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

Google

Office of Science and Technology Policy
Request for Information on Advancing Privacy-Enhancing Technologies
Doc. No.: 2022-12432; 87 Fed. Reg. 35250
July 8, 2022

Google welcomes the opportunity to provide comments in response to the Office of Science and Technology Policy Request For Information on Advancing Privacy-Enhancing Technologies (PETs).[1] Our comments describe Google's approach to PETs, including current applications. We conclude with policy recommendations, including public support, funding for research, regulatory incentives, technical standards and practices, and expert guidance for consideration in development of a holistic and flexible national strategy that facilitates responsible use of PETs.[2]

Emerging Privacy Tools And Techniques

New and emerging tools and techniques offer effective ways to safeguard and enhance privacy and security while enabling society to unlock the immense benefits that can be obtained from responsible use of data and technology across contexts, including research and commercial applications. As part of our continuing investments to drive innovation and make a safer ecosystem for Internet users, Google has invested significant effort in developing PETs, implementing them across our products, and making most of our research and tools open source so that anyone, anywhere can benefit from our advancements and contribute to progress in the field of PETs.[3] We also make a number of packaged solutions available to enterprise customers through Google Cloud, enabling them to take advantage of advanced privacy technology without having to make significant independent investments in research and tooling.[4] We discuss some specific PETs Google is investing in below.

---

[1] Google uses the terms "privacy-enhancing technologies" (PETs) and "privacy-preserving technologies" (PPTs) interchangeably and may use either term in the research cited herein.
[2] https://research.google/research-areas/security-privacy-and-abuse-prevention/.
[3] In the commercial context, advancing PETs that both improve privacy and provide utility, will lead to more businesses and organizations embracing PETs. This in turn can improve privacy and data protection outcomes, and further enhance trust in the organizations that use PETs and in the data sharing economy at large. This is one of the driving principles behind Google's Privacy Sandbox Project, which seeks to collaborate with industry by applying PETs to evolve existing digital ads practices.
[4] For example, confidential computing extends encryption to situations where the data is in use, unlocking new possibilities for collaboration, while preserving confidentiality of underlying

Differential Privacy

Differential privacy (DP) is an anonymization technology that adds specifically crafted noise to data or computations and provides a mathematical framework to quantify and understand the privacy guarantees of a system or an algorithm. DP is flexible and can be applied at the point of data collection (e.g. during survey collection), release (e.g. sharing a dataset), and model training (e.g. learning from data) to prevent identification of individuals. Because of its flexibility and ability to guard against privacy attacks such as reconstruction and memorization, it is suitable for many applications.

DP also can be used to share and analyze data in a privacy-respecting way across silos to train more robust machine learning (ML) models, while retaining privacy guarantees. For example, medical researchers may want to develop a ML model that predicts the prevalence of a disease such as cancer. Combining medical x-ray data across multiple institutions could produce a more accurate model but expose private patient data. If a model is trained on the aggregate data with DP, it could be more accurate and also provide a mathematically strong assurance about the privacy of the individuals' data.

At Google, we use DP in different applications including Android keyboard text prediction with provable privacy guarantees, traffic optimization on Google Maps, and the open source release of mobility metrics due to COVID-19. To democratize access and enable collaboration and use that advances the state of the art, we make our DP research and technologies available through open-sourced projects such as the Differential Privacy Library (for data analysis) and TensorFlow Privacy (for ML).

It is important to note that DP is not suited for every application: use of DP can contribute to inefficiency and slower processing and impact utility and model accuracy. Consequently, DP is not appropriate for long-tail analysis. For instance, if a researcher wants to answer questions on a small population, the relative negative impact of the noise that DP introduces will likely be large. The impact of DP on fairness and equity is an area that merits further research. The

---

data. We make this advanced technology accessible to customers through Confidential VMs and Confidential GKE Nodes. Another tool, Data Loss Prevention (DLP) helps customers protect against privacy breaches by scanning customer data against over 150 known information types to automatically identify, classify, mask, tokenize, and transform sensitive elements. DLP can also help measure how well quasi-identifiers are preserving data privacy through properties such as k-anonymity and l-diversity. And, because privacy requirements are increasingly baked into compliance frameworks, Google Cloud also offers tools that make it easier for customers to easily and confidently apply requisite controls to their data sets. Assured Workloads provides customers an easy mechanism for applying FedRAMP High, FedRAMP Moderate, IL4, and CJIS compatible controls to their workloads. For more, please see our overview of products and services relevant to data controllers and our privacy resource center.

long-tail problem becomes more acute when training ML models. By erasing the long-tail, DP prevents models from learning the behavior of small groups. Language translation provides a helpful example. If a small group speaks a rare language, then a model might not be able to learn that language, and would not be able to provide recommendations in that language. Thus, recommendations for advancing DP must consider how its application can interact with other important objectives, while privacy research continues to try to close these gaps.

