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

VMware

July 8, 2022

Dr. Alondra Nelson
Director, Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, DC 20504

Dear Dr. Nelson:

VMware appreciates the opportunity to submit these comments in response to the Office of Science and Technology Policy's Request for Information (RFI) on Privacy-Enhancing Technologies (PET). Below we provide a high-level overview of VMware's research and development approach and vision relating to the data economy as well as answers to several of the questions posed in the RFI.

VMware's Vision: Confidential Computing Technologies as PET Foundation for the Data Economy
The data economy The data economy is defined by Wikipedia as "a global digital ecosystem in which data is gathered, organized, and exchanged by a network of vendors for the purpose of deriving value from the accumulated information". The data economy is estimated to comprise 1% of US GDP[1]. Importantly, there is a growing understanding that shared data is more valuable than unshared data; for example, Gartner reports that "Data and analytics leaders who share data externally generate three times more measurable economic benefit than those who do not"[2]. Data sharing increases the importance of both data security (the ability to keep secrets) and data privacy (the ability to control who sees those secrets and under what conditions).

Inhibitors of a data economy Consumers and enterprises are understandably reluctant to share data without privacy assurances, arguably holding back the full potential of the "data economy". There have been numerous instances of unauthorized data sharing[3] that have reduced trust by data owners. We hypothesize that technical privacy guarantees will unleash a more extensive and effective data economy.

Desired future state We believe that the future state should be one in which data owners retain full privacy controls **implemented with trustworthy technology foundations, not merely operating on trust**. Data producers/owners should have technically grounded control over who sees their data, when it is available for use, how it is combined with other data, when it is deleted, etc. Enterprises that participate in the data economy, and who implement appropriate technological controls, will become

[1] https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/FTQ_1_Jan_2019.pdf
[2] https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business
[3] https://iapp.org/resources/article/u-s-state-data-breach-lists/

not just “trusted” but “trustworthy”. Ideally, a virtuous cycle of deployment will be induced: the first enterprises to deploy trustworthy technology may become preferred service providers, and those who do not will be viewed with a skeptical eye by an increasingly large number of customers/partners.

Beneficiaries Data producers and owners will benefit, both intangibly (e.g., from superior services resulting from the controlled sharing of data) and tangibly (e.g., from the possibility of directly monetizing the controlled sharing of data in a data economy). Enterprises that offer products and services based on shared, combined data will benefit economically through competitive advantage and more customized services. Companies that provide infrastructure for the data economy will benefit. A workforce that specializes in implementations of the data economy will benefit. However, there could also be dislocations. Corporations whose products and services rely on the availability of uncontrolled data might find that it is less available and/or that they would need to compensate data owners for the use of that data. In addition, there could be costs associated with implementing technical privacy controls; it is our hypothesis that the cost of such controls is less than the economic and societal benefit of a robust data economy business opportunity. Finally, the widespread deployment of confidential computing has security advantages beyond use in the data economy; thus national security will eventually be strengthened.

Technology

Confidential computing as a foundation for security and privacy. It is our hypothesis that “confidential computing” (CC) technologies are necessary foundations for enabling comprehensive data security and privacy. Confidential computing is a principled, hardware-based security mechanism for distributed computing. CC protects the integrity and confidentiality of processing and data, wherever programs run, from malicious programs or careless insiders. CC provides confidentiality for data in use and enables, for the first time, a general-purpose end-to-end encryption of data (at rest, in transit, during processing) that was not feasible prior to the recent introduction of “trusted execution environments” (TEEs) such as Intel’s Software Guard Extensions (SGX), AMD’s Secure Encrypted Virtualization, Intel’s Trust Domain Extensions (TDX), and Arm “Realms”. Through isolation and encryption, CC provides a general-purpose means of grounding data security in the data economy.

