Guest Editorial
Privacy in Retrieval, Computing, and Learning

Sennur Ulukus, *Fellow, IEEE*, Salman Avestimehr, *Fellow, IEEE*, Michael Gastpar, *Fellow, IEEE*,
Syed Jafar, *Fellow, IEEE*, Ravi Tandon, *Senior Member, IEEE*,
and Chao Tian, *Senior Member, IEEE*

THE increasing prevalence of massive datasets makes the outsourcing of storage and computation tasks to distributed servers a necessity. This raises a number of concerns regarding the security and integrity of stored information, the privacy of accessing desired information, the communication overhead of distributed systems, the latency, reliability, and complexity of distributed computing, and privacy in distributed training and learning systems. Recent breakthroughs from coding, communication, and information-theoretic perspectives have opened up exciting new research avenues for these topics. There are many theoretical and practical open problems. This Special Issue is dedicated to communication theory, coding theory, information theory, signal processing, and networking aspects of privacy in information retrieval, privacy in coded computing over distributed servers, and privacy in distributed learning.

The Special Issue starts with a guest editor-authored tutorial overview article [A1], in which Ulukus *et al.* review privacy in retrieval, computing, and learning, describe some of the commonly used techniques, and survey the state-of-the-art. The tutorial paper is then followed by 20 technical papers.

In [A2], Liu *et al.* propose a privacy-preserving distributed algorithm to maximize cache hit rates of devices in the edge networks. This is a challenging problem since content popularities are often dynamic, complicated, and unobservable. To approach this problem, authors formulate the maximization of cache hit rates on devices as distributed problems under the constraints of privacy preservation and then introduce a privacy-preserving federated learning method for popularity prediction.

In [A3], Sasidharan and Thomas consider a distributed learning scenario, in which the edge nodes are available intermittently and are connected via low-bandwidth links. The edge nodes communicate local gradients to helper nodes, and these helpers forward messages to the central node after possible aggregation. In this setting, the authors propose a new scheme for gradient aggregation in distributed machine learning. Their scheme makes use of a well-known class of pyramid codes, thus expanding the realm of application of locally repairable codes to distributed learning. This article also establishes a trade-off between communication costs at edge nodes and at helper nodes.

In [A4], Li and Song study the federated multi-armed bandit problem under a master–worker, a decentralized and a hybrid structure. Several algorithms are proposed, and the performance is analyzed in terms of the regret. The three measures, i.e., data privacy, communication cost, and regret, enjoy different behaviors under these algorithms, which reflects a complex trade-off relation among them.

In [A5], Schlegel *et al.* consider the problem of linear inference on local data over a mobile edge computing network. A novel coding scheme based on Shamir's secret sharing algorithm is proposed to provide information-theoretic privacy against a given number of colluding edge servers while minimizing the overall latency in the presence of straggling servers.

In [A6], Naim *et al.* consider a setting that involves a single server and multiple users, where each user holds a discrete value and belongs to one of the k distinct groups. The goal is to allow the server to find the aggregated values in each group, under communication and privacy constraints. A novel algorithm is proposed to accomplish this task, which distinguishes itself from existing approaches by taking an interactive approach.

In [A7], Zhu *et al.* explore how multiple users may jointly, privately, and efficiently retrieve a file from a secure distributed database when the storage is coded and the set of servers includes unresponsive and/or Byzantine servers. A solution is proposed based on a combination of interference alignment and Lagrange encoding.

In [A8], Yakimenka *et al.* studied a relaxed notion of single-server PIR, where, instead of perfect privacy and perfect retrievability, some information leakage and distortion are allowed in order to reduce the download cost. The optimal trade-off between rate, distortion, and leakage is characterized

Sennur Ulukus is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: ulukus@umd.edu).

Salman Avestimehr is with the Electrical and Computer Engineering Department, University of Southern California, Los Angeles, CA 90007 USA (e-mail: avestimehr@ee.usc.edu).

