SEPIA: Aggregation of Network Measurements Using Multiparty Computation

Martin Burkhart*, Mario Strasser, Dilip Many*, Xenofontas Dimitropoulos
ETH Zurich, Switzerland, (*student authors)
{burkhart, strasser, dmany, fontas}@tik.ee.ethz.ch

A number of network security and monitoring problems can substantially benefit if a group of involved organizations aggregates private data to jointly perform a computation. For example, aggregation of private data is useful for alert signature extraction, collaborative anomaly detection, multi-domain traffic engineering, detecting traffic discrimination, and collecting network performance statistics. However, all these approaches face a delicate privacy versus utility trade-off. Some private data typically have to be revealed, which prohibits the acquisition of many data providers, while data anonymization, used to remove sensitive information, complicates or even prohibits developing good solutions [3]. Moreover, the ability of anonymization techniques to effectively protect privacy is questioned by recent studies [2]. One possible solution to this privacy-utility tradeoff is multiparty computation (MPC).

For almost thirty years, MPC [4] techniques have been studied for solving the problem of jointly running privacy-preserving computations on data distributed among multiple organizations. However, MPC techniques are typically impractical in terms of computation and communication cost. For this reason, they have mainly attracted theoretical interest in the last decades and only recently a real-world sugar-beet auction [1] was demonstrated. Adopting MPC techniques to network monitoring and security problems introduces the additional challenge of dealing with voluminous input data that require processing in *near real-time*. This is not presently possible with existing general-purpose MPC frameworks.

Therefore, we design, implement, and evaluate SEPIA, a library for efficiently aggregating multi-domain network data using MPC in the semi-honest adversary model. The foundation of SEPIA is a set of optimized MPC operations, implemented with performance of parallel execution in mind. On top of these comparison operations, we design and implement novel MPC protocols tailored for network security and monitoring applications. The *event correlation* protocol identifies events that occur frequently in multiple domains. The protocol is generic having several applications, for example, in alert correlation or in identification of multi-domain network traffic heavy-hitters. In addition, we introduce SEPIA's *entropy* and *distinct count* protocols that compute the entropy of traffic feature distributions and find the count of distinct feature values, respectively. These metrics are used frequently in traffic analysis applications, e.g., in network anomaly detection. We implement these protocols along with a vector addition protocol to support additive operations on timeseries and histograms.

Our evaluation of SEPIA's performance shows that SEPIA protocols run in near real-time with 5-minute windows, with up to 140 input providers and 9 computation nodes. Compared to implementations using existing general-purpose MPC frameworks, our protocols are significantly faster requiring, for example, 3 minutes for a task that takes 2 days with a general-purpose framework. Moreover, we run SEPIA on traffic data of 17 networks collected during the global Skype outage in August 2007 and show how the networks can use SEPIA to troubleshoot and timely detect such anomalies. Finally, we discuss novel applications in network security and monitoring that SEPIA enables.

References

[1] P. Bogetoft, D. Christensen, I. Damgård, M. Geisler, T. Jakobsen, M. Krøigaard, J. Nielsen, J. Nielsen, K. Nielsen, J. Pagter, et al. Secure multiparty computation goes live. In *Financial Cryptography*, 2009.

[2] P. Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *57 UCLA Law Review*, 2010. Available at `http://ssrn.com/abstract=1450006`.

[3] P. Porras and V. Shmatikov. Large-scale collection and sanitization of network security data: risks and challenges. In *Workshop on New security paradigms (NSPW)*, 2006.

[4] A. Yao. Protocols for secure computations. In *IEEE Symposium on Foundations of Computer Science*, 1982.

Index Termsâ€”secure multiparty computation, performance, measurement. 1. Introduction. The main goal of Secure Multiparty Computation is to allow several parties calculating a joint function. The corresponding inputs of each party are kept private and no Trusted Third Party should be required for the calculation.Â In this subsection we introduce the secure multiparty computation framework, which we used for our tests. Then software under test is introduced. Here, we adjusted an demo application, which is part of the framework, and those adjustments are explained.Â SEPIA: Privacy-preserving Aggregation of Multi-domain Network Events. and Statistics. Proceedings of the 19th USENIX Conference on Security, p. 15, 2010. It sometimes makes sense to use multiparty secure computation for cryptographic computations as well. For example, there might be several reasons why we would want to â€œsplitâ€ a secret key between several parties, so no party knows it completely. Some proposals for key escrow (giving government or other entity an option for decrypting communication) suggested to split a cryptographic key between several agencies or institutions (say the FBI, the courts, etc..) so that they must collaborate in order to decrypt communication, thus hopefully preventing unlawful access.Â Similarly multiparty secure computation can be used to achieve distributed cryptography with finer access control mechanisms. Proving the fundamental theorem Security is guaranteed by secure multiparty computation protocols using well known secu-rity schemes: Shamirâ€™s secret sharing, perturbation-based, and various encryption schemes. Dierential privacy of the nal result is achieved by distributed Laplace perturbation mechanism (DLPA). Partial random noise is generated by all participants, which draw random variables from Gamma or Gaussian distributions, such that the aggregated noise fol-lows Laplace distribution to satisfy dierential privacy.Â SEPIA: Privacy-preserving aggregation of multi-domain network events and statistics. In 19th USENIX Security Symposium, August 2010.Â Ecient and provably secure aggregation of encrypted data in wireless sensor networks. ACM Trans. Sen. Netw., 5(3):20:1â€“20:36, June 2009. Secure multiparty computation goes live. In Financial Cryptography and Data Security, FCâ€™09, pages 325â€“343. Springer, 2009.Â Sepia: Privacy-preserving aggregation of multi-domain network events and statis-tics. In USENIX Security SYMPOSIUM. USENIX, 2010.Â In Passive and Active Network Measurement, volume 5448 of Lecture Notes in Computer Science, pages 229â€“238. Springer, 2009. 18. Jaideep Vaidya and Chris Clifton.