Measurement and attestation: under-appreciated aspect of confidential computing. Two under-appreciated properties of confidential computing are “measurement” and “attestation”, which determine both the programs (applications) and the hardware that can be "trusted". Through these mechanisms, CC grounds data privacy capabilities that could unleash a vibrant data economy. Without measurement and attestation, unauthorized programs could be arbitrarily used on private data by malicious insiders, careless operators, and/or predatory participants in a data economy. With measurement and attestation, it becomes possible to control who can see or modify the data and under what circumstances.

Hardcoded privacy and chains of trust. The security concept of “confidentiality” is a building block for privacy in the sense that privacy controls must be correctly implemented through encryption, certificate management, and proper operation of the software that directly manipulates the decrypted data. Many details need to be successfully addressed in order to implement privacy controls, i.e., to specify who is

given access to data and under what circumstances. In general, formal policy languages are not yet used to specify data privacy controls which means that manual techniques must be used to verify that a given system obeys an implicit privacy specification. Not only are these manual mechanisms error prone, but they are also impossible to verify automatically as use evolves. The full range of details that needs to be considered is enormous, even without the risks associated with coding errors and malicious exploits. Confidential computing does not inherently address these complexities, but it does offer a foothold for addressing these problems that cannot be easily subverted.

One important contemporary cloud security problem that has privacy implications involves the simple use case of protecting a cloud application (and its data) from a cloud provider hosting that application. This is a case where in principle a careless or malicious operator has the potential to introduce a privacy violation because both the cloud provider and the application developer are a "chain of trust". One general class of privacy violation that needs to be addressed is preventing unauthorized access, sharing and manipulation by a cloud provider; in other words, to remove that cloud provider from the chain of trust.

To leverage the foothold provided by confidential computing, VMware has researched and implemented a "certifier" framework that simplifies the range of issues that need to be addressed in confidential system design, such that trust policy is separated from implementation and (with few additional lines of code) the developer's role in ensuring security is reduced to providing a correctly written program and specifying an access policy that represents their intentions. It is VMware's intention to contribute this certifier framework to the open source community, to enable both "hard-coded" privacy such as trustworthy cloud computing as well "data economy" applications that we hope will eventually support sophisticated end-user "Data Use Controls".

Data use controls (DUCs). The measurement and attestation of programs opens the door to a privacy-preserving data economy infrastructure. A given program can be inspected to ensure that it does not programmatically divulge, reveal, or misuse data. Measurement and attestation of this program can then ensure that only acceptable programs operate on data in question. However, initially, data providers will not have machine-readable specifications concerning rights and limitations on use of their data. Therefore, that specification will necessarily be encoded in the logic of the attested program. A better option exists, in which each data provider can explicitly specify the uses and restrictions that apply to their data, and those directives are enforced automatically and monitored. The bridge to this next step involves policy languages that have been researched under the term "data use controls" (DUCs).[4] By bundling data with their DUCs, a data provider's policy concerning the permitted uses of their data could potentially be unconditionally enforced by the attested program. Such a program could be verified to correctly implement data use controls in a trustworthy and verifiable manner. Over time, there is potential for automated verification techniques over increasingly sophisticated DUCs policy languages, which are already being shown[5] as capable of supporting the full generality of data privacy requirements such as the Health Insurance Portability and Accountability Act (HIPAA) and the EU's

[4] https://www.usenix.org/conference/hotosxiii/do-you-know-where-your-data-are-secure-data-capsules-deployable-data-protection

[5] *cf* https://arxiv.org/abs/1909.00077

General Data Protection Regulation (GDPR). As described below, DUCs represent an important direction for use-inspired basic research and in some cases may be ready for research translation.