Federated Learning And Analytics

Federated Learning (FL) is a data minimization technology developed at Google that enables state-of-the-art ML without centralized data collection. Using FL, organizations can train ML models with information from real-world interactions with people and improve their functionality over time, without needing to collect and store the underlying user data. Instead, the system computes model adjustments on the raw data wherever it resides (e.g. on users' devices) and only makes the aggregate model available to downstream systems and engineers.

In its original form, FL enables ML models to learn and improve over time without personally identifiable information ever leaving a user's device. Google uses FL to power smart text selection features in the Android operating system, "Hey Google" detection by Google Assistant, and Smart Reply in Android Messages, enabling more powerful and personalized services, and features that are accessible in no- or low-connectivity settings, while limiting the data that leaves a user's device. New applications of FL extend these capabilities across enterprises. For example, FL can be used to enable collaboration among companies and institutions, by allowing them to run analyses on their combined data without requiring the raw data to be shared. It can also be used to power privacy-preserving ML and analytics in data centers, allowing ML models to be trained with data from multiple silos without combining the data into large, centralized datasets. Google makes FL accessible to researchers and developers through the open-sourced TensorFlow Federated framework.

Because data is not collected in a data center when applying FL, we cannot use traditional methods to assess the quality of the model on real-world data. Federated Analytics (FA) addresses this by using the same federated infrastructure to compute metrics on device, and average those metrics to get population-level summaries of how models perform. For example, FA is used to show how accurate next word prediction models are in Gboard. FA also can be used to compute aggregate answers to data science queries over decentralized datasets, such as the data distributions or event frequencies.

Federated technologies have significant data minimization benefits and can also be combined with techniques like DP and Secure Multiparty Computation (see below) to further enhance privacy. But, they also come with tradeoffs. In on-device deployments, federated technologies

are limited by the available storage and computing power, meaning they may not be appropriate for certain applications. Furthermore, not collecting data in a centralized location limits organizations' ability to run certain types of analyses that require access to the raw data such as debugging or some types of fairness audits. As for the limitations impacting DP, these considerations should be taken into account in recommendations and further research to advance PETs.

Fully Homomorphic Encryption And Secure Multiparty Computation

Google uses a number of cryptographic protocols[5] to help users stay safe across the Internet. For example, fully homomorphic encryption (FHE) enables computation on encrypted data without revealing the underlying data. Some recent advancements in the field of FHE make it easier to implement. However, its use still requires cryptographic expertise which poses a barrier to greater access and use. Lowering the barriers to FHE research and adoption was a driving reason behind our work last year to open source a solution that enables any developer to create a program that can process encrypted data without decrypting it.[6]

Secure Multiparty Computation (SMPC) is another advanced cryptographic technique that provides utility from private data while strongly protecting the privacy and security of individual-level data. Generally it can provide guarantees that no entity learns anything more than their specified output, with proof that learning more is equivalent to breaking a cryptographically difficult problem (e.g. factoring, encryption), which is computationally infeasible and thus extremely unlikely to be achieved.

Google has explored use in several applications: Private-Join-And-Compute to allow two organizations to compute statistics across their private datasets; Secure Aggregation to enable a server to learn a federated learning model update from a group of users, while keeping each individual user contribution confidential; Private Information Retrieval to equip user devices to retrieve database entries from a server, while protecting the retrieved entry from the server; and Private Set Membership to equip user devices to check membership of an identifier in a set held by a server, while protecting the queried identifier from the server and the database from the users. We have also deployed SMPC through Exposure Notifications Private Analytics to allow health authorities to collect metrics about COVID-19 exposure notifications without learning the specific contribution from any individual device.

