Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on

Advancing Privacy Enhancing Technologies

Kotevska, Olivera

RFI Response: Advancing Privacy-Enhancing Technologies (35250)

Attention:

Jeri Hessman

pets-rfi@nitrd.gov

Submitted by:

Olivera Kotevska
Research Scientist
Oak Ridge National Laboratory
1 Bethel Valley Rd
Oak Ridge, TN 37831-6013
PO Box 2008, MS6013

Introduction

This white paper summarizes my feedback regarding the "Advancing Privacy-Enhancing Technologies" in response to the subject request for information.

Dr. Olivera Kotevska received her Ph.D. from the University of Grenoble Alpes in France and was an international guest researcher at the National Institute of Standards and Technologies (NIST) Gaithersburg, MD, during her doctorate. She started working at Oak Ridge National Laboratory (ORNL) in 2019 and is currently a research scientist. Her research is in the area of privacy algorithms for complex systems. She received internal ORNL funding for her research of privacy algorithms for streaming data and edge computing and additionally works on the projects for privacy protection of machine learning models. She has a few publications on these topics in reputable IEEE conferences [1-3], DOE white paper [4-5], and DOE open-source license [6].

Specific Recommendations

The current strategic vision does an excellent job of identifying and describing issues that must be addressed by the U.S. in the development of privacy enabled technologies. The descriptions and issues identified remain accurate and timely. As the vision development progress, it would be valuable to identify specific recommendations, policies, and

investments that will be important in sustaining U.S. leadership in privacy preservation. The following paragraphs summarize these recommendations in topical areas.

- **U.S. Leadership.** In the last few years, privacy has been identified as an integral part of today's and future technologies. As the technology progresses toward interconnectivity [7], voice-activated systems, automated and autonomous intelligent decisions, data becomes crucial for their operation. This explosion of data is increasing rapidly and exponentially, collecting sensitive information from healthcare to critical infrastructure. In this direction, the EU developed the first privacy regulation called GDPR [8] to protect users from unauthorized data collection, sharing and publishing. They even had a few court cases regarding data privacy with top tech companies to enforce its use. However, the U.S. has the CCPA [9], HIPAA [10], ECPA [11], and COPPA [12] created by congress and signed into law, but these are for a specific state, a specific type of application, or target group. Many international companies follow the GDPR standard, and it is not clear that the U.S. has the leadership in this position internationally. Privacy must be part of any data sharing, collection, and transfer to protect sensitive information from potential intruders.

  The plan should provide information regarding U.S. leadership in privacy preservation and regulations applicable on national and international levels. The development of privacy regulations that protect users and entities (e.g., power grid companies) should be developed and act as an example for other countries (e.g., EU and China) so U.S. can establish strongly their leadership position.

- **National Security Issues.** There have been cases where major U.S. entities had cybersecurity and data breach attacks. Some of them are U.S. Electric Grid'17, U.S. Federal Aviation Administration'15, U.S. Healthcare Network, SolarWinds supply chain data breach'20, First American Financial Corporation data breach'19, and many more. Because of these cases, the importance of privacy preservation technologies was addressed by the recent report published by the National Academy of Engineering [2]. For example, the electric grid is quite vulnerable to cyber-security attacks, and privacy preservation algorithms can provide an additional layer of data protection. Future applications (e.g., self-driving cars, transportation, energy automation, remote surgery, and patient monitoring) transfer raw unprotected data over the Internet. Privacy technology can protect sensitive data as close to the source as possible before it gets transferred to the cloud.

  The plan should recommend national and international norms for privacy preservation in critical infrastructure, economic situations, and healthcare. The privacy preservation technologies should be incorporated into the existing solutions and act as an option for the desired level of privacy protection.

- **Regulations.** As mentioned in the U.S. leadership section, some state and federal regulations exist in the U.S. The private sector in the U.S. lacks transparency in its technical specification if they follow any regulations and their response to the attacks on the Internet of Things (IoT) devices [13-14]. While in the E.U., regulations are mandatory, and this is clear, so the tech companies that operate in the E.U. need to follow them. This information creates trustworthiness by delivering this information to the users.

