Mobile Privacy-Preserving Crowdsourced Data Collection in the Smart City

Joshua Joy[*],Ciaran McGoldrick[†],Mario Gerla[‡]
UCLA
Email:[*]jjoy@cs.ucla.edu,[†]ciaran@cs.ucla.edu,[†]gerla@cs.ucla.edu

***Abstract*—Smart cities rely on dynamic and real-time data to enable smart urban applications such as intelligent transport and epidemics detection. However, the streaming of big data from IoT devices, especially from mobile platforms like pedestrians and cars, raises significant privacy concerns.**

Future autonomous vehicles will generate, collect and consume significant volumes of data to be utilized in delivering safe and efficient transportation solutions. The sensed data will, inherently, contain personally identifiable and attributable information - both external (other vehicles, environmental) and internal (driver, passengers, devices).

The autonomous vehicles are connected to the infrastructure cloud (e.g., Amazon), the edge cloud, and also the mobile cloud (vehicle to vehicle). Clearly these different entities must cooperate and interoperate in a timely fashion when routing and transferring the highly dynamic data. In order to maximise the availability and utility of the sensed data, stakeholders must have confidence that the data they transmit, receive, aggregate and reason on is appropriately secured and protected throughout. There are many different metaphors for providing end-to-end security for data exchanges, but they commonly require a management and control sidechannel.

This work proposes a scalable smart city privacy-preserving architecture that enables each node (e.g. vehicle) to divulge (contextually) local privatised data. The system is shown to scale gracefully to IoT scope deployments.

I. Introduction

Researchers are becoming increasingly interested in studying smart city behaviors, like pedestrians, drivers and traffic, city resources (e.g., energy) and city environment (e.g., pollution, noise). These studies are commonly based on Open Shared Data made available by several Smart City testbeds around the country. To this end, Open Data Science enables researchers to collect the data, analyze and process it with Data Mining and Machine Learning techniques and create accurate models that allow them to credibly validate smart city design methodologies. These systems enable the collection of data from sensors, cameras embedded in the "smart city" (e.g., smart building, smart transport, smart instrumented crowds) which can be used to derive models of behavior, predict trends, optimize system management and detect the onset of attacks.

There is now an increasing demand that research addressing these challenges be performed in more realistic environments. In other words, researchers will need to deploy their technologies in real vehicles, in real roads and cities (or, at smaller-scale, on-campus roads used for general purposes), to demonstrate that they are not mere simulation and pilot testbed toys and do scale to urban dimensions. As smart city research and systems require testing and validation in such uncontrolled environments, considerable attention must be paid to the validity of the experiments and the integrity and privacy of the data gathered through them. Since the experiments must be performed on massive scale in public places, it is prudent to anticipate malicious agents who either wish to make illicit use of the data gathered or seek to inject false data.

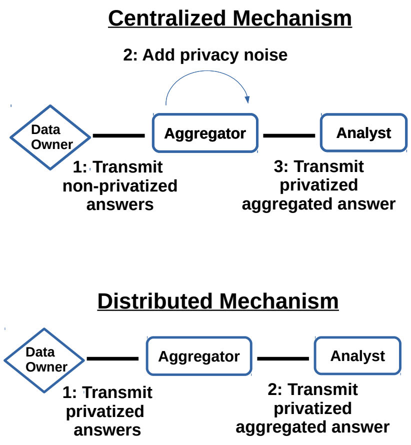
Centralized Mechanism
2: Add privacy noise
Data Owner
Aggregator
Analyst
1: Transmit non-privatized answers
3: Transmit privatized aggregated answer
Distributed Mechanism
Data Owner
Aggregator
Analyst
1: Transmit privatized answers
2: Transmit privatized aggregated answer

Figure 1: Illustrates that the data owner has control and consent over the privatization release as opposed to the centralized mechanism which requires strong trust assumptions regarding the aggregator adding differentially private noise.

As this research will have significant impact on the economy and safety of smart environments, the security challenges to realistic in-the-field experiments carried out in the smart city must be addressed. Rather than requiring each researcher working in the area to start afresh on securing his experimental infrastructure, there should be a common "privacy" infrastructure that all researchers can adopt. This approach has several advantages, including allowing researchers to concentrate on the issues they care about, rather than addressing issues of cybersecurity that, while important, are not relevant to the research questions they seek to answer. A shared cybersecure infrastructure can assure that security holds for a representative set of technologies and approaches, rather than requiring each project to perform a separate security evaluation.

