Differentially Private Noisy Search with Applications to Anomaly Detection (Abstract)

Daniel M. Bittner
Rutgers University
dbittner@cs.rutgers.edu

Anand D. Sarwate
Rutgers University
anand.sarwate@rutgers.edu

Rebecca N. Wright
Rutgers University
rebecca.wright@rutgers.edu

We consider the problem of privacy-sensitive anomaly detection—screening to detect individuals, behaviors, areas, or data samples of high interest. What defines an anomaly is context-specific; for example, a spoofed rather than genuine user attempting to log in to a web site, a fraudulent credit card transaction, or a suspicious traveler in an airport. The unifying assumption is that the number of anomalous points is quite small with respect to the population, so that deep screening of all individual data points would potentially be time-intensive, costly, and unnecessarily invasive of privacy. Such privacy violations can raise concerns due sensitive nature of data being used, raise fears about violations of data use agreements, and make people uncomfortable with anomaly detection methods. Anomaly detection is well studied, but methods to provide anomaly detection along with privacy are less well studied. Our overall goal in this research is to provide a framework for identifying anomalous data while guaranteeing quantifiable privacy in a rigorous sense. Once identified, such anomalies could warrant further data collection and investigation, depending on the context and relevant policies.

In this research, we focus on privacy protection during the deployment of anomaly detection. Our main contribution is a differentially private access mechanism for finding anomalies using a search algorithm based on adaptive noisy group testing. To achieve this, we take as our starting point the notion of *group testing* [1], which was most famously used to screen US military draftees for syphilis during World War II. In group testing, individuals are tested in groups to limit the number of tests. Using multiple rounds of screenings, a small number of positive individuals can be detected very efficiently. Group testing has the added benefit of providing privacy to individuals through plausible deniability—since the group tests use aggregate data, individual contributions to the test are masked by the group. We follow on these concepts by demonstrating a search model utilizing adaptive queries on aggregated group data.

Our work takes the first steps toward strengthening and formalizing these privacy concepts by achieving *differential privacy* [2]. Differential privacy is a statistical measure of disclosure risk that captures the intuition that an individual's privacy is protected if the results of a computation have at most a very small and quantifiable dependence on that individual's data. In the last decade, there has been an explosion of research in differential privacy, with recent practical adoption underway by high-profile companies such as Apple, Google, and Uber.

In order to make differential privacy meaningful in the context of a task that seeks to specifically identify some (anomalous) individuals, we introduce the notion of anomaly-restricted differential privacy. Using ideas from information theory, we show that noise can be added to group query results in a way that provides differential privacy for non-anomalous individuals and still enables efficient and accurate detection of the anomalous individuals. Our method ensures that using differentially private aggregation of groups of points, providing privacy to individuals within the group while refining the group selection to the point that we can probabilistically narrow attention to a small numbers of individuals or samples for further attention. To summarize:

- We introduce a new notion of anomaly-restriction differential privacy, which may be of independent interest.
- We provide a noisy group-based search algorithm that satisfies the anomaly-restricted differential privacy definition.
- We provide both theoretical and empirical analysis of our noisy search algorithm, showing that it performs well in some cases, and exhibits the usual privacy/accuracy tradeoff of differentially private mechanisms.

Potential anomaly detection applications for our work might include spatial search for outliers: this would rely on new sensing technologies that can perform queries in aggregate to reveal and isolate anomalous outliers. For example, this could lead to privacy-sensitive methods for searching for outlying cell phone activity patterns or Internet activity patterns in a geographic location.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Homeland Security under award 2009-ST-061-CCI002 and contract HSHQDC-16-A-B0005/HSHQDC-16-J-00371, the Defense Advanced Research Projects Agency (DARPA) and Space and Naval Warfare Systems Center, Pacific (SSC Pacific) under contract No. N66001-15-C-4070, and the National Science Foundation under award 1453432.

REFERENCES

[1] Robert Dorfman. 1943. The Detection of Defective Members of Large Populations. *The Annals of Mathematical Statistics* 14, 4 (December 1943), 436–440. http://www.jstor.org/stable/2235930

[2] Cynthia Dwork, Frank Mcsherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC '06: In Proceedings of the 3rd Theory of Cryptography Conference*. 265–284.

AISec'17, November 3, 2017, Dallas, TX, USA

ACM ISBN 978-1-4503-5202-4/17/11.
https://doi.org/10.1145/3128572.3140456