Computer Networks 57 (2013) 3728–3742

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Elementary secure-multiparty computation for massive-scale collaborative network monitoring: A quantitative assessment

A. Iacovazzi [a,*], A. D'Alconzo [b], F. Ricciato [c], M. Burkhart [d]

[a] DIET, Sapienza Universiy of Rome, Via Eudossiana 18, 00184 Rome, Italy
[b] FTW – Forschungszentrum Telekommunikation Wien, Donau-City-St. 1, 1220 Vienna, Austria
[c] DII, University of Salento, Campus Ecotekne, Via per Monteroni, 73100 Lecce, Italy
[d] Department of Computer Science, ETH, Universitätstrasse 6, 8092 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 10 July 2012
Received in revised form 1 July 2013
Accepted 21 August 2013
Available online 4 September 2013

Keywords:
Secure multi party computation
Cooperative traffic monitoring
Applied cryptography
Privacy

ABSTRACT

Recently, Secure-Multiparty Computation (SMC) has been proposed as an approach to enable inter-domain network monitoring while protecting the data of individual ISPs. The SMC family includes many different techniques and variants, featuring different forms of "security", i.e., against different types of attack (er), and with different levels of computation complexity and communication overhead. In the context of collaborative network monitoring, the rate and volume of network data to be (securely) processed is massive, and the number of participating players is large, therefore *scalability* is a primary requirement. To preserve scalability one must sacrifice other requirement, like *verifiability* and *computational completeness* that, however, are not critical in our context. In this paper we consider two possible schemes: the Shamir's Secret Sharing (SSS), based on polynomial interpolation on prime fields, and the Globally-Constrained Randomization (GCR) scheme based on simple blinding. We address various system-level aspects and quantify the achievable performance of both schemes. A prototype version of GCR has been implemented as an extension of SEPIA, an open-source SMC library developed at ETH Zurich that supports SSS natively. We have performed a number of controlled experiments in distributed emulated scenarios for comparing SSS and GCR performance. Our results show that additions via GCR are faster than via SSS, that the relative performance gain increases when scaling up the data volume and/or number of participants, and when network conditions get worse. Furthermore, we analyze the performance degradation due to sudden node failures, and show that it can be satisfactorily controlled by containing the fault probability below a reasonable level.

1. Introduction and motivations

Since its inception the Internet has been exposed to global threats: spam, large-scale malware infections, DDoS attacks and botnets are all examples of global phenomena insensitive to any administrative network boundary. Besides threats, the popularity of global Over-The-Top (OTT) services and peer-to-peer applications has increased the risk of "global failures" that impact customers and networks of multiple ISPs, e.g., like the worldwide Skype outage in 2007 [1,2].

Despite the global nature of threats and failures, the operation and management of the network infrastructure remains almost entirely localized within each ISP's domain, and so do the detection, prevention and reaction processes. The contrast between global problems and local response plays heavily in favor of the former. Most operators concede that some degree of coordination (and

* Corresponding author. Tel.: +39 0644585365; fax: +39 064744481.
E-mail addresses: alfonso.iacovazzi@uniroma1.it (A. Iacovazzi), a.dalconzo@ftw.at (A. D'Alconzo), fabio.ricciato@unisalento.it (F. Ricciato), burkhart@tik.ee.ethz.ch (M. Burkhart).

1389-1286/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comnet.2013.08.017

collaboration) across ISPs, at least in the stage of detecting and diagnosing the problem, would be highly beneficial. The simplest use-case would be to enable each ISP to complement the detailed view of its own “internal” network, obtained by the local monitoring process, with a condensed view of the “external” situation. The combination of the two views would improve the effectiveness of the alarming and troubleshooting process along several dimensions: lower rates of false positives, lower delay, lower cost. More advanced forms of inter-domain collaboration could involve sharing malware information (e.g., with a newly learned signature) or the coordinated activation of local countermeasures (e.g., new firewall rules).

In order to be accepted by ISPs any form of collaborative model must fulfill some fundamental requirements. First, ISPs will not share their raw data due to business sensitivity and/or user privacy regulations. Second, they will want to preserve their anonymity when it comes to disclosing information about critical events that have impacted their domain like failures and/or attacks.