In this work the vehicle example will be used to illustrate

	Output Pertubation	**Input Pertubation**	**Trusted Aggregator**	**Untrusted Aggregator**	**Pollution Protection**	**Anonymity**	**Stream Analytics**	**Sparse Datasets**
PINQ[3]	X		X					
PDDP[3]		X	X		X			X
SplitX[2]		X	X		X	X		X
Rappor[9]		X		X				
PCDH-LU[1]		X		X			X	
Safe Zones[11]	X		X				X	
Target Model		X		X	X	X	X	X

Table I: Privacy-Preserving Stream Analytics Design Space

secure and resilient IOT data gathering, as vehicles are extremely rich sensor platforms, run mature applications and touch many critical smart city issues, from citizens' welfare to energy and environment. In dealing with vehicles and their associated data, a central issue will be privacy e.g. of location data. Using techniques that go under the generic name of Differential Privacy [5], [7], [6], a vehicle that wishes to remain anonymous will obfuscate(fuzz) its position so that subsequent observations cannot reveal the exact origin and destination of the trip, yet permits analyzers to extract meaningful analysis from the aggregate data. The methodology is not restricted to vehicles, and can be extended to many other massively distributed urban computing observation platforms, from intelligent homes and smart energy grids to pedestrians and drones, etc.

The goal of this work is to develop an integrated architecture that allows resilient and secure gathering of Internet Of Things (IOT) data in the face of impairments and attacks that can originate at different layers of the network protocol stack. Moreover, privacy and confidentiality can be efficiently assured as required by the application. The role of the urban infrastructure (in the form of mobile cloud and edge cloud) will be elaborated. The architecture is realised and demonstrated in an important emerging IOT scenario, the autonomous vehicle.

II. Background

Table I gives an overview of various privacy-preserving properties of related work and our proposed target model. We now explain the desired properties and motivate our target architecture.

Perturbation techniques are classified into two basic categories: input perturbation and output perturbation [7].Input perturbation techniques are those where the underlying data are randomly modified, and answers to questions are computed using the modified data resulting in noisy answers. Output perturbation techniques are those where (correct) answers to queries are computed exactly from the real data, but noisy versions of these are reported.

Figure 1 demonstrates the difference between a trusted aggregator and untrusted aggregator.

A trusted aggregator sits between the data owners and the analysts, verifies and may modify the query, and returns the aggregate sum to the analysts [5], [7]. A trusted aggregator must blindly add noise and add the exact amount of noise (no more no less). If not enough differential private noise is added, the differential privacy guarantees do not hold breaking the privacy preserving guarantees. If too much diffferential private noise is added, the answers will be too noisy and not useful.

An untrusted aggregator computes and publishes some statistics based on the privatized answers received from the data owners. The data may live indefinitely as it is the data owner, rather than the aggregator, which adds the differential private noise. Thus, even if the aggregator were to be breached, no privacy violations would occur.

It's important that a single malicious adversary is unable to submit a single answer (e.g., very large answer) that substantially distorts the aggregate sum. Safeguards against this attack is commonly referred to as pollution protection.

Data owners also desire to remain anonymous within the set of all other participating data owners (anonymity set). The anonymity guarantee ensures that private answers submitted are unable to be linked back to the specific data owner by either the aggregator or a malicious adversary performing traffic analysis.

To ensure timely results support for stream analytics is desirable. This ensures that data is processed continuously and released on the order of minutes.

Finally, it is important that the accuracy of the results is able to scale from small datasets (e.g., small IoT deployments) to large scale IoT systems. This means that both sparse datasets and heavy hitter algorithms should be supported by the system.

Thus, it can be seen that our target model should include the desired properties of input pertubation, untrusted aggregator, pollution protection, anonymity, stream analytics, and sparse datasets. This ensures protection against data breaches at the aggregator, graceful scalability for large scale IoT deployments, and high accuracy.

III. Related

Privacy-preserving analytics has been an active area of research in recent years [11], [13], [1], [3], [8], [9], [16], [17].