Michael Gastpar is with the School of Computer and Communication Sciences, EPFL, 1015 Lausanne, Switzerland (e-mail: michael.gastpar@epfl.ch).

Syed Jafar is with the Department of Electrical Engineering and Computer Science, University of California at Irvine, Irvine, CA 92697 USA (e-mail: syed@ece.uci.edu).

Ravi Tandon is with the Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ 85721 USA (e-mail: tandonr@email.arizona.edu).

Chao Tian is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77842 USA (e-mail: chao.tian@tamu.edu).

Digital Object Identifier 10.1109/JSAC.2022.3142570

for large file sizes, and a construction based on linear programming is proposed for arbitrary file sizes.

In [A9], Obead *et al.* study a generalization of PIR, which allows a user to compute a linear combination of the messages that are stored in a coded fashion across distributed servers while keeping the combining coefficients private. The capacity is characterized for MDS coded storage.

In [A10], Budkuley *et al.* study the problem of commitment over a class of channels referred to as reverse elastic channels, which is a model of channel uncertainty. The authors establish a number of capacity theorems. They also present a conjecture on the commitment capacity of a symmetric channel instance.

In [A11], Li *et al.* develop a novel server cooperation strategy. In their model, the servers both cooperate and collude. The authors explain how their strategy can be put to use in the problems of secure distributed matrix multiplication (SDMM) and (linear) private information retrieval (PIR).

In [A12], Allaix *et al.* establish more general results for the problem of private information retrieval in quantum models. Achievable schemes are developed based on linear and MDS codes. The article also presents new converse bounds.

In [A13], Heidarzadeh *et al.* introduce the problem of private linear transformation (PLT) to generalize the problems of private information retrieval and private linear computation. Capacity results are established for two different special cases (referred to as JPLT-I and JPLT-II).

In [A14], Wan *et al.* present a novel framework for secure and distributed computation of linearly separable functions. Achievable schemes are proposed which can trade-off between local computational capabilities, common randomness, and the number of stragglers. In addition, converse results establishing the optimality of some of the schemes are also derived.

In [A15], Ye and El Rouayheb formulate and study the problem of intermittent private information retrieval, motivated by the need for privacy in location-based applications. The queries within this context involve correlated requests over time, and privacy requirements may only be necessary for some parts of requests. The authors develop a combination of obfuscation and PIR-based techniques for this problem while adapting them to the correlation structure between the requests.

In [A16], Shariatnasab *et al.* consider active deanonymization attacks within the context of bipartite networks and study the fundamental privacy limits. Specifically, attack algorithms are proposed by leveraging techniques from feedback communication with the goal of minimizing the number of queries needed for deanonymization. Theoretical analysis for stochastic models is presented together with simulation results.

In [A17], Hasircioglu *et al.* study the problem of secure distributed matrix multiplication and adapt bivariate polynomial codes for this scenario. These codes provide information-theoretic security guarantees and are shown to further speed up distributed matrix multiplication and reduce average computation time (compared to existing approaches in the literature), by exploiting partial work done by stragglers.

In [A18], Yan and Tuninetti address the problem of cache-aided robust, secure, demand-private scalar linear function retrieval in a multi-server setup. The authors use the key-superposition technique to simultaneously satisfy the constraints of content security against eavesdroppers, privacy of user demands against colluding users, and privacy of user demands against servers.

In [A19], Kurt *et al.* propose an algorithm which uses observed data sequence to detect network anomalies while maintaining data privacy and limiting the risk of false alarms. The emphasis of the article is on finding a model-free (data-driven) solution for anomaly detection since estimating the nominal model behavior is intractable due to large network size, and time-changing anomalous behavior.

In [A20], Song and Hayashi explore the connection between two important security primitives—symmetric private information retrieval and secret sharing, and establish a weak equivalence result between the two problems. The two primitives both have many applications, and the equivalence results found further connect them improving our understanding.