  The plan should include the creation of a law to address the mandatory privacy protection regulations across domains. Regulation could be two-fold: government policies giving the general direction in handling privacy and data protection or self-regulation to ensure that industries adopt best practices in cybersecurity and data minimization.

  Research Investments. The U.S. government agencies (e.g., NSF, DARPA, NIH) invested in the scientific foundations of privacy preservation. Particularly in the area of fundamental development of privacy algorithms, databases, and healthcare. However, most of the research has been performed by universities and partly by the national laboratories and industry. There are a few ways to advance future privacy-enhancing technologies (PET).
  - Taking the algorithms in the research community and making them in practical application.
  - Development of software solution for mapping the privacy regulations (*mentioned in the Regulation section*) into rules and integrating them into the existing application platforms. These include:
    - Utilize analytics to automated policy configuration and management.
    - Explaining current access policies over data.
    - Explaining the impact of changes to access policy.
  - Future research directions in privacy
    - Privacy for obfuscating data sources in images and other data types.
    - Better explanations and visualizations of privacy preservation.
    - Development of better privacy measurement and utility so that exploratory analytics is truly possible.
    - Enable meaningful control over data and mechanism of deciding how much privacy the user can have (*mentioned in the Industry and Application part*).
    - Development of mechanisms for checking if current data sharing matches historical records across the same user and other users and learning an optimal data sharing policy.
    - Incorporating privacy into decision-making, machine, statistical algorithms, and fairness analysis.

    - Development of trustworthiness in privacy algorithms is needed as it will encourage the users to share sensitive information and use the new technology.
    - Development of lightweight solutions that are robust to more powerful adversaries.
    - Development of solutions resilient to privacy attacks, especially under distributed settings.
- Research opportunities in science
    - Development of privacy layer for supercomputing resources.
    - Development of various surrogate datasets for testing privacy advances the development of robust privacy algorithms.
    - Development of distributed machine learning approaches (e.g., federated learning) with privacy capabilities to encourage data sharing between different entities. National laboratories have a challenge in accessing diverse datasets, so a solution that can guarantee the protection of sensitive information will encourage data sharing with the industry.
- The next generation of privacy-protecting systems.
    - As we progress into new computational architectures, e.g., quantum Internet and quantum computing, a neuromorphic adaptation of the privacy algorithms is needed.
    - Incorporate diverse privacy models and incorporate a socio-cognitive approach to improve privacy.

- **Industry and Applications.** Industry plays a crucial role in developing privacy-enabled technologies and making these capabilities available for use. In a few cases, the industry has performed fundamental research, e.g., Apple, Google, and Samsung. However, it is unclear if and what type of privacy preservation is applied by reading the specifications. They are long, unclear, and written in a language most people cannot understand. Most people do not read them, and very rarely is there an option for the user to choose their privacy preservation preferences. The relationship between industry and government in the U.S. is not clearly defined.

  The plan should address this issue explicitly and explore how to coordinate federal research investments with industry investments. Policies could be three-fold, i.e., a legal code prepared by lawyers and interpreted by courts, human-readable code easily understandable by consumers, and machine-readable code embedded in the IoT software. Data protection should be built into the IoT solution from the outset throughout the development life cycle as part of the principle of 'privacy by design.' Create awareness of the level of exposure associated with connectivity through interaction with IoT and ensure consumers are afforded opportunities to access and control their data.

- **Education.** A recent study showed that people have a minimal understanding of privacy preservation and complicated manuals, and programs make them give up easily.

  The plan should create educational programs, materials, and delivery methods in high-school, universities, and companies for privacy awareness and introduce a basic understanding of existing regulations and how to choose products with privacy features.