In distributed data analytics systems [13], [3], [2], data owners maintain their personal data locally, and have exclusive control over their own data. These systems enable analysts to analyze data owners' locally maintained data in a distributed manner under the differential privacy guarantees. However, the systems [3] rely on expensive public-key crypto systems which do not scale well in a large-scale IoT environment and are not suitable for stream analytics. Subsequent system [13] addresses the scaling issue by avoiding using expensive crypto systems. It, however, suffers from the pollution attack where even a single malicious data owner can substantially distort the query result without detection. The recent system [2] defeats the pollution attack, however, it involves some level of synchronization among system components preventing its usage for stream analytics. Nevertheless, all these systems require trusted servers, and deal with only the "one-shot" queries whereby data owners' data is assumed to be static and the aggregate result remains unchanged during the course of the query execution.

Data analysts care about timeliness as it is common that user data is constantly changing. It is crucial that analysts can analyze user data continuously. Dwork et al. [8] proposed one of the first proposals to make the transition from differentially private static analytics to stream analytics. This system introduces the concept of pan-privacy. Data owners reveal their data relying on the pan-private mechanism to defend against intrusions. However, this system updates the query result only after data owners' data changes significantly, and does not support stream analytics over an unlimited time period.

Chan et al. [1] operate with an untrusted aggregator, and support differentially private continual monitoring of multiple distributed streams over a sliding window. Similarly, Friedman et al. [11] propose a concept of safe zones, and enable the monitoring of arbitrary functions over statistics derived from distributed data streams in a differentially private manner. However, these systems are tailored for heavy-hitter monitoring only (i.e., they can only report on a fraction of the data), and require some form of synchronization which limits their scalability.

Recently, RAPPOR [9] enables analysts to collect various types of statistics from a large number of data owners' devices in such a way that provides the differential privacy guarantees for individual data owners. However, RAPPOR is designed for heavy-hitter collection, and does not deal with the situation where data owners' answers to the same query are changing over time. Therefore, RAPPOR does not fit well with the IoT stream analytics.

IV. Goals

We propose to enable scalable and privacy-preserving data analytics for personal identifiable information (PII) that protects individuals' privacy yet enables analysts to leverage available data to improve urban scenarios, healthcare, traffic regulation (congestion avoidance, accident prevention), and crime control (locating lost/stolen vehicles). Vehicular networks are targeted as a challenging exemplar. Our vision is to tap into underutilized real time data - both vehicular (personal (seatbelt use, car seat use, etc) and public (pollution, traffic condition, etc.)).This data is continuously generated and (generally) cannot be stored in central cloud infrastructures for two reasons – users' privacy concerns, and the volume and timeliness requirements from the generated data.

Trust Assumptions. Analysts are potentially malicious. Analysts may attempt to de-anonymize data owners, link data owners' query requests and answers, or remove the differential private (*DP*) noise added to query answers. Data owners are also potentially malicious. They may generate false or illegitimate answers in an attempt to distort the query results learned by analysts. We assume that there are no Sybil data owners (i.e., data owner create multiple false identities) and delegate to other procedures (e.g., vehicular network security based on public-key infrastructure) for the elimination of Sybil nodes/identities.

Finally, the aggregators are also potentially malicious. Alternatively, they may malfunction and not follow the specified protocols faithfully. This may invalidate the query results, but will not be able to violate the privacy of individual users due to the utilization of input perturbation. Prior work must synchronize the addition of differentially private noise [2] and must be trusted to safeguard the unprivatized data and properly release the privatized answers faithfully [3], [2], leaving these prior systems vulnerable to data breaches.

Performance Properties Current state-of-the-art systems provide differential privacy guarantees. However, they fall short in applications with "one-shot" queries. One-shot query systems have strong and unrealistic assumptions – user data is assumed to be static and the aggregate result is assumed unchanged spanning hours and days. Recently, data analysts have begun to care about timeliness, especially when user data is constantly changing. This is particularly true in mobile networks where the mobiles (eg. vehicles) may upload new data every minute.

The privacy-preserving software should scale well both on the data owners' mobile devices as well as on the server system components (i.e., the aggregator). The data owner-side operations should not incur much overhead to support even the most resource-constrained devices, e.g., mobile phones and low-capacity sensors. The server-side operations should be efficient enough to continuously support a large data owner base with millions of data owners on the order of minutes; additionally, the servers should tolerate data owner churn as data owners may be intermittently disconnected.

Privacy Properties. Data owners' personal data resides on their own devices. Data owners have exclusive control over their own data, and are able to determine whether or not to answer queries over their own data.

