Chair of
Communication and Distributed Systems

RWTH AACHEN UNIVERSITY

Jan Henrik Ziegeldorf

Designing Digital Services with Cryptographic Guarantees for Data Security and Privacy

Reports on Communications and Distributed Systems
Editor: Prof. Dr.-Ing. Klaus Wehrle

VOL 16

Designing Digital Services with Cryptographic Guarantees for Data Security and Privacy

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Informatiker

Jan Henrik Ziegeldorf

aus Schwerte, Deutschland

Berichter:

Prof. Dr.-Ing. Klaus Wehrle
Prof. Dr. rer. nat. Björn Scheuermann

Tag der mündlichen Prüfung: 08.12.2017

Reports on Communications and Distributed Systems

edited by
Prof. Dr.-Ing. Klaus Wehrle
Communication and Distributed Systems,
RWTH Aachen University

Volume 16

Jan Henrik Ziegeldorf

Designing Digital Services with Cryptographic Guarantees for Data Security and Privacy

Shaker Verlag
Aachen 2018

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: D 82 (Diss. RWTH Aachen University, 2017)

Printed in Germany.

ISBN 978-3-8440-5837-6
ISSN 2191-0863

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

In the past two decades, tremendously successful digital services have been built that collect, process, and monetize massive amounts of personal user data, up to the point where data is proclaimed the oil of the 21st century. Along come serious threats to data security and privacy that significantly increase the demand for effective protection, e.g., as manifested in the growth of encrypted Internet traffic. Communication security protocols, however, protect data against external attackers and do not address the root cause of almost all privacy threats, the need to share sensitive data with third parties. These third parties may illicitly process data beyond its original purpose of collection or be hacked and forced to provide data access. Countering these threats requires the development of Privacy Enhancing Technologies that complement or replace traditional communication security protocols.

We identify Secure Multiparty Computation (SMC) as a rigorous approach not only to provide data security and privacy protection, but even to reconcile privacy interests with seemingly adverse public and business interests. However, the potential of SMC is foremost on the theoretical level – it is often dismissed for being too inefficient and impedimentary for real-world applications. This thesis bridges the gap between the theoretical strength of SMC and the feeble realization of its potential in practice. To this end, we conduct a qualitative and quantitative analysis of SMC frameworks and abstract three research challenges: i) Extending the functionality and ii) increasing the efficiency of SMC as well as iii) customizing it to challenged environments. We choose a use case-driven research methodology to address these questions, which allows us to motivate and validate all our contributions in practice.

First, we motivate the problem of financial privacy in cryptocurrencies and propose decentralized mixing as a solution. We recognize the advantages of securing mixing operations with SMC and contribute secure protocols to technically realize our novel approach. As a result, our mixing system achieves stronger security and privacy guarantees than prior works while remaining highly scalable and fully compatible with the prevalent designs of decentralized cryptocurrencies such as Bitcoin.

Second, we propose efficient SMC designs for different classification algorithms to address data security and privacy issues in pattern recognition and machine learning. The evaluation of our classifiers shows that they are secure, accurate, and outperform the state of the art. We demonstrate three real-world use cases that prove applicability of our classifiers but also motivate their deployment in challenged environments. Thus, we present two additional approaches, bandwidth optimizations and secure outsourcing, to bring our secure classifiers to these scenarios.

Finally, we investigate secure outsourcing as a general strategy to customize SMC to challenged deployment and operation scenarios by the example of computing set intersections, a universal building block in many real-world applications and a well studied SMC problem. We present efficient schemes with negligible overheads for the outsourcers and demonstrate their applicability in two comprehensive case studies, privacy-preserving crowd-sensing and genetic disease testing in the cloud.

In summary, the contributions made in this thesis widen the technical solution space for practical data security and privacy protection in data-driven digital services.