Recently, Secure-Multiparty Computation (SMC) has been proposed as an approach to enable inter-domain network monitoring while protecting the data of individual ISPs [3,4]. With SMC the collaboration paradigm shifts from “local computation on shared data” to “shared computation on local data”. The SMC family includes many different techniques and variants, featuring different forms of “security”, i.e., against different types of attack (er) and with different levels of computation complexity and communication overhead. In the context of collaborative network monitoring, the rate and volume of network data to be (securely) processed is massive, and the number of participating players might be large, therefore *scalability* is a primary requirement. To preserve scalability one must sacrifice other requirements, like *verifiability* and *computational completeness* that, however, do not appear to be critical in this context. In fact, since SMC players map to ISPs, it is reasonable to exclude the presence of “active attackers” and assume that all players follow the “honest-but-curious” model. Therefore, we restrict the focus onto non-verifiable techniques that are much simpler and scalable than verifiable ones.

In a previous work [5] (see also the extended version [6]) we have shown that any “Elementary SMC” (E-SMC) scheme that supports only simple additions with *private inputs and public output* is sufficient to support a set of primitive operations that are likely relevant for inter-ISP collaboration, e.g., Conditional Counting, Voting, Histogramming, Set Union, Anonymous Publishing and even Anonymous Scheduling. The point made in [5,6] is that private addition can become very powerful when combined with local transformations of the inner data, e.g., involving probabilistic data structures like Bloom filters and bitmaps. Whenever intermediate results – which are necessarily public in E-SMC – are not regarded as sensitive, such primitives can be chained into structured “private workflows” that safeguard the privacy of the input data as well as the anonymity of each player. We claim that a large part, if not all, of the procedures needed to support collaborative inter-domain network monitoring can be reduced to elementary secure additions.

Given this framework, the central design problem reduces to finding the most scalable way to implement elementary secure additions. In this paper we consider two possible schemes: the Shamir’s Secret Sharing (SSS), based on polynomial interpolation on prime fields, and the Globally-Constrained Randomization (GCR) scheme based on simple blinding [5]. The goal of this paper is to address the system-level aspects and quantify the achievable performance of both schemes. An attractive system-level feature of GCR is the possibility of pushing all the communication and processing overhead into a preliminary offline preparation phase, leaving the online computation phase as fast and lightweight as a cleartext addition. In order to compare quantitatively the performance of the two schemes in a fair way, we have implemented a prototype version of GCR in SEPIA [4], an open-source platform that supports SSS natively, and then performed a number of controlled experiments in emulated scenarios.

The contributions of this work are:

1. We discuss a number of system-design features of GCR that enable massive-scale implementation. That is, how to split the computation into *offline* randomization and *online* aggregation phases, and how to efficiently handle joining/leaving of players.
2. We assess the sensitivity of GCR performance to a number of system design parameters, as well as to the network conditions.
3. We compare quantitatively the performance of a GCR-based implementation of additive E-SMC versus a SSS-based implementation.
4. We investigate the resilience of the GCR scheme to node failures by leveraging theoretical analysis and emulation results.

The rest of this paper is organized as follows. Section 2 describes the reference scenario and the assumed adversary model. We review the GCR scheme and its features in Section 3. Section 4 contrasts GCR and SSS from a theoretical point of view. Sections 5 and 6, illustrate the implementation of GCR within SEPIA and the emulation setup, respectively. In Section 7 we assess the dependency of the GCR performance from system parameters and network conditions, and we contrast it with the performance attained by SSS. In Section 8 we investigate the impact of players fault on the GCR performance. Finally, related work is discussed in Section 9, and in Section 10 we summarize our conclusions.

2. Reference scenario

In the collaborative inter-ISP scenario, a set of ISPs holds a set of monitored data collected locally, like e.g., traffic statistics, network logs, records of security incidents. Based on these data, each ISP performs statistical and behavioral analysis of the hosts interacting with its network and to identify possible threats such as spam campaigns, worms spread-out, and Distributed Denial of Service (DDoS) attacks. Unfortunately, each ISP holds only partial information corresponding to its particular standpoint inside the global Internet. As pointed out already in [4], each ISP