The system should satisfy the local differential privacy mechanism as well as provide anonymous guarantees robust to traffic analysis. The local differential privacy mechanism ensures that each data owners response should be independently generated such that a single data owner's response does not rely on the addition of differentially private noise from another data owner. Thus, even if an active attack were to be made against a single data owner or the aggregator, the differential privacy guarantee holds true. The anonymity property ensures that any answer submitted is unable to be linked to a single data owner.

V. Architecture

A. System Model

The system under study is a mobile IoT system that collects data samples from the environment (external) and from the data owner (internal). Personal data is stored in each device and privately released upon the device answering a query from analysts. Privacy is a concern, but so is real time response

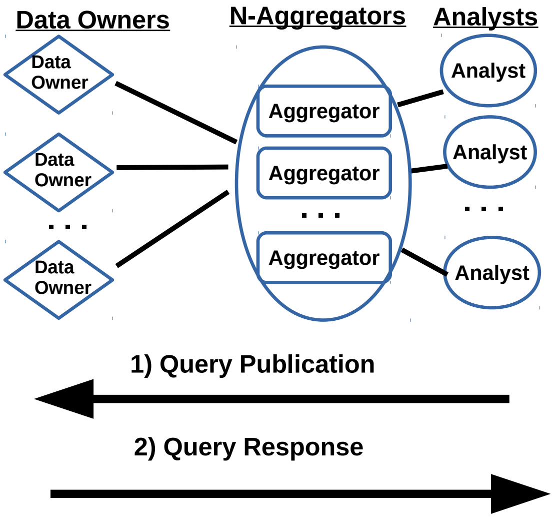

Figure 2: System architecture. Data owners privately and anonymously respond in real-time to the analysts? queries. The privatized responses are continuously aggregated into small batches and then transmitted to each corresponding analyst.

and quality of information. Privacy, latency and utility are the main focus of the study. The system model consists of three components - IoT devices, aggregators, and analysts as seen in Figure 2.

The mobile IoT devices considered in our study are vehicles, because vehicles are the mobile agent(s) with the richest assortment of on board sensors, with efficient communications to peers and to infrastructure and with (effectively) unlimited power supply. Vehicles can communicate either with other moving vehicles or with fixed network nodes placed alongside the road, commonly referred to as road-side units (RSUs). The typical use of an RSU is to provide moving vehicles with access to an infrastructure network, as well as infrastructure-based services.

While there have been various prior work which proposed privacy mechanisms [18], [15], [7], we choose differential privacy (*DP*) as differential privacy has gained broad acceptance. Recent data analytics systems achieve *DP* guarantees in a distributed setting [6], [13], [3], [2] where each data owner holds their own personal data, rather than in a centralized database. Specifically, in these systems, the query answers are first generated at each data owner's own device independently, and then the *DP* noise is added to the aggregate answer (either collaboratively by the data owners or by a centralized aggregator). However, existing distributed systems require strong trust assumptions regarding the aggregation mechanism and require expensive zero-knowledge proofs to defend against pollution attacks (e.g., where one data owner can distort the aggregate sum with a single answer).

We are interested in Distributed *DP* (local differential privacy) mechanisms whereby each data owner holds their own data and is independently responsible for adding differentially private noise. By satisfying local *DP*, strong trust assumptions are removed, pollution attacks are thwarted, and data owners are guaranteed *DP* regardless of the amount of noise added by other data owners or system components.

1) **Analysts.** An analyst formulates queries, distributes the queries to data owners via the aggregator, and receives the query results from the aggregator. Multiple analysts can operate in the system simultaneously. Each query is signed by the analyst. Examples of an analyst may be the *Los Angeles Department of Transportation*, *Center for Disease Control*, or *National Institutes of Health*.
2) **Data Owners.** A data owner is a user device or agent under the user's control. It stores the continuously-generated personal user data locally. Each data owner device requests queries from specific analysts, formulates answers to these queries using the data owners' local data, and responds with privatized and anonymized answers. All the query requests and answers are anonymized and are TLS transmitted via aggregators.
3) **Aggregator.** The data owners transmit their privatized and anonymized answers to the aggregator via TLS. The aggregator interfaces with analysts, aggregates the small batches of responses, and provides the privacy-preserving stream analytics service.