The role of cryptographic approaches to secure multiparty computation. Cryptographers are sometimes skeptical of hardware-based security mechanisms such as TEEs in part because they may contain opaque implementation details; in principle, the manufacturer is in the chain of trust, and can introduce security flaws deliberately or mistakenly. There is a significant body of research in cryptographic approaches to privacy that do not require trust in the hardware manufacturer (for example, secure multi-party computation, homomorphic and semi-homomorphic encryption, and zero-knowledge proofs). However, there are often performance and complexity challenges with these approaches. In addition, present research has not addressed the incorporation of DUCs into these approaches. Nevertheless, these cryptographic approaches can potentially enhance specific privacy challenges and complement CC as a form of defense in depth even if they do not become a general-purpose foundation for the data economy. More research is needed.

Blockchain platforms. Blockchain-based infrastructures are often proposed as a foundation for digital commerce and may play a role in the "data economy". Blockchain platforms involve a combination of technical components including consensus mechanisms (e.g., proof-of-work, proof-of-stake, proof-of-authority), immutable data structures (e.g., Merkle trees), and transaction languages that execute commercial exchange. These foundations have been used to create marketplaces for digital objects (e.g., non-fungible tokens or NFTs). Blockchains can also be useful as a foundation for other important aspects of a data economy, such as auditing of configuration changes, security compliance, and data access transparency. Although blockchains have been identified as one foundation for privacy-preserving operations such as financial "know your customer" inquiries, the full machinery of modern blockchains may not be needed to support the data economy. For example, it is possible for two parties to execute commercial agreements without involving immutable data structures, complex consensus mechanisms, or formal transaction languages. Therefore, the door should be left open to new discoveries about the role of blockchain machinery in the data economy, but blockchains do not need to be an assumed foundation.

Technology priorities and challenges. To accelerate the data economy, the first priority is to encourage the widespread deployment of TEEs that underpin confidential computing. Concurrently, incentives and programs should encourage the use of CC technologies for privacy protection in the context of enterprises that participate in the data economy. Third, research should be accelerated in the area of data use controls. Initially, enterprises will likely need to specify their customers' intent on behalf of the customers, but eventually the control of that specification should devolve to the originator/owner of the data. NIST has played an important leadership role in promulgating the idea and approach to "zero trust" security[6]; the agency could arguably play a similar leadership role in developing the architecture and practical approach to a data economy. Ultimately, it will be desirable to standardize DUCs policy languages to increase interoperability.

[6] https://csrc.nist.gov/News/2022/planning-for-a-zero-trust-architecture-white-paper

With the above context, we now turn to commentary on selected RFI Questions:

1. Specific research opportunities to advance PETs

A modified version of Donald Stokes' 4-quadrant research framework[7] can help characterize the initiatives needed to enable a data economy. To complement Stokes' original research framework, NSF's new Technology and Innovation Partnerships (TIP) directorate has proposed a new dimension for *research translation* which we represent as "TIP's dimension"; this dimension focuses on the work of innovation, i.e., translating inventions into practice. A comprehensive governmental initiative would invest in a range of research programs touching on each quadrant and dimension.

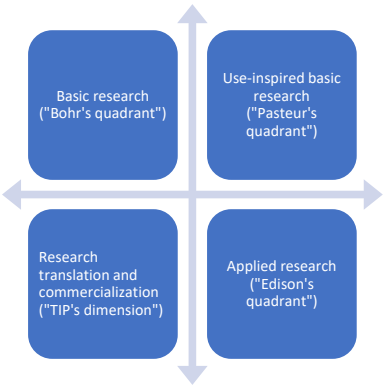

Figure 1: A modified version of Donald Stokes' framework

Confidential computing is sufficiently advanced that it can be immediately put into practice, while new research can subsequently enhance and deepen the latent and emerging opportunities. Therefore, VMware suggests pursuing each quadrant in parallel.

Research translation and commercialization: vendor-neutral, standardized infrastructure for CC and DUCs. It would be valuable to encourage development of common infrastructure for confidential computing and data use controls across producers and exchangers of data. Today, there are a plethora of TEEs under production, and, every cloud provider has a different approach to confidential computing (in particular to the level of support that is provided for key privacy-related features of confidential computing - measurement and attestation).

