Research Statement

Privacy is a fundamental human right [3]. The internet is a massive surveillance apparatus. My research attempts to reconcile the need for (provable) privacy with the reality of the entire world transforming to be internet-based with all the privacy risks that entails. In this transition, the digital economy prevalently depends on collecting (often sensitive) information about users, which digital service providers usually sell (or otherwise exploit) for uses that are certainly not in the consumers' best interests. When we search for (or consume) any information on online social media, e-commerce sites, video-streaming sites, etc. the service providers can track our search queries and interests and can infer personally identifiable information [12]. For instance, the digital content distributors often amass a fine-grain media-consumption data of a user to learn sensitive information like political orientation, sexual orientation, etc., which users may not like to share with anyone.

On the other hand, the market of the digital economy is worth three trillion dollars [7], and online service providers cannot suddenly stop accumulating said data. Moreover, most of the e-companies that support search engines show personalized ads to run their business model to earn a profit, and often they assemble this information to provide a customized user experience. My research agenda investigates the trade-off between the utility of the services and the privacy of the consumers. I have been investigating the practicality of employing cryptographic techniques to develop secure systems that protect user privacy as well as make online service providers interested in deploying privacy-preserving variants of their services. I want to build impactful and robust privacy-enhancing technologies (PETs) that are practical, scalable, and economically viable for online service providers to deploy them to safeguard the confidentiality of their consumers, providing the same set of utilities.

Below, I first describe my various research projects during my doctoral study; some of them have appeared in top-tier conferences of computer security (ACM CCS, PoPETS), followed by my short- and long-term goals for the future works.

Dissertation research

In my doctoral thesis, I study private information retrieval (PIR) as a vital step towards my long-term goal of building privacy-preserving e-business. PIR solves the ubiquitous problem of safeguarding the privacy of users' (reading) access patterns to remote, untrusted databases. PIR is an extremely powerful cryptographic primitive, yet existing PIR techniques mainly implement inconvenient database abstractions and impose expensive performance overheads; as a result, no PIR-based system has ever been deployed at scale in the wild. Modern database systems expose many different "views" of the underlying data to their users. Thus, to make PIR appealing to the users of such database systems, PIR systems should themselves support a variety of expressive and realistic query types. Likewise, the underlying cryptographic protocols must be sufficiently practical to let the benefits of supporting private queries overshadow the associated costs.

Expressive private queries (ACM CCS 2017): Most of the PIR works consider an elementary setting that allows a user to retrieve a single bit or a fixed-size block from the database by the index of the item, which we call positional query. But, in reality, modern applications interact with the underlying database by a variety of abstractions; many of them are by key-value pair and SQL-based queries. For instance, a user wants to know the most popular Bangladeshi restaurant near her location from Google Maps is unlikely to know (or care to know) the precise table and row within Google's POI database. Instead, the user is more interested in an abstraction that can facilitate the search "5 most popular Bangladeshi restaurants near me" on Google Maps without having to concern herself with the underlying database layout. We introduced a new *indexes of queries* technique [15] to support expressive query, which offers some advantages compared to the prior works [17].

Querying through any variants of the *indexes of queries* constructions is an instance of sparse matrix-vector multiplication, an extremely parallelizable workload much fit to run on GPU devices. Our results of the microbenchmark experiments showed that in the worst case, the new constructions incur overhead that is insignificant compared to the whole PIR process. Still, most of the cases, our

techniques can reduce the upload cost and server-side computation cost compared to the positional query-based PIR. We further evaluated the feasibility of deploying our technologies over a real-world dataset *IACR Cryptology ePrint archive*.

Optimal-rate private queries (PoPETS 2019): We demonstrated a novel family of so-called **1**-private, many-server PIR protocols [16] exhibiting unprecedented performance concerning *every* cost metric—download, upload, computation, and round complexity—typically considered in the PIR literature. With two servers, our protocols match the performance of the fastest known information-theoretic PIR [6] and computational PIR [5] protocols, with more servers, our constructions are much faster relative to them. We have implemented an open-source library libbitmore [11] to examine the efficiency of our constructions, setting the parameters at scale and evaluating the performance of the head-to-head comparison among the optimal protocols as well as Percy++ v1.0 [10] and RAID-PIR v0.9.5 [8] multi-server PIR libraries considered as benchmarks in the PIR literature. We found that the slowest member of our family is faster than the fastest mode of Percy++ v1.0 [10] and RAID-PIR v0.9.5 [8] and folklore protocols of Chor et al. [6] and Boyle et al. [5].