Additionally, it should be noted that the anonymity scheme still provides a form of accountability that prevents Sybil attacks. All anonymous users will be authenticated, for example using PKI (managed by the analysts for example), such that only authenticated data owners are allowed to upload data. All data owners form an anonymity set, such that a single data owner is anonymous with the set of all other uploading data owners. The proposed privacy preserving solutions are an integral part of the overall architecture and offer greatly reduced processing overhead and latency.

Query Formulation. An example query may be "what is the distribution of real-time speed across the Los Angeles metro area?". Here, the analyst can define 22 indexes on speed: '0', '1∼10', '11∼20', ⋯, '181∼190', '191∼200', and '>200'. If a vehicle is moving at 15 mph, the data owner on that vehicle answers '1' for the third index, and '0' for all others.

Thus, each query specifies a few fields as follows:

$$Query := \langle Query_{ID}, Analyst_{ID}, Sensors_{1\to n}, p, q, epoch, T_{end} \rangle \tag{1}$$

Each query is uniquely identified by the query *ID* and analyst *ID*. The query is formulated as an *n*-dimensional bit vector over the IoT sensors.

$Sensors_{1\to n}$ denotes the n sensors to be monitored. Upon querying each sensor, the retrieved sensor value is then matched against the formulated query indexes. Each data owner's answer is in the form of a '1' or '0' per index, depending on whether or not the answer falls within that range. Different ranges should not overlap.

p and q denote the two coin flipping probabilities used in the randomized response mechanism later at data owners. They describe the probabilities of the first coin and the second coin

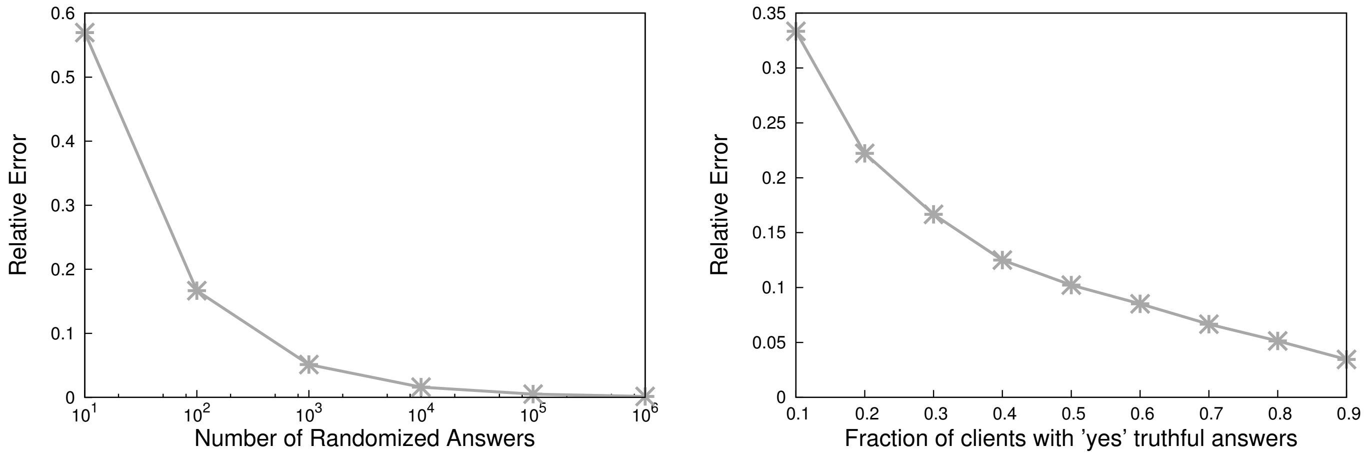

Figure 3: The two coin flipping parameters p and q are set to achieve the highest possible utility. (a) The fraction of the population with the sensitive attribute is 80%. (b) We also test with varying fractions having the sensitive attribute (different fractions of original "yes" answers), and the results follow the same pattern. The total number of data owners is 1000.

coming up heads. The analyst requests a certain utility (express by p,q) and the data owners may refuse if too risky.

epoch denotes the execution interval of the query, i.e., how often the query needs to be executed at a data owner. T_{end} denotes the query end time, i.e., when answers to the query will no longer be accepted.

Sanity Check. The data owner?s device checks all the queries received from analysts and discards any query if its privacy cost ε (computed based on the query's p and q) exceeds a certain configurable threshold. The data owner is also able to restrict query execution based on a personal sensor blacklist. Analyst reputation may also be used to blacklist certain analyst queries from being executed.