The short paragraphs above are intended to present ideas for expanded or added content to the revised national Privacy-Enhancing Technologies plan. If the Council or Subcommittee would like additional information or has questions on any of the topics above, please contact Dr. Olivera Kotevska.

References

[1]. Alfonso, S., Danfeng (Daphne) Yao, Kotevska, O., Measurement of Local Differential Privacy Techniques for IoT-based Streaming Data. In 2021, *18th IEEE Conference on Privacy, Security, and Trust (PST) (pp. 1-20). IEEE. Acceptance rate: 24%.*

[2]. Kotevska, O., Alamudin, F., Stanley, C., Optimal Balance of Privacy and Utility with Differential Privacy Deep Learning Frameworks. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 425-430). IEEE. *Acceptance rate for regular paper: 16%.*

[3]. Peralta-Peterson, M., Kotevska, O., Effectiveness of Privacy Techniques in Smart Metering Systems. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 675-678). IEEE. *Acceptance rate for short paper: 24%.*

[4]. O. Kotevska, C. Stanley, R. Michael, B. Kay, A. Sarwate, R. Kannan, G. Tourassi. Challenges with Sensitive Data in Distributed Graph. *Published* to DOE ASCR Workshop on Cybersecurity and Privacy for Scientific Computing Ecosystems. Link: https://custom.cvent.com/DCBD4ADAAD004096B1E4AD96F3C8049E/files/event/f9a6d32df34c43e5980b54f79ad848bc/738372bbcdd246ac9a9c00a177c075ac.pdf

[5]. J.R. Michael, C. Stanley, R. Adamson, O. Kotevska, Addressing the Limitations to Distributed Learning Containing Sensitive Data. *Published* to DOE ASCR Workshop on Cybersecurity and Privacy for Scientific Computing Ecosystems. Link: https://custom.cvent.com/DCBD4ADAAD004096B1E4AD96F3C8049E/files/event/f9a6d32df34c43e5980b54f79ad848bc/2e42c4c5dbe74b578e47188036244a10.pdf

[6]. Kotevska, Olivera. *Privacy Algorithms for Streaming Data.* Computer Software. https://github.com/ORNLStreamingPrivacy/Algorithms.git. Web.

[7]. National Academy of Engineering 2019. *Privacy and Security in the 21st Century: Who Knows and Who Controls? Proceedings of a Forum*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25575.

[8]. General Data Protection Regularization (GDPR). Link: https://gdpr-info.eu. Last visited: 06.29.2022. The Fourth Industrial Revolution. Link: https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab. Last visited: 06.29.2022.

[9]. California Consumer Privacy Act (CCPA). Link: https://oag.ca.gov/privacy/ccpa. Last visited: 06.29.2022.

[10]. Health Insurance Portability and Accountability Act (HIPAA). Link: https://www.hhs.gov/hipaa/index.html. Last visited: 06.29.2022.
[11]. Electronic Communications Privacy Act (ECPA). Link: https://epic.org/ecpa/
[12]. Children's Online Privacy Protection Rule (COPPA). Link: https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa. Last visited: 06.29.2022.
[13]. IoT Devices Most Vulnerable to Internal Security Threats. Link: https://www.iotworldtoday.com/2022/02/22/iot-devices-most-vulnerable-to-internal-security-threats/. Last visited: 06.29.2022.
[14]. Top 6 Hackable Medical IoT Devices. Link: https://www.criticalinsight.com/resources/news/article/top-6-hackable-medical-iot-devices. Last visited: 06.29.2022.
[15]. Apple Differential Privacy Team. (2017, December). Learning with Privacy at Scale.
[16]. Erlingsson, Ú., Pihur, V., & Korolova, A. (2014, November). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (pp. 1054-1067).
[17]. Nguyên, T. T., Xiao, X., Yang, Y., Hui, S. C., Shin, H., & Shin, J. (2016). Collecting and analyzing data from smart device users with local differential privacy. *arXiv preprint arXiv:1606.05053*.