Thus, a different approach is needed to remove each specific cloud provider from the chain of trust, which creates an onerous burden for most cloud users. As mentioned above, VMware intends to bring certain vendor-neutral technologies to the open source community in order to help with this problem. Much work will remain, however, ranging from new data exchange interfaces for data lakes and warehouses, to common policy languages for expressing privacy desires.

[7] https://en.wikipedia.org/wiki/Pasteur%27s_quadrant

Patterned on its work in "zero trust" security[8], NIST could be of significant help in developing best practices and incremental approaches to vendor-neutral, standardized data exchange infrastructure that could be used by enterprise that wish to become privacy leaders on behalf of their employees and customers.

Research translation and commercialization: seeding the data economy with exemplar products and services Although many processor vendors have introduced hardware abstractions for confidential computing, industry is only beginning to exploit these capabilities for security and privacy. A key goal of research translation should be to proliferate creative new privacy-related use cases and implementations of confidential computing for purposes of privacy.

One mechanism to advance this goal would be to introduce new Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) initiatives that encourage the development of confidential computing products and services that leverage secure multi-party data sharing. In addition, the National Science Foundation (NSF)'s new Technology, Innovation and Partnerships (TIP) directorate could contemplate launching innovative public-private partnerships to stimulate the data economy and accelerate research translation in CC and DUCs, leveraging not only the Convergence Accelerator model but also looking at other models including joint solicitations.[9]

Use inspired basic research: data use controls and other advances As mentioned elsewhere, foundational PET research has already been funded and is being advanced by NSF as well as digital infrastructure providers such as VMware. Examples of this research include portable TEE abstractions, separation of attestation policy and implementation, certification services for end-to-end trust establishment in distributed confidential computing systems, and research into data use controls policy languages. Some of this work may lead to standards and implementation in open source under the guidance of forums such as the Confidential Computing Forum. However, there is additional use-inspired basic research that would be highly desirable in the long run. Important topics include DUCs frameworks; parameterizing automated CC "attestation" with data use controls; mapping prose data use policies into machine-readable DUCs specifications; identifying synergies between CC and other PETs; and extending CC to emerging hardware such as machine learning (ML) accelerators, remote memory, tagged architectures, etc.

Applied research There is a fair bit of applied research that is needed to deploy confidential computing, including, for example, performance and correctness work related to specific implementations. Although industry will likely carry out much of this applied research without government funding,

[8] https://csrc.nist.gov/News/2022/planning-for-a-zero-trust-architecture-white-paper

[9] VMware regularly engages in such public-private research partnerships with the NSF and has co-funded joint solicitations in Security (https://www.nsf.gov/pubs/2016/nsf16582/nsf16582.htm), Edge (https://www.nsf.gov/pubs/2018/nsf18540/nsf18540.htm), Sustainability (https://nsf.gov/pubs/2020/nsf20594/nsf20594.htm), and Next G Telecommunications https://www.nsf.gov/pubs/2021/nsf21581/nsf21581.htm).

government can support the applied research necessary in other ways, such as by serving as a convening body for industry, academia, and government to discuss applied research developments.

Basic research Finally, there is a need for additional basic research in the area of privacy. Of necessity, the term is a subjective, and is viewed and experienced differently by different populations. For example, Prof. Seny Kamara (Brown University) has called attention to the importance of considering marginalized communities in the context of cryptography.[10] The NSF Social, Behavioral, and Economics (SBE) directorate is positioned to engage in relevant social science and ethnography research, to identify societal impacts of privacy or its absence, and to understand economic and incentive structures leading to a vibrant data economy. In addition, NSF CISE directorate is well-positioned to engage in more basic research on novel PETs.

3. Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

Sectors: Although there is a segment of the economy that does (or could be seeded) to specifically support a "data economy" (e.g., data aggregators), the "data economy" could also be an aspect of many existing industries. Sectors that could participate in a CC-based data economy (and benefit from PETs more broadly) include:

- **Companies that aggregate and analyze data** as their core business
- **Web-enabled enterprises** that wish to remove themselves from the chain of trust in respecting privacy controls
- Enterprises in industries like **finance** and **healthcare** that require data aggregation but have regulatory requirements around privacy-preserving data sharing
- **Government agencies** (**defense, intelligence**, etc) that have multilevel security challenges

Workloads: Privacy-preserving workloads that could be implemented with today's confidential computing capabilities range from the scientific (e.g., privacy-preserving genomics analysis) to commerce (e.g., privacy-preserving auctions) to new businesses (e.g., consumer monetization of privacy data). Moreover, confidential computing offers a general-purpose and performant way to implement many of the workloads being discussed today in the context of cryptographic analytics (e.g., secure multi-party computation, homomorphic and semi-homomorphic encryption, private set intersection, zero knowledge proofs) and certain forms of federated machine learning. Programs aimed at exploring the use cases should help motivate others by showing feasibility and inspiring new commercial opportunities for trustworthy data analytics.

7. Risks related to PETs adoption

[10] https://www.youtube.com/watch?v=Ygq9ci0GFhA

We see a number of "incubation risk" unknowns involved in promoting the data economy. Many of these can potentially be mitigated, but as with all initiatives it will be important to monitor and adjust the nation's investments and plans accordingly as more is learned about these risks.

Risk: The concept of "privacy" is extremely broad, and there may be aspects that cannot be supported with technology.

- Mitigation: Nevertheless, the end-to-end encryption enabled by confidential computing, combined with technical data user controls over data dissemination, will be a useful advance for many classes of data. In addition, basic research over the next few years may reveal unexpected techniques and insights to close selected gaps.

Risk: IOT and "unpermissioned" data collection can circumvent technical user controls when third parties are collecting the data.

- Mitigation: Nevertheless: (1) even if user controls can be circumvented, that limitation doesn't mean other user data shouldn't be protected and user-controlled; (2) with the right ecosystem hacks "trustworthy" companies may offer those controls voluntarily; (3) this is an area where regulatory interventions may usefully complement technology; and finally (4) basic research may reveal useful new techniques and approaches (cf. the Stanford Secure Internet of Things center's ideas around auditing IoT devices[11]).

Risk: Derived data is particularly challenging to control from a user control perspective and from a differential privacy perspective.

- Mitigation: Attestation and data use controls could potentially be extended to expose data only to programs whose derivations are "acceptable". In addition, provenance information could be included in the DUCs to retain controls over the use of derived data.

Risk: The data economy sounds like an idea that presumes Blockchain, not just confidential computing – is there a risk-increasing dependency?

- Mitigation: Our view is that Blockchain could be one of several mechanisms for exchanging value, managing identity, etc. but need not be an assumed underpinning of any solutions. Not every data economy application needs strict consistency or immutable ledgers, for example. Even so, perhaps the machinery of digital commerce (NFTs) could be part of a data economy solution.

Risk: Unlike pure algorithmic cryptographic approaches (secure multiparty computation, etc.) confidential computing puts the hardware vendor in the chain of trust. Some may argue that the community should focus on algorithmic approaches to data exchange instead.

- Mitigation: Trust in the processor company is a form of fate sharing. If one can't trust the processor, all is lost anyway. Moreover, even if cryptographic approaches are used, the input data ought to be controlled with confidential computing and DUCs. Our view

[11] https://dl.acm.org/doi/10.1145/3081333.3081342

is that there are many techniques that can be added to confidential computing as a form of defense in depth, although more research is needed into the specifics.

Risk: The community has discovered vulnerabilities in certain TEEs and it will take time to deploy successor technologies. Moreover, new vulnerabilities are likely to emerge.

- Mitigation: Our view is that there is enough momentum around TEEs that the vendors are likely to improve over time with widespread adoption. Meanwhile even vulnerable TEEs are hard to exploit, so there is significant CC value-added even in the presence of exploits.