Generally, applications involving sensitive data or processing (e.g. financial, health), and especially those that would involve combining data from multiple sources, would all greatly benefit from the use of SMPC. As noted above, these could vary in operation between organization-to-organization SMPC, aggregate metric collection by a single organization, or

[5] https://security.googleblog.com/2019/02/protect-your-accounts-from-data.html.
[6] https://developers.googleblog.com/2021/06/our-latest-updates-on-fully-homomorphic-encryption.html.

devices using MPC to privately leverage a server's expertise. SMPC is especially a good fit for simple functionalities like retrieval, counting, or aggregate statistics like averages, regressions or low-depth ML.

However, SMPC can add computational and communication costs; additional engineering and implementation burden; and, similar to FHE, requires substantial subject matter expertise to implement accurately. While this is a steadily advancing research and development area, adoption to date is limited to very specific use cases and largely in the confines of academia and some startup businesses and large tech companies. Because of these limitations, complex functionalities that are difficult to run even without SMPC (e.g. building an internet-scale knowledge graph) are likely to be an infeasible fit for SMPC in the near and medium-term because of the overheads introduced.

Opportunities And Recommendations

We welcome the development of a comprehensive and multifaceted national strategy to significantly expand the use of PETs to benefit individuals and society. We agree with the RFI's assessment[7] that despite PETs' significant benefits, they have yet to achieve widespread adoption. Below we describe opportunities for government action and provide recommendations to achieve the important goal of greater privacy of, and utility from, data. Specifically, we would encourage a national strategy to:

- Champion the use of PETs through public commentary, policies, and government applications;
- Support federal investments in open fundamental and application-specific research and development, through additional funding, technical research, additional prize challenges, and other mechanisms for driving the development of PETs;
- Adopt a flexible, risk-based approach, that accounts for factors such as the sensitivity of data, available protections, and costs;
- Incentivize the development and adoption of PETs through smart regulation;
- Ensure that recommendations align with recognized privacy standards, and support the development of standards and recommended practices for implementation of PETs; and
- Provide guidance to facilitate the responsible use of PETs.

Champion Use Of PETs

PETs should be a staple of the Administration's public commentary about privacy and security. Administration officials should use their public platforms to raise awareness of PETs, share

[7] https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies.

helpful explanations, and encourage their use, particularly outside of the tech sector.[8] A national strategy should also encourage the use of PETs in government contracts and acquisitions, where feasible, and incorporate the use of PETs in existing and future government data collection, processing, and sharing practices.

Support Open Fundamental And Application-Specific Research

A national strategy should support fundamental research to develop and refine PETs and technologies through research organizations like the National Science Foundation and Defense Advanced Research Projects Agency, educational institutions, and through prize challenges like the one recently created by the US and UK governments.[9] Government funding should address the dearth of expertise required for continued development and use of PETs and expand pathways to enable more people from unserved and underserved communities to build the necessary skills.

A national strategy should also support a strong open-source ecosystem, including by helping to contribute libraries and frameworks, that ease the development and deployment costs of PETs.

Some PETs are promising yet relatively nascent, such as synthetic data generation and zero shot learning, which could encourage the use of less sensitive data, where feasible. Support for research into these technologies can contribute to their advancement and help make their use more widespread.

Adopt A Flexible, Risk-Based Approach, That Accounts For Factors Such As The Sensitivity Of Data, Available Protections, And Costs

While the diversity of PETs deployed across Google to enable secure and private data use helps illustrate the value of a national strategy, it also underscores the imperative of flexibility. No single PET is a panacea. A number of factors influence which PETs are appropriate for a particular product or use case. These can include the sensitivity of data, whether data is individual or aggregated, how data is being used, the intended outcome, the impact on functionality, and the size of and resources available to the implementer, among others. PETs

[8] For example, if a national strategy were to encourage companies to use SMPC to "exchange" nothing more than the output (since the intermediates are provably protected), more companies would see the benefit of the technology, and invest in its development, (including by making open-source contributions).

[9] https://www.whitehouse.gov/ostp/news-updates/2022/06/13/u-s-and-uk-governments-collaborate-on-prize-challenges-to-accelerate-development-and-adoption-of-privacy-enhancing-technologies/

may also impose costs, whether in the form of increased expense to the implementing party[10] or reduced functionality to end users. Because research and development into PETs is ongoing, the cost-benefit analysis is evolving.