In [A21], Hong *et al.* consider a distributed computing framework for matrix multiplication where some of the workers are Byzantine, i.e., they send wrong computations to the master node. The article provides suitable solutions for identifying the Byzantine workers by proposing the use of locally testable codes together with a hierarchical group testing algorithm.

Appendix: Related Articles

[A1] S. Ulukus, S. Avestimehr, M. Gastpar, S. Jafar, R. Tandon, and C. Tian, "Private retrieval, computing, and learning: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142358.

[A2] S. Liu, C. Zheng, Y. Huang, and T. Q. S. Quek, "Distributed reinforcement learning for privacy-preserving dynamic edge caching," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142348.

[A3] B. Sasidharan and A. Thomas, "Coded gradient aggregation: A tradeoff between communication costs at edge nodes and at helper nodes," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: doi: 10.1109/JSAC.2022.3142356.

[A4] T. Li and L. Song, "Privacy-preserving communication-efficient federated multi-armed bandits," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142374.

[A5] R. Schlegel, S. Kumar, E. Rosnes, and A. G. I. Amat, "Privacy-preserving coded mobile edge computing for low-latency distributed inference," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142295.

[A6] C. Naim, R. G. L. D'Oliveira, and S. El Rouayheb, "Private multi-group aggregation," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142357.

[A7] J. Zhu, Q. Yan, and X. Tang, "Multi-user blind symmetric private information retrieval from coded servers," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142352.

[A8] Y. Yakimenka, H.-Y. Lin, E. Rosnes, and J. Kliewer, "Optimal rate-distortion-leakage tradeoff for single-server information retrieval," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142296.

[A9] S. A. Obead, H.-Y. Lin, E. Rosnes, and J. Kliewer, "Private linear computation for noncolluding coded databases," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142362.

[A10] A. Budkuley, P. Joshi, M. Mamindlapally, and A. K. Yadav, "On reverse elastic channels and the asymmetry of commitment capacity under channel with under channel elasticity," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142304.

[A11] J. Li, O. Makkonen, C. Hollanti, and O. Gnilke, "Efficient recovery of a shared secret via cooperation: Applications to SDMM and PIR," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142366.

[A12] M. Allaix, S. Song, L. Holzbaur, T. Pllaha, M. Hayashi, and C. Hollanti, "On the capacity of quantum private information retrieval from MDS-coded and colluding servers," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142363.

[A13] A. Heidarzadeh, N. Esmati, and A. Sprintson, "Single-server private linear transformation: The joint privacy case," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142293.

[A14] K. Wan, H. Sun, M. Ji, and G. Caire, "On secure distributed linearly separable computation," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142373.

[A15] F. Ye and S. El Rouayheb, "Intermittent private information retrieval with application to location privacy," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142301.

[A16] M. Shariatnasab, F. Shirani, and E. Erkip, "Fundamental privacy limits in bipartite networks under active attacks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142299.

[A17] B. Hasırcıoğlu, J. Gómez-Vilardebó, and D. Gündüz, "Bivariate polynomial codes for secure distributed matrix multiplication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142355.

[A18] Q. Yan and D. Tuninetti, "Robust, private and secure cache-aided scalar linear function retrieval from coded servers," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142360.

[A19] M. N. Kurt, Y. Yılmaz, X. Wang, and P. J. Mosterman, "Online privacy-preserving data-driven network anomaly detection," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142302.

[A20] S. Song and M. Hayashi, "Equivalence of non-perfect secret sharing and symmetric private information retrieval with general access structure," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142375.

[A21] S. Hong, H. Yang, and J. Lee, "Hierarchical group testing for Byzantine attack identification in distributed matrix multiplication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 3, Mar. 2022, doi: 10.1109/JSAC.2022.3142364.