Risk: Industry lacks sufficient support for a common data economy architecture (common infrastructure APIs, interfaces, etc).

- Mitigation: Addressing this risk would be the point of a comprehensive public-private initiative. Our view is that we need to find an "ecosystem hack" to bootstrap this industry. We also believe it would be valuable for NIST to do for the data economy what it has done for the concept of "Zero Trust".

Risk: Today's cryptography may be broken by Quantum techniques and/or flawed implementations

- Mitigation: In addition to the use of post-Quantum techniques where feasible, we advocate crypto agility at the enterprise scale to allow for comprehensive system updates as flaws are found and to assist with mapping appropriate ciphers to data based on data classification – VMware is doing research in this space[12] and recently demonstrated quantum safe crypto agility at VMworld 2021[13].

9. Existing barriers, not covered above, to PETs adoption

As mentioned above, Prof. Seny Kamara's thoughtful keynote "Crypto for the People"[14] articulated the need for research (and approaches to research) that would better address the needs of marginalized communities. It would be desirable if data use controls and confidential computing could be leveraged by marginalized communities to provide new opportunities for user control, provenance, access transparency, and forensics related data misuse. However, research is needed to understand these opportunities.

10. *Other information that is relevant to the adoption of PETs*.

To guide the development of a data economy, it may be helpful to develop a maturity model for security and privacy in the data economy. A potential draft scaffolding is as follows. Many industry players are currently working to achieve "level 1", and there are selected instances of "level 2".

[12] https://research.vmware.com/projects/cryptographic-agility
[13] https://octo.vmware.com/its-time-for-crypto-agility/
[14] https://www.youtube.com/watch?v=Ygq9ci0GFhA

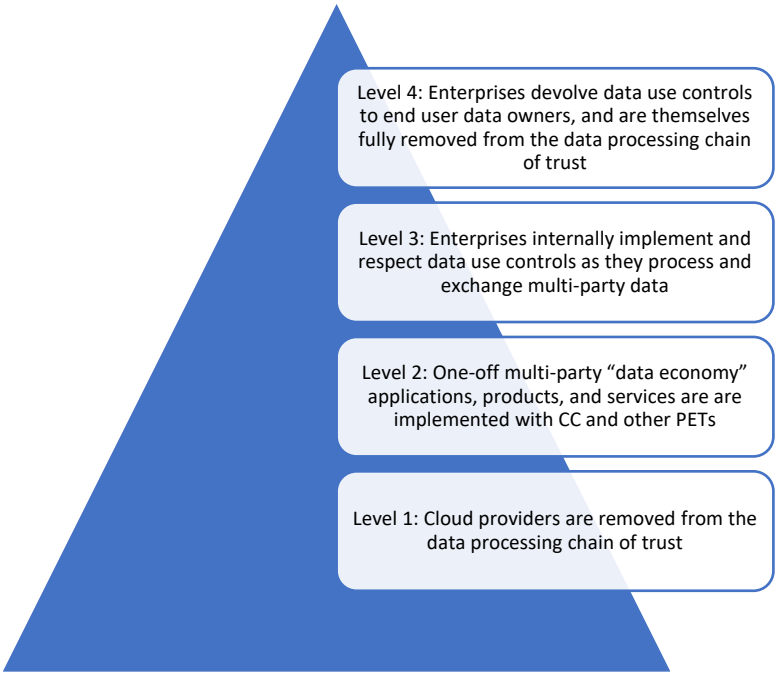

Figure 2 A maturity model of privacy controls in the data economy

Conclusion

Thank you again for this opportunity to provide comments on PETs as the Office of Science and Technology Policy embarks on the development of a national strategy on privacy-preserving data sharing and analytics. VMware looks forward to continuing the conversation and contributing to this important effort.

Sincerely,

Chris Ramming
Senior Director, Research and Innovation

VMware Inc.
3401 Hillview Avenue
Palo Alto, CA 94304