Kurzfassung

In den letzten Jahrzehnten wurden erfolgreiche digitale Services entwickelt, die gewaltige Mengen persönlicher Nutzerdaten sammeln, verarbeiten und monetisieren. Damit einher gehen ernsthafte Bedrohungen der Datensicherheit und Privatsphäre sowie ein erhöhter Bedarf an Schutzmechanismen, wie sich beispielsweise im Zuwachs an verschlüsseltem Internetverkehr manifestiert. Etablierte Protokolle für Kommunikationssicherheit schützen jedoch nur gegen externe Bedrohungen und behandeln dabei nicht die Ursache fast aller Privatsphärebedrohungen, die Notwendigkeit sensitive Daten mit Dritten zu teilen. Solch dritte Parteien können z.B. die gesammelten Daten zweckentfremdet weiterverarbeiten oder zur Herausgabe der Daten gezwungen werden. Dies motiviert die Entwicklung Privatsphäre-erhaltender Technologien, die traditionelle Ansätze der Kommunikationssicherheit ersetzen oder komplementieren.

Wir identifizieren Secure Multiparty Computation (SMC) als rigorosen Ansatz, um Datensicherheit und Privatsphäre zu schützen und sogar bestehende Privatsphäreinteressen mit gegenläufigen Geschäftsinteressen zu versöhnen. Allerdings gilt SMC als zu ineffizient und hinderlich für praktische Anwendungen – die vorliegende Dissertation überbrückt die Kluft zwischen dem theoretischen Potenzial von SMC und dessen Realisierung in der Praxis. Unsere quantitative und qualitative Analyse existierender SMC Frameworks zeigt, dass dazu drei Herausforderungen überwunden werden müssen: i) Die Erweiterung der Funktionalität und ii) die Steigerung der Effizienz von SMC sowie iii) die Anpassung von SMC auf neue Einsatzszenarien. Um diese Probleme zu lösen, wählen wir einen anwendungsgetriebenen Forschungsansatz, der es uns erlaubt, unsere Beiträge praktisch zu motivieren und zu validieren.

Zuerst zeigen wir Privatsphäreprobleme im Bereich digitaler Währungen auf und untersuchen dezentralisierte Mix-Systeme als Lösung. Wir erkennen und motivieren die Vorteile, solche Systeme durch SMC abzusichern, und entwerfen anschließend die erforderlichen Protokolle. Unser Ansatz gewährt nicht nur stärkere Sicherheits- und Privatsphäregarantien als bisherige, sondern bleibt dabei skalierbar sowie vollständig kompatibel zu den Designs vorherrschender Kryptowährungen wie z.B. Bitcoin.

Im Folgenden entwerfen wir effiziente SMC Protokolle für verschiedene Klassifikationsalgorithmen, um bestehende Datensicherheits- und Privatsphärebedrohungen im Bereich des maschinellen Lernens zu lösen. Unsere Klassifikatoren sind sicher, numerisch genau und vor allem deutlich schneller als vorherige Ansätze, wie wir u.a. anhand von drei realen Anwendungsfällen zeigen. Unsere Bandbreitenoptimierung und Protokolle zur sicheren Auslagerung von Berechnungen erlauben es zusätzlich, mit den typischen Ressourcenbeschränkungen in diesen Szenarien umzugehen.

Schließlich untersuchen wir das sichere Auslagern von Berechnungen als generelle Lösungsstrategie, um SMC auf ressourcenbeschränkte Umgebungen anzupassen. Als konkretes Problem betrachten wir die sichere Berechnung von Schnittmengen, eines der meist untersuchten SMC Probleme und universaler Baustein für viele reale Anwendungen. Wir präsentieren effiziente Designs und demonstrieren ihre Anwendung anhand zwei umfassender Fallstudien im Bereich Crowd-Sensing und Gentests.

Insgesamt erweitern unsere Beiträge den technischen Lösungsraum praktikabler Daten- und Privatsphäreschutzmechanismen für datengetriebene digitale Services.

Acknowledgements

So many people have directly or indirectly influenced my dissertation in so many different ways that I cannot name them all. I am deeply grateful to all colleagues, friends, and family, who accompanied me along this journey. From its beginnings as a diploma student at COMSYS, over to Philips Research, Eindhoven, and back to COMSYS for good, as well as through all the little ”detours” that made work and life as a PhD student so much more interesting and worthwhile – I have enjoyed your company, wisdom, and support both professionally and privately and I sincerely hope you stick around for what is to come.

Contents