Sennur Ulukus (Fellow, IEEE) received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University and the Ph.D. degree in electrical and computer engineering from WINLAB, Rutgers University. She is currently the Anthony Ephremides Professor in Information Sciences and Systems with the Department of Electrical and Computer Engineering, University of Maryland (UMD) at College Park, where she also holds a joint appointment with the Institute for Systems Research (ISR). Prior to joining UMD, she was a Senior Technical Staff Member at AT&T Labs-Research. Her research interests are in information theory, wireless communications, machine learning, signal processing, and networks, with a recent focus on private information retrieval, age of information, group testing, distributed coded computing, machine learning for wireless, energy harvesting communications, physical layer security, and wireless energy and information transfer. She is a Distinguished Scholar–Teacher with UMD. She received the 2003 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 IEEE Communications Society Best Tutorial Paper Award, the 2020 IEEE Communications Society Women in Communications Engineering (WICE) Outstanding Achievement Award, the 2020 IEEE Communications Society Technical Committee on Green Communications and Computing (TCGCC) Distinguished Technical Achievement Recognition Award, a 2005 NSF CAREER Award, the 2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Outstanding Teaching Award. She was a Distinguished Lecturer of the IEEE Information Theory Society from 2018 to 2019.

Salman Avestimehr (Fellow, IEEE) received the B.S. degree in electrical engineering from the Sharif University of Technology in 2003 and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California at Berkeley in 2005 and 2008, respectively.

He is currently the Dean's Professor, the Inaugural Director of the USC-Amazon Center on Secure and Trusted Machine Learning (Trusted AI), and the Director of the Information Theory and Machine Learning (vITAL) Research Laboratory, Electrical and Computer Engineering Department, University of Southern California. He is also an Amazon Scholar at Alexa AI. His research interests include information theory, large-scale distributed computing and machine learning, secure and private computing/learning, and federated learning. He has received a number of awards for his research, including the James L. Massey Research and Teaching Award from IEEE Information Theory Society, an Information Theory Society and Communication Society Joint Paper Award, a Presidential Early Career Award for Scientists and Engineers (PECASE) from the White House (President Obama), a Young Investigator Program (YIP) Award from the U.S. Air Force Office of Scientific Research, a National Science Foundation CAREER Award, the David J. Sakrison Memorial Prize, and several best paper awards at conferences. He has been an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY and the General Co-Chair of the 2020 International Symposium on Information Theory (ISIT).

Michael Gastpar (Fellow, IEEE) received the Dipl.El.-Ing. degree from the Eidgenössische Technische Hochschule (ETH), Zürich, Switzerland, in 1997, the M.S. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 1999, the Doctorat ès Science degree from the Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland, in 2002. He was also a student in engineering and philosophy at the Universities of Edinburgh and Lausanne.

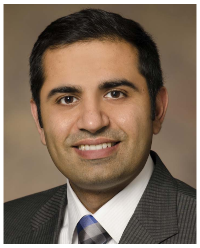
From 2003 to 2011, he was an Assistant and tenured Associate Professor with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley. Since 2011, he has been a Professor with the School of Computer and Communication Sciences, EPFL. He was also a Professor at the Delft University of Technology, The Netherlands, and a Researcher with the Mathematics of Communications Department, Bell Labs, Lucent Technologies, Murray Hill, NJ, USA. His research interests are in network information theory and related coding and signal processing techniques, with applications to sensor networks and neuroscience.

Dr. Gastpar received the IEEE Communications Society and Information Theory Society Joint Paper Award in 2013 and the EPFL Best Thesis Award in 2002. He was an Information Theory Society Distinguished Lecturer (2009–2011), an Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY (2008–2011), and he has served as a Technical Program Committee Co-Chair for the 2010 and 2021 International Symposia on Information Theory (Austin, TX, USA, and Melbourne, Australia).

Syed Jafar (Fellow, IEEE) received the B.Tech. degree from IIT Delhi, India, in 1997, the M.S. degree from Caltech, USA, in 1999, and the Ph.D. degree from Stanford, USA, in 2003, all in electrical engineering.