Tunicate onion descriptors of Tor (in preparation): This is an ongoing project to solve a real-world privacy problem on Tor; i.e., a low-latency anonymous communication system that provides anonymity for both user and service provider. Tor offers a protocol to hide the IP address of a web service from its clients and calls it hidden (now onion) services. We present *Tunica* protocols that employ PIR to prevent frustrating attempts to generate popularity histograms over *hidden service descriptors* and thereby develop an intersection-style attack [4] to deanonymize Tor clients. So that the *hidden service directories (HSDirs)* cannot distinguish among requests from Tor clients for different *hidden service descriptors*. We propose two Tunica constructions who leverage PIR to mitigate the said information leakage: the first variant, $\text{Tunica}_{\text{DPF}}$, is based on 2-server PIR from distributed point functions [9]. In contrast, the second variant, $\text{Tunica}_{\text{LWE}}$, is based on single-server XPIR [2] from the standard hard problem of *learning-with-errors over rings on ideal lattices*.

Future research plan

I plan to initiate collaboration with other researches from disparate disciplines (of computer science, psychology, law and policymaking, business, and human-computer interaction) with the ultimate goal of transitioning the theoretical contributions I made (and would make) into practice. Also, I like to team with industries to solve real problems they encounter in the direction of deploying a privacy-preserving version of their services. Such that the online resource providers can protect users from ill-tracking, fraudsters, identity thieves, intrusive advertising companies, repressive law enforcement and governments, and other potential cyber attackers. Next, I discuss some potential projects I desire to research short- and long-term basis along with my academic career.

Short-term research goals

I spotlight a list of upcoming projects that I aim for pursuing over the next couple of years. I am confident that the novelty, scalability, practicality, impact, and multidisciplinary nature of my visionary projects highlighted below prove myself as a passionate academician to excel in my research goals.

PIR in deployment: I want to work towards the elimination of the barriers of deploying PIR in the wild in multiple directions. Firstly, I like to endow PIR with more novel features comparable to the advanced "views" modern database systems offer users to access the underlying data. For example, I will extend the *indexes of queries* technique to support new (aggregate) queries like COUNT and SUM. With this feature, a user can ask *"how many papers published in a user-defined keyword in any particular year"* to the *IACR Cryptology ePrint Archive* dataset. Secondly, I want to achieve a remarkable milestone towards the ultimate practicality by turning our optimal-rate **1**-private PIR protocol [16] into a threshold scheme of $\boldsymbol{t}$-private constructions. The attainment in this research direction will prepare

PIR more painstaking to solve real privacy problems in emerging technologies, e.g., cryptocurrencies, IoT, autonomous vehicles, smart cities, etc.

Censorship circumvention: The anonymous communication system Tor can be used as a tool to circumvent censorship. A vital challenge to employ this tool is to distribute the so-called Tor bridges to censored users. There are several bridge distribution mechanisms [13], but all of them are implemented and evaluated in a simulated scenario. In collaboration with Dr. Sadia Afroz (ICSI) and Prof. Damon McCoy (NYU), we plan to assess the state-of-the-art proxy (Tor bridge) distribution mechanisms with each other and inspect the performances in the live Tor.

Privacy-preserving machine learning (PPML): Recently, I attended a 3-days-long Bootcamp on private artificial intelligence (AI) hosted by Microsoft Research. Where I, with three other Ph.D. candidates, Laia Amoros (Aalto University, Finland), Keewoo Lee (Seoul National University, Republic of Korea), and Caner Tol (Worcester Polytechnic Institute, USA), proposed a novel solution to ensure privacy during the trade of machine learning (ML) models between multiple parties leveraging homomorphic encryption (HE). Machine-Learning-as-a-Service (MLaaS) is a lucrative business, e.g., it will be worth 20 billion dollars by 2024, where big tech companies provide AI services to customers. One of the services allows customers to purchase an ML model on demand. Before the final transaction of purchase, without revealing the parameters of the ML model, the seller of the model wants to ensure that once the model is given to the buyer, she will buy it. Similarly, the buyer wants to ensure that the model is valid and accurate before purchasing it. This project inspires us to conduct more research on watermarking complex ML models beyond current developments [1] and convert them to HE-friendly versions like [14]. Moreover, this opportunity intrigues me to employ cryptographic techniques to solve other explicit and implicit leakages in current ML-based (and data-intensive) e-services.

Transitioning privacy-preserving e-services into practice: To redefine the online service providing systems, all stakeholders of the services need to be aware of preserving privacy. Users should have the necessary awareness to demand the service providers support customizable privacy. I will try to collaborate with researchers from human-computer interaction and psychology disciplines to inspect unsolved privacy problems and vulnerabilities of popular online services from the perspective of end users. Until recently, the most successful PETs deployed in the wild, e.g., the anonymous communication system Tor and private messaging-calling app Signal, are non-profit and have no business model. So, there is a potential scope to investigate the feasibility study of protecting user rights without compromising the functionalities of the services. I wish to cooperate with researchers from business disciplines to explore how online service providers can still earn a comparable profit through a business model that supports the privacy of their consumers. Ultimately, I plan to work closely with industry partners to better understand their constraints and privacy problems. Afterward, I will devise next-generation of PETs that are tailor-made to solve their business problems and practically deployable.