A national strategy to advance the development and adoption of PETs must be similarly flexible to account for the myriad of use cases and organizations seeking to implement them. Rather than prescribe particular PETs or applications of PETs, a strategy should be risk-based, and provide guidance on appropriate use cases and safeguards, taking contextual factors and national and international standards into account.

A strategy that focuses only upon the most advanced technologies or upon particular organizations or use cases would exclude most organizations and fail to provide a comprehensive vision for the embrace of PETs. On the other hand, overly prescriptive, technology-specific approaches can discourage organizations from taking advantage of the most advanced PETs if regulations fail to keep up with research and development.

Incentivize The Development And Adoption Of PETs Through Smart Regulation

Drafters of privacy and data protection law in the US and around the world have recognized the need to balance the protection of data and privacy with important societal goals that can be advanced through the use of data. This objective is reflected in the appropriate exceptions for de-identified or anonymized data across many different pieces of privacy and data protection legislation. We recommend that the national strategy reflect this approach and encourage the use of PETs by granting similar exceptions to those found in privacy and data protection regulations. This is important given the costs PETs can pose to organizations, particularly those less-resourced. If organizations do not have a reasonable degree of certainty that they will benefit from implementing PETs, such as through reductions in compliance costs for exempt data, then they may be discouraged from taking on the increased cost of implementing these technologies.

It is also important to ensure that laws and policies do not inadvertently prevent or disincentivize the use of PETs in products. For example, requirements for the auditing or disclosure of datasets effectively require that those datasets be collected and stored centrally, precluding the use of techniques like FL and data minimization. Similarly, requirements around data accuracy should include clear exceptions to allow the introduction of noise into data to utilize DP.

[10] Such costs can include technical hardware, additional compute power and time, hiring of expert personnel, and the time and expense of changing organizational collection and use of data.

Ensure That Recommendations Align With Recognized Privacy Standards, And Support The Development Of Standards And Recommended Practices For Implementation Of PETs

National and international consensus standards are essential to the health of the global technology ecosystem, promoting cross-border and cross-application interoperability and minimizing barriers to trade and innovation. They can also help to ensure that PETs are implemented responsibly and build a common understanding of the privacy-enhancing benefits and tradeoffs of these technologies. A national strategy should align recommendations with widely accepted privacy standards, including the National Institute of Standards and Technology Privacy Framework, and International Organization for Standardization and Institute of Electrical and Electronics Engineers standards,[11] and where needed, support further development of national and international standards and recommended technical practices.

Provide Guidance To Facilitate Responsible Use Of PETs

Additional guidance and recommended practices for using PETs and managing their potential tradeoffs (e.g. between privacy and utility) would help organizations use them effectively, and encourage broader adoption. These recommendations may be general or technique-specific. For example, differential privacy is governed by the privacy loss parameter, epsilon, which affects the accuracy of the data, but also the privacy properties of the result. A lower epsilon means more privacy, but increases data loss. A higher epsilon is good for data utility, but a significant degradation of privacy. Practitioners have struggled to find the right balance between utility and privacy in these situations, and some guidance on how to balance competing equities when using PETs, particularly across various applications and contexts, would be welcome. Additionally, more information and guidance on how best to use some underutilized PETs such as SMPC and other advanced cryptographic techniques would also very likely expand adoption.

In addition, guidance around joint use of PETs to protect against a broader variety of privacy risks (e.g. FL and DP used together provide both data minimization and anonymization benefits) would be particularly welcome. A national strategy should make every effort to avoid an either-or approach to PETs and encourage joint usage where appropriate.

Conclusion

PETs play a critical role in the digital ecosystem by offering effective ways to safeguard and enhance individual privacy and data protection while allowing society to unlock the immense

[11] ISO/IEC PWI 6089 Guidance on addressing privacy protections for artificial intelligence systems may be a particularly strong opportunity.

benefits that can be obtained from the increased use and study of data. The protection of individual privacy is critical, and must also be balanced with other important societal goals that can be advanced through the use of data; PETs like DP, FL, FHE, and SMPC can help to strike that balance. PETs remain a top priority for Google as we seek to offer more secure, private experiences to our users on Google products and services and across the Internet, while also seeking to unlock the value data can bring to improving our products and helping all Americans. We welcome the opportunity to support OSTP's continued work to prepare a national strategy, and facilitate the responsible development and deployment of PETs.