B. Privacy Mechanism

Randomized response [19] was originally created by social scientists as a privacy mechanism to allow individuals to answer sensitive questions without providing truthful answers all the time yet still allow analysts to collect meaningful statistical user data. Utilizing the randomized response method, data owners locally randomize their truthful answers to analysts' sensitive queries, and respond only with the privatized (locally randomized) answers to the analysts. Randomized response mechanism also satisfies the differential privacy guarantee as well as provides the optimal sample complexity for local differential privacy mechanisms [4].

There are many different randomized response mechanisms in the literature. In this section, we present only the mechanism described in [10] because it offers a superior balance between the utility and the privacy guarantee of randomized responses, as compared to other mechanisms [19], [14], [12].

a) Mechanism Description: Suppose each data owner has two independent coins. With coin flipping, the first coin comes up heads with probability p, and the second coin comes up heads with probability q. Without loss of generality, in this paper, heads is represented as "yes" (i.e., 1), and tails is represented as "no" (i.e., 0).

The data owner flips the first coin. If it comes up heads, the data owner responds with the truthful answer; otherwise, the data owner flips the second coin and reports the result of this second coin flipping. Eventually, if there are N randomized answers where Y_R of them are "yes", then the estimated number of the original truthful "yes" answers (i.e., without randomization), Y_E, can be computed as:

$$Y_E = \frac{Y_R - (1-p) \times q \times N}{p} \tag{2}$$

The intuition behind randomized response is that it provides "plausible deniability", i.e., any truthful answer can produce a response either "yes" or "no", and data owners retain strong deniability for any answers they respond. If the first coin always comes up heads, there is high utility yet no privacy. Conversely, if the first coin is always tails, there is low utility though strong privacy.

1) Utility of Randomized Response: Suppose the real and the estimated counts of data owners' truthful "yes" answers are $\tilde{Y}$ and Y_E, respectively. Then, the utility is defined as the relative error η — the magnitude of the difference between the real count and the estimated count, divided by the magnitude of the real count.

$$\eta = |\frac{Y - Y_A}{Y}| \tag{3}$$

Here, smaller relative error η means higher utility of the randomized responses, and vice versa.

VI. Evaluation

Privacy systems must carefully balance the trade-off between privacy and utility. Increasing privacy causes a corresponding decrease in utility and vice-versa. In distributed systems it becomes more difficult to balance this trade-off since there is no centralized system to add the minimum differentially private noise required.

The system evaluation involves varying the numbers and fractions of data owners participating. In general, we seek to show that as more data owners participate the utility improves quickly due to the law of large numbers. This is particularly suitable for stream analytics whereby we can reasonably expect a large amount of data.

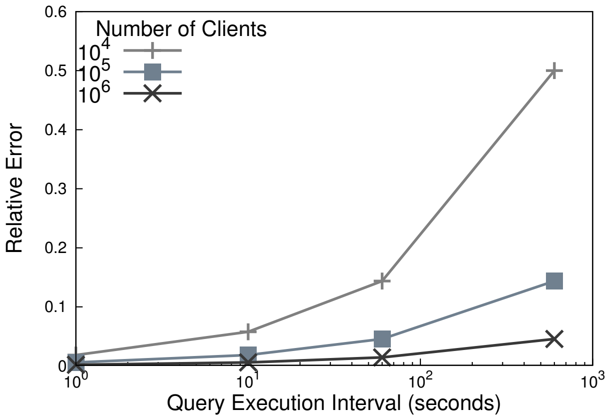

Figure 4: Utility of query results. Here, the utility is measured by the query result's relative error.

Figure 4 shows that, in a typical large-scale distributed IoT application scenario with one million devices each answering a query every 10 seconds, the system can produce query results every one second with the relative error smaller than 0.5%. As compared to prior systems [11], [1] we achieve both strong local privacy guarantees as well as high utility without requiring synchronization amongst system components or devices as each device independently privatizes their answer. The experimental results also shows that the utility improves with the increase of the number of devices, as well as with the decrease of the query execution interval. This is because increasing the number of devices or reducing the query execution interval will put more answers into a batch. As implied in Figure 3, more answers in a batch lead to a lower relative error (i.e., higher utility). In certain scenarios, if there are just a few devices or the devices answer queries less often, the system needs to extend the batching interval to put more answers into each batch for achieving high utility, although this may incur longer latency.

VII. Conclusion

In this paper, a highly scalable, efficient, privacy preserving Differential Privacy scheme is introduced for use in future and emerging IoT and data stream analytics scenarios. The scheme privatizes the data on the source device, thereby obviating many traditional attack and compromise modalities. Moreover the computational complexity of the system is low, ensuring tractable implementation in both low-power IoT scenarios, as well as more capable aggregator nodes. The system premise and functionality are validated through experimentation.

References

[1] Chan, T. H., Li, M., Shi, E., and Xu, W. Differentially private continual monitoring of heavy hitters from distributed streams. In *Privacy Enhancing Technologies - 12th International Symposium, PETS 2012, Vigo, Spain, July 11-13, 2012. Proceedings* (2012), S. Fischer-Hübner and M. K. Wright, Eds., vol. 7384 of *Lecture Notes in Computer Science*, Springer, pp. 140–159.

[2] Chen, R., Akkus, I. E., and Francis, P. Splitx: high-performance private analytics. In *ACM SIGCOMM 2013 Conference, SIGCOMM'13, Hong Kong, China, August 12-16, 2013* (2013), D. M. Chiu, J. Wang, P. Barford, and S. Seshan, Eds., ACM, pp. 315–326.

[3] Chen, R., Reznichenko, A., Francis, P., and Gehrke, J. Towards statistical queries over distributed private user data. In *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2012, San Jose, CA, USA, April 25-27, 2012* (2012), S. D. Gribble and D. Katabi, Eds., USENIX Association, pp. 169–182.

[4] Duchi, J. C., Wainwright, M. J., and Jordan, M. I. Local privacy and minimax bounds: Sharp rates for probability estimation. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.* (2013), C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, Eds., pp. 1529–1537.

[5] Dwork, C. Differential privacy. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II* (2006), M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Eds., vol. 4052 of *Lecture Notes in Computer Science*, Springer, pp. 1–12.

[6] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings* (2006), S. Vaudenay, Ed., vol. 4004 of *Lecture Notes in Computer Science*, Springer, pp. 486–503.

[7] Dwork, C., McSherry, F., Nissim, K., and Smith, A. D. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings* (2006), S. Halevi and T. Rabin, Eds., vol. 3876 of *Lecture Notes in Computer Science*, Springer, pp. 265–284.

[8] Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. Differential privacy under continual observation. In *STOC* (2010).

[9] Erlingsson, Ú., Pihur, V., and Korolova, A. RAPPOR: randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014* (2014), G. Ahn, M. Yung, and N. Li, Eds., ACM, pp. 1054–1067.

[10] Fox, J. A., and Tracy, P. E. *Randomized response: a method for sensitive surveys*. Beverly Hills California Sage Publications, 1986.

[11] Friedman, A., Sharfman, I., Keren, D., and Schuster, A. Privacy-preserving distributed stream monitoring. In *21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, California, USA, February 23-26, 2014* (2014), The Internet Society.

[12] Greenberg, B. G., Abul-Ela, A.-L. A., Simmons, W. R., and Horvitz, D. G. The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association 64*, 326 (1969), 520–539.

[13] Hardt, M., and Nath, S. Privacy-aware personalization for mobile advertising. In *the ACM Conference on Computer and Communications Security, CCS'12, Raleigh, NC, USA, October 16-18, 2012* (2012), T. Yu, G. Danezis, and V. D. Gligor, Eds., ACM, pp. 662–673.

[14] Kuk, A. Y. Asking sensitive questions indirectly. *Biometrika 77*, 2 (1990), 436–438.

[15] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. In *Proceedings of the 22nd International Conference on Data Engineering, ICDE 2006, 3-8 April 2006, Atlanta, GA, USA* (2006), L. Liu, A. Reuter, K. Whang, and J. Zhang, Eds., IEEE Computer Society, p. 24.

[16] Rastogi, V., and Nath, S. Differentially private aggregation of distributed time-series with transformation and encryption. In *SIGMOD* (2010).

[17] Shi, E., Chan, T. H., Rieffel, E. G., Chow, R., and Song, D. Privacy-preserving aggregation of time-series data. In *NDSS* (2011).

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

[19] Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association 60*, 309 (1965), 63–69.