Long-term research goals

The dynamics of the internet change rapidly. There would be new security and privacy problems in current and upcoming online services. Day by day, the internet is getting flourished by the emergence of lightweight computing devices, e.g., the internet-of-things (IoT), self-driving vehicles, and smart cities; the privacy of end users of these products will be at more risk. In a longer period, I will continue investigating existent and new cryptographic techniques to solve the challenging security and privacy problems in the foreseeable future. To conclude, I envision my longer-term research projects will lead the two following movements.

Provably private e-business: I dream of an internet where each service provider ensures provable privacy to respect the fundamental rights of its consumers. To acquire the web of privacy-preserving e-businesses, I will work in three directions. At first, I desire to build a feasible privacy-preserving variant of a given e-service by knowing the constraints in the business model. Then I like to team with the e-business authority to redesign its services so that we can confirm similar user experiences

as well as profits while protecting the privacy of the users. Finally, I plan to work for the end users to make them aware of the necessity of protecting their privacy.

Internet freedom: Furthermore, I plan to work as an academic activist to route cryptographic ingenuity to free internet resources to everyone. I desire to concentrate more on internet censorship circumvention as well as research to support anonymity and transparency on the internet.

References

[1] Yossi Adi, Carsten Baum, Moustapha Cisse, Benny Pinkas, and Joseph Keshet. Turning your weakness into a strength: Watermarking deep neural networks by backdooring. In *Proceedings of the 27th USENIX Conference on Security Symposium*, SEC'18, pages 1615–1631, USA, 2018. USENIX Association.

[2] Carlos Aguilar-Melchor, Joris Barrier, Laurent Fousse, and Marc-Olivier Killijian. XPIR: Private information retrieval for everyone. *Proceedings on Privacy Enhancing Technologies (PoPETS)*, 2016(2):155–174 (April, 2016).

[3] UN General Assembly. Universal declaration of human rights (217 [iii] a), 1948.

[4] Alex Biryukov, Ivan Pustogarov, and Ralf-Philipp Weinmann. Trawling for Tor hidden services: Detection, measurement, deanonymization. In *Proceedings of IEEE S&P 2013*, pages 80–94, Berkeley, CA, USA (May, 2013).

[5] Elette Boyle, Niv Gilboa, and Yuval Ishai. Function secret sharing. In *Advances in Cryptology: Proceedings of EUROCRYPT 2015 (Part II)*, volume 9057 of *LNCS*, pages 337–367, Sofia, Bulgaria (April, 2015).

[6] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *Proceedings of FOCS 1995*, pages 41–50, Milwaukee, WI, USA (October, 1995).

[7] Patrick Delices. The digital economy. *Journal of International Affairs*, 64(1):225–226, 2010.

[8] Daniel Demmler, Amirr Herzberg, and Thomas Schneider. RAID-PIR; version 0.9.5 [computer software]. Available from: `https://github.com/encryptogroup/RAID-PIR` (October, 2016).

[9] Niv Gilboa and Yuval Ishai. Distributed point functions and their applications. In *Advances in Cryptology: Proceedings of EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 640–658, Copenhagen, Denmark (May, 2014).

[10] Ian Goldberg, Casey Devet, Wouter Lueks, Ann Yang, Paul Hendry, and Ryan Henry. Percy++ / PIR in C++; version 1.0 [computer software]. Available from: `git://git-crysp.uwaterloo.ca/percy` (October, 2014).

[11] Ryan Henry and Syed Mahbub Hafiz. libbitmore; version v0.0.1 [computer software]. Available from: `https://www.github.com/rh3nry/libbitmore` (July, 2019).

[12] Arvind Narayanan and Vitaly Shmatikov. Myths and fallacies of "personally identifiable information". *Commun. ACM*, 53(6):24–26 (June, 2010).

[13] Milad Nasr, Sadegh Farhang, Amir Houmansadr, and Jens Grossklags. Enemy at the gateways: Censorship-resilient proxy distribution using game theory. In *Proceedings of the Network and Distributed System Security Symposium*, NDSS'19, 2019.

[14] M. Sadegh Riazi, Mohammad Samragh, Hao Chen, Kim Laine, Kristin E. Lauter, and Farinaz Koushanfar. XONN: xnor-based oblivious deep neural network inference. In *28th USENIX Security Symposium, USENIX Security 2019, Santa Clara, CA, USA, August 14-16, 2019*, pages 1501–1518, 2019.

[15] **Syed Mahbub Hafiz** and Ryan Henry. Querying for queries: Indexes of queries for efficient and expressive it-pir. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 1361–1373, New York, NY, USA, 2017. ACM.

[16] **Syed Mahbub Hafiz** and Ryan Henry. A bit more than a bit is more than a bit better: Faster (essentially) optimal-rate many-server pir. *Proceedings on Privacy Enhancing Technologies*, 2019(4):112 – 131, 2019.

[17] Frank Wang, Catherine Yun, Shafi Goldwasser, Vinod Vaikuntanathan, and Matei Zaharia. Splinter: Practical private queries on public data. In *Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation*, NSDI'17, pages 299–313, Berkeley, CA, USA, 2017. USENIX Association.