He is currently a Chancellor's Professor of electrical engineering and computer science with the University of California at Irvine, Irvine, CA USA. His industry experience includes positions at Lucent Bell Labs and Qualcomm. His research interests include multiuser information theory, wireless communications, and network coding.

Dr. Jafar was a recipient of the New York Academy of Sciences Blavatnik National Laureate in Physical Sciences and Engineering, the NSF CAREER Award, the ONR Young Investigator Award, the UCI Academic Senate Distinguished Mid-Career Faculty Award for Research, the School of Engineering Mid-Career Excellence in Research Award, and the School of Engineering Maseeh Outstanding Research Award. His coauthored papers have received the IEEE Information Theory Society Paper Award, the IEEE Communication Society and Information Theory Society Joint Paper Award, the IEEE Communications Society Best Tutorial Paper Award, the IEEE Communications Society Heinrich Hertz Award, the IEEE Signal Processing Society Young Author Best Paper Award, and various conference best paper awards. He received the UC Irvine EECS Professor of the Year award six times from the Engineering Students Council, a School of Engineering Teaching Excellence Award, and a Senior Career Innovation in Teaching Award. He was a University of Canterbury Erskine Fellow in 2010, an IEEE Communications Society Distinguished Lecturer for 2013–2014, and an IEEE Information Theory Society Distinguished Lecturer for 2019–2020. He is a Thomson Reuters/Clarivate Analytics Highly Cited Researcher. He served as the Technical Program Committee Co-Chair for the 2018 ISIT, Vail, CO, USA. He served as an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS from 2004 to 2009, for IEEE COMMUNICATIONS LETTERS from 2008 to 2009, and for IEEE TRANSACTIONS ON INFORMATION THEORY from 2009 to 2012.



Ravi Tandon (Senior Member, IEEE) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur (IIT Kanpur), in 2004, and the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park (UMCP), in 2010. He is currently the Litton Industries John M. Leonis Distinguished Associate Professor with the Department of ECE, The University of Arizona. Prior to joining The University of Arizona in Fall 2015, he was a Research Assistant Professor at Virginia Tech with positions in the Bradley Department of ECE, Hume Center for National Security and Technology, and at the Discovery Analytics Center with the Department of Computer Science. From 2010 to 2012, he was a Post-Doctoral Research Associate at Princeton University. His current research interests include information theory and its applications to wireless networks, communications, security and privacy, machine learning, and data mining. He was a recipient of the 2018 Keysight Early Career Professor Award, the NSF CAREER Award in 2017, and the Best Paper Award at IEEE GLOBECOM 2011. He currently serves as an Editor for IEEE TRANSACTIONS ON INFORMATION THEORY, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE TRANSACTIONS ON COMMUNICATIONS.

Chao Tian (Senior Member, IEEE) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2000, and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2003 and 2005, respectively.

He was a Post-Doctoral Researcher at the Ecole Polytechnique Federale de Lausanne (EPFL) from 2005 to 2007, a Member of Technical Staff—Research at AT&T Labs—Research, NJ, USA, from 2007 to 2014, and an Associate Professor with the Department of Electrical Engineering and Computer Science, The University of Tennessee Knoxville, from 2014 to 2017. He joined the Department of Electrical and Computer Engineering, Texas A&M University, in 2017. His research interests include data storage systems, multi-user information theory, joint source-channel coding, signal processing, and computing algorithms. He received the Liu Memorial Award at Cornell University in 2004 and the AT&T Key Contributor Award in 2010, 2011, and 2013. His authored and coauthored papers received the 2014 IEEE ComSoc DSTC Data Storage Best Paper Award and the 2017 IEEE Jack Keil Wolf ISIT Student Paper Award. He was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2012 to 2014 and an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS from 2016 to 2021, and is currently an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY.