queries, i.e. $\{x_k\}$, we establish the following models:

$$x_k = x_{k-1} + \omega \ , \tag{10}$$

$$\omega \sim \mathcal{N}(0, Q) \ . \tag{11}$$

This constant process model indicates that adjacent values from the original time series should be consistent except for a white Gaussian noise ω, called the *process noise*, with variance Q. The value of Q indicates the uncertainty of the process model. In other words, the larger Q is, the less likely the process is constant.

The noisy observation, which is obtained from the Laplace Perturbation mechanism, can be represented as follows:

$$z_k = x_k + \nu_k \ , \tag{12}$$

$$\nu_k \sim Lap(\Delta f / \alpha) \tag{13}$$

where ν is called the *measurement noise*. Clearly, the noisy observation z_k is the true state plus the perturbation noise. We adopt the following Gaussian approximation:

$$\nu_k \sim \mathcal{N}(0, R) \ , \quad R \propto (\Delta f)^2 / \alpha^2 \tag{14}$$

which has been shown in [18] to be computationally efficient and to minimize posterior estimation error.

We outline the Filtering procedure in Algorithm 1. It consists of two recursive operations: *prediction* and *correction*. At every time stamp k, the *prediction* step of the *Filtering* module generates a predicted value $\hat{x}_k^-$ as follows:

$$\hat{x}_k^- = \hat{x}_{k-1} \tag{15}$$

where $\hat{x}_{k-1}$ represents the released count at time $k-1$.

Upon receiving the perturbed count z_k, a posterior estimate can then be derived at the *correction* step based on the prediction as well as the noisy observation:

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-) \ . \tag{16}$$

The value K_k, called *Kalman Gain* is defined as

$$K_k = P_k^-(P_k^- + R)^{-1} \ . \tag{17}$$

and is adjusted with every measurement in order to minimize the posterior error. The detailed definitions and derivations

Algorithm 1 Filtering

Input: Noisy measurements $\mathbf{Z}= \{z_k\}$
Output: Released series $\mathbf{R}= \{r_k\}$

1: **for** each k **do**
2: *Prediction:* $\hat{x}_k^- = r_{k-1}$
3: $z_k \leftarrow$ Laplace Perturbation
4: *Correction:* $\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-)$
5: $r_k \leftarrow \hat{x}_k$

Table I
DATA SETS

Set	Mean	Standard Dev.	Trend	Seasonality
$s01$	90.2	33.3	Yes	Mild-None
$s02$	29.9	5.6	No	Medium
$s03$	1.19	5.76	No	Mild-None
$s04$	6	4.3	Yes	Very
$s07$	150	26.635	No	Mild-None
$s11$	301.1	78.8	Yes	Medium

as well as *Prediction* and *Correction* implementations are omitted here for brevity, as they can be found in [6], [18].

It is shown in [6] that releasing $\{\hat{x}_k\}$ rather than $\{z_k\}$ greatly improves the accuracy of the shared aggregates. We will empirically evaluate the utility of shared data in the context of outbreak detection in the next section.

V. EVALUATIONS

We implemented the proposed framework as well as three outbreak detection algorithms, i.e. C1, C2, and C3, in Java. All experiments were conducted using a 2.90 GHz Intel Core i7 PC with 8GB RAM. We used simulated data sets provided by CDC EARS[2], all simulating 6 years of daily count data. We chose 6 data sets, i.e. `s01-04`, `s07`, and `s11`, with different characteristics to evaluate our proposed solution. The descriptions of the 6 data sets are listed in Table I. From each data set, 10 iterations were used in the evaluation and the average utility results are reported.

Unless specified, the default parameter settings are as follows: $\alpha = 1$, $R \propto (\Delta f)^2/\alpha^2$, $Q[s01] = 100$, $Q[s02] = 100$, $Q[s03] = 0.001$, $Q[s04] = 1$, $Q[s07] = 100$, $Q[s11] = 100$. Note that the Q values are commonly derived by offline tuning and our settings may not be optimal.

A. Accuracy of Released Data

We first study the trade-off between privacy and accuracy in our proposed framework, in comparison to the baseline Laplace Perturbation algorithm (LPA) with default sensitivity $\Delta f = T$ and practical sensitivity $\Delta f = 2$. The LPA algorithms apply perturbation at every time stamp, and release perturbed values, i.e. $\{z_k\}$ as in Equation (12). In contrast, our solution adopts a tighter sensitivity bound, i.e. $\Delta f = 2$, and releases the posterior estimates, i.e. $\{\hat{x}_k\}$ as

[2]http://www.bt.cdc.gov/surveillance/ears/datasets.asp

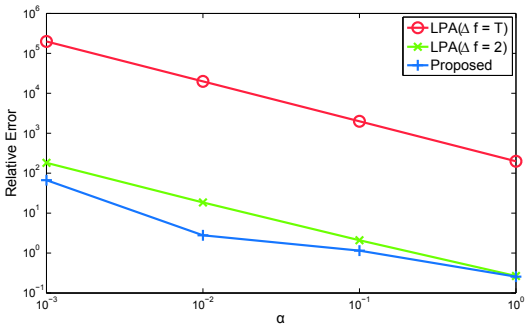


Figure 2. Privacy vs. Accuracy with `s04` Data Set

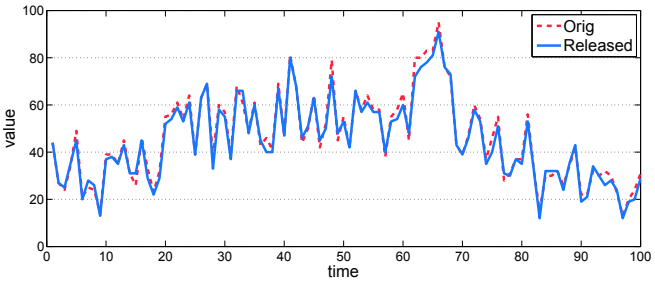


Figure 3. Original vs. Released Series with `s04` Data Set

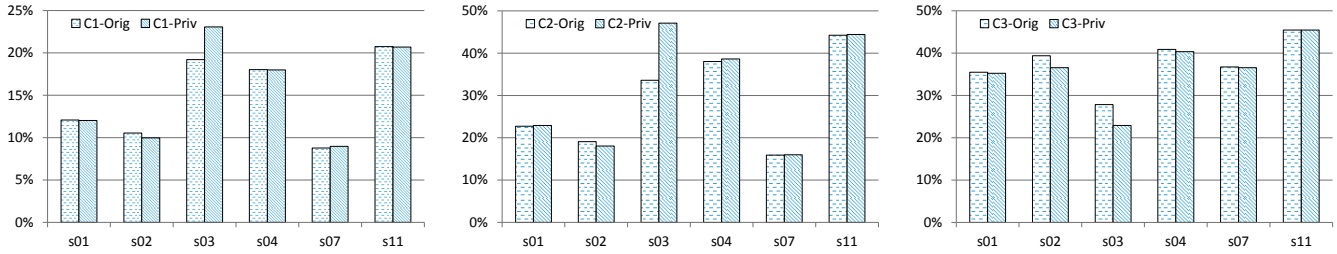
in Equation (16). We measure the accuracy of the released time series by average relative error, as proposed in [18]. We plotted the trade-off curves with data set `s04` in Figure 2 and note that other data sets have similar trends.

As we increase the privacy budget α from 0.001 to 1, all methods show improvement in accuracy of released data series, due to the reduced perturbation noise. As can be seen, methods that adopt practical sensitivity outperform baseline `LPA` with default sensitivity, reducing the relative error by several orders of magnitude. Moreover, our proposed solution constantly outperforms LPA($\Delta f = 2$). We conclude that FAST filtering techniques can improve the accuracy of released data on top of already reduced perturbation error.

We further examine the private, released series by our solution and compare it with the original `s04` data set. The released data series was generated with $\alpha = 1$. As is shown in Figure 3, the data series released by our privacy-preserving framework closely follows the original data line at all time stamps. This shows that our framework provides highly accurate data release while providing differential privacy guarantee.

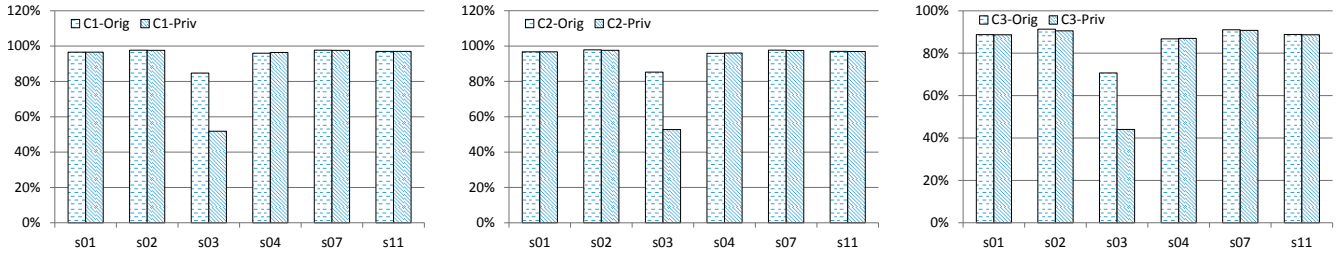
B. Outbreak Detection Evaluation

We study the usefulness of the private, released data series provided by our solution with EARS outbreak detection algorithms. We generated released data from all six data sets with the privacy parameter $\alpha = 1$ and ran `C1`, `C2`, and `C3` on both original data sets and private, released data sets. For `C1`, `C2`, and `C3` algorithms, we set $r_k = x_k$ if detection is performed on original data with no privacy preservation and



(a) C1 Algorithm (b) C2 Algorithm (c) C3 Algorithm

Figure 4. Sensitivity Performance with *Original* vs. *Private* Data



(a) C1 Algorithm (b) C2 Algorithm (c) C3 Algorithm

Figure 5. Specificity Performance with *Original* vs. *Private* Data

$r_k = \hat{x}_k$ if using our privacy-preserving framework. Same rules apply to the baseline counts.

Two common data mining metrics, i.e. *Sensitivity* and *Specificity*, are studied. Sensitivity, also referred to as recall, measures an algorithm's ability to identify positive results, while specificity relates to an algorithm's ability to identify negative results. A *true positive* is a correctly identified outbreak day while a *true negative* is a correctly identified non-outbreak day. Note that the ground truth is provided along with the simulated data sets. We summarize the utility results in Figure 4 and Figure 5.

As can be seen, using the data released by our privacy-preserving solution does not significantly degrade the performance of any outbreak detection algorithm. For instance in Figure 4(c), `C3` achieves 35.48% sensitivity with the original `s01` data set, and 35.25% with the private released data. We notice that with data set `s03`, where the original data values exhibit relatively high variation and no clear trend, using the private released data incurs more detected positives, i.e. higher sensitivity and lower specificity. We interpret this phenomenon as a result of model misfit, since the process model for the underlying time series is constant, as in Equation (10). As a future work direction, in-depth study about the dynamics of real data sets can be performed in order to establish an accurate internal model in the Filtering component.

VI. Conclusion and Discussions

We have presented a privacy-preserving framework for anomaly detection based on continual aggregate statistics. It provides differential privacy guarantee and adopts recently proposed FAST [6] filtering techniques to improve the accuracy of released aggregates. We demonstrated the usability of our framework and described the technical details of each component through a case study on epidemic outbreak detection. We further showed that a practical estimate for the global sensitivity can be derived with public survey statistics, which in both theory and practice reduces the level of perturbation noise inflicted by differential privacy mechanism. Empirical studies with simulated data sets confirmed that our solution greatly improves the accuracy of released data series compared to the baseline Laplace Perturbation algorithm (LPA). Evaluations on three widely used outbreak detection algorithms showed that our solution preserves the utility of released aggregates despite different data dynamics, enabling real-time anomaly detection while providing provable privacy guarantee.

Several research questions about differentially private anomaly detection remain open and would be interesting to investigate in the future. 1) The aggregate function at every time stamp may be highly sensitive, such as min and max, or the anomaly detection algorithm may require raw attribute values rather than aggregates, such as for outlier detection. It is shown in [18] that FAST filtering techniques can be generalized to share any aggregates supported by differential privacy. For anomaly detection, the domain specific sensitivity analysis for different types of aggregate functions is left open, which is needed to derive a tighter upper bound for the global sensitivity, especially when the time period T is long. 2) The privacy principal is not constant, i.e. needs to be determined given any anomaly detection task. In this paper, we addressed through a case study the problem of

privately detecting Influenza epidemic outbreak detections, where user-level privacy guarantee was provided. It can be relaxed to provide the event-level privacy guarantee, as in [3], [4]. However, for a different anomaly detection task which monitors the activities of each user individually, such as credit card fraud detection, only even-level privacy is suitable and can be provided by differential privacy. 3) A fundamental question is if a variant of differential privacy can be defined to allow for bounded or expected sensitivity, without excluding certain data contributors as we did. The rationale behind is that the worst case where $\Delta f = T$ rarely happens in reality, especially when T is large. 4) Can we decompose the monitoring period to shorter intervals and apply offline differentially private transformations to each segment of the aggregate series? In other words, shall we trade the timeliness of anomaly detection for accuracy?

We believe that anomaly detection with individual privacy preservation is an important direction for future research and plan to present our approach towards addressing some of the open questions in follow-up work.

ACKNOWLEDGMENT

This research is supported by NSF under grant CNS-1117763 and AFOSR DDDAS program under grant FA9550-12-1-0240.

REFERENCES

[1] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to non-interactive database privacy," in *STOC*, 2008.

[2] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *In Proceedings of the 3rd Theory of Cryptography Conference*. Springer, 2006, pp. 265–284.

[3] T.-H. H. Chan, E. Shi, and D. Song, "Private and continual release of statistics." in *ICALP (2)*, ser. Lecture Notes in Computer Science, S. Abramsky, C. Gavoille, C. Kirchner, F. M. auf der Heide, and P. G. Spirakis, Eds., vol. 6199. Springer, 2010, pp. 405–417.

[4] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," ser. STOC '10. New York, NY, USA: ACM, 2010, pp. 715–724.

[5] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *SIGMOD*, 2010.

[6] L. Fan and L. Xiong, "Real-time aggregate monitoring with differential privacy," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, ser. CIKM '12. New York, NY, USA: ACM, 2012, pp. 2169–2173.

[7] L. C. Hutwagner, W. W. Thompson, G. M. Seeman, and T. Treadwell, "A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines," *Statistics in Medicine*, vol. 24, no. 4.

[8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009.

[9] M. Jackson, A. Baer, I. Painter, and J. Duchin, "A simulation study comparing aberration detection algorithms for syndromic surveillance," *BMC Medical Informatics and Decision Making*, vol. 7, no. 1, p. 6, 2007.

[10] J. Caberera, B. Ravichandran, and R. Mehra, "Statistical traffic modeling for network intrusion detection," in *Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 2000. Proceedings. 8th International Symposium on*, 2000, pp. 466–473.

[11] C. for Disease Control and Surveillance. (2007) Early aberration reporting system.

[12] D. Hawkins and D. Olwell, *Cumulative Sum Charts and Charting for Quality Improvement*, ser. Information Science and Statistics Series. Springer Verlag, 1998.

[13] D. F. Stroup, G. D. Williamson, J. L. Herndon, and J. M. Karon, "Detection of aberrations in the occurrence of notifiable diseases surveillance data," *Statistics in Medicine*, vol. 8, no. 3, pp. 323–329, 1989.

[14] M. L. GATTON, L. A. KELLY-HOPE, B. H. KAY, and P. A. RYAN, "Spatial-temporal analysis of ross river virus disease patterns in queensland, australia," *The American Journal of Tropical Medicine and Hygiene*, vol. 71, no. 5, pp. 629–635, 2004.

[15] F. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," in *SIGMOD*, 2009.

[16] C. A. Bradley, H. Rolka, D. Walker, and J. Loonsk, "Biosense: implementation of a national early event detection and situational awareness system." *MMWR: Morbidity Mortality Weekly Report*, vol. 54, pp. 11 – 19, 2005.

[17] F.-C. Tsui, J. U. Espino, V. M. Dato, P. H. Gesteland, J. Hutman, and M. M. Wagner, "Technical description of rods: A real-time public health surveillance system," *Journal of the American Medical Informatics Association*, vol. 10, no. 5, pp. 399–408, 2003.

[18] L. Fan and L. Xiong, "An adaptive approach to real-time aggregate monitoring with differential privacy," *IEEE Transactions on Knowledge and Data Engineering*, vol. 99, no. PrePrints, p. 1, 2013.