2022-11-01

Multi-regulation computing: examining the legal and policy questions that arise from secure multiparty computation

Multi-Regulation Computing: Examining the Legal and Policy Questions That Arise From Secure Multiparty Computation

Julissa Milligan Walsh*
Julissa.Papers@gmail.com
Boston University
Boston, MA, USA

Mayank Varia
varia@bu.edu
Boston University
Boston, MA, USA

Aloni Cohen
aloni@uchicago.edu
University of Chicago
Chicago, IL, USA

Andrew Sellars
sellars@bu.edu
Boston University
Boston, MA, USA

Azer Bestavros
best@bu.edu
Boston University
Boston, MA, USA

ABSTRACT

This work examines privacy laws and regulations that limit disclosure of personal data, and explores whether and how these restrictions apply when participants use cryptographically secure multiparty computation (MPC). By protecting data during use, MPC offers the promise of conducting data science in a way that (in some use cases) meets or even exceeds most people's conceptions of data privacy. With MPC, it is possible to correlate individual records across multiple datasets without revealing the underlying records, to conduct aggregate analysis across datasets which parties are otherwise unwilling to share for competitive reasons, and to analyze aggregate statistics across datasets which no individual party may lawfully hold.

However, most adoptions of MPC to date involve data that is *not* subject to privacy protection under the law. We posit that a major impediment to the adoption of MPC—on the data that society has deemed most worthy of protection—is the difficulty of mapping this new technology onto the design principles of data privacy laws. While a computer scientist might reasonably believe that transforming any data analysis into its privacy-protective variant using MPC is a clear win, we show in this work that the technological guarantees of MPC do not directly imply compliance with privacy laws. Specifically, a lawyer will likely want to ask several important questions about the pre-conditions that are necessary for MPC to succeed, the risk that data might inadvertently or maliciously be disclosed to someone other than the output party, and what recourse to take if this bad event occurs.

We have two goals for this work: explaining why the privacy law questions are nuanced and that the lawyer is correct to proceed cautiously, and providing a framework that lawyers can use to reason systematically about whether and how MPC implicates data privacy laws in the context of a specific use case. Our framework revolves around three questions: a definitional question on whether the encodings still constitute 'personal data,' a process question about whether the act of executing MPC constitutes a data disclosure event, and a liability question about what happens if something goes wrong. We conclude by providing advice to regulators and suggestions to early adopters to spur uptake of MPC. It is our hope that this work provides the first step toward a methodology that organizations can use when contemplating the use of MPC.

1 INTRODUCTION

A U.S. Senate bill 681 that was introduced in 2019, called the "Student Right to Know Before You Go," envisioned a data system that would calculate innovative new metrics related to post-secondary education. For example, one metric described in the bill is the "annual earnings from employment, of students who enrolled in the institution of higher education ... disaggregated by program of study and credential received; the State in which the student is employed; and completion status" [56]. Essentially, the idea was to identify a data linkage between:

(1) University student records, which contain information about declared majors and graduation status, and
(2) IRS income tax filings, with employment status and wages.

Both of these datasets are subject to stringent data privacy laws. However, the relevant data privacy laws appear to be in conflict: it is extremely difficult for the U.S. federal government to release tax data (unless one obtains consent from individual taxpayers), and conversely the federal government is prohibited from receiving student unit records.[1] So: if the bill had passed, the next question would have been: how could this linkage be performed? To address this issue, Senate bill S.681 stated that the envisioned data system would "use *secure multiparty computation* technologies or [another technology that] delivers greater student privacy and security" [56] (emphasis added).

The technological hope. Secure multiparty computation, which we will abbreviate as MPC, enables data science without data sharing. Using cryptography, MPC allows participants to encode and federate their data across several computing parties—such as cloud providers—in such a way that no individual computing party can decode the data, and yet collectively the computing parties can perform data linking and aggregation. MPC has been an active area of research for 40 years, with substantial improvements over the

*The opinions in this paper reflect the views and work of the author as a Visiting Clinical Assistant Professor at Boston University between 2018-19.

[1] 20 U.S.C. § 1501c. In fact, the Department of Education already publishes a College Scorecard which reports historical income statistics by institution and field of study (with differential privacy [34]). The IRS addressed the data access problem by including only data from students who received federal financial aid, whose data the IRS already has as a result [53].

past decade in the efficiency and usability of software frameworks for secure data analysis (e.g., [7, 11, 14, 24, 29, 30, 33, 40, 59]).

As a result, companies, non-profit organizations, and government agencies are considering the value of MPC to perform commercially or socially beneficial analyses of non-disclosable data. There are a variety of successful deployments of MPC technology to date, such as: providing sexual assault survivors with a privacy-respecting way to identify fellow survivors based on the ideas of the #MeToo movement [39], training a machine learning model for smartphone keyboard predictions [10], measuring the gender wage gap in Boston [38], safeguarding cryptographic key material [44, 51], and calculating public health metrics about the spread of COVID-19 using smartphones [22].

Upon further inspection, a curious pattern emerges from these successful tech transitions. Each example involves data that are potentially considered very sensitive by the contributors of the data. But *legally*, these same data were not subject to a data privacy statute or regulation that restricts its use, processing, or disclosure.

The law and policy challenge. We focus in this work on data privacy laws at the U.S. federal level, where there is a patchwork of regulations that apply to specific sectors, industries, and/or types of data. For example, there exist separate statutes related to the specific privacy and security standards of: covered healthcare entities[2] and their business associates,[3] educational institutions that hold student records,[4] financial institutions that hold customer's non-public personal information;[5] consumer reports that are compiled and disclosed by consumer reporting agencies,[6] and data about children under the age of 13.[7] We occasionally touch upon state laws that introduce another layer of privacy questions, and cross-border data transfers that may implicate other regimes like the European Union's GDPR and its objective of data protection by design and by default.[8] All of these laws specify some kind of personally identifiable or protected health information that is subject to privacy protections; for generality we will refer to all such information as *personal data* in this work.

One thing that all of these laws have in common is the difficulty of adjudicating whether a new technology comports with the privacy regulation. As we will see, MPC is particularly difficult to analyze because its security guarantees and assumed preconditions don't map cleanly onto the concepts embedded within privacy laws, which have traditionally made a fundamental (yet implicit) assumption that meaningful data use requires access to the data in clear text. Frustratingly, this legal dilemma occurs even though MPC often can be superior to privacy law in terms of the actual protections it provides to individual privacy.

These challenges are compounded when data is governed by *multiple* legal regimes instead of just one. For example, it may be impossible to comply with all regulations if each party cannot disclose any aspect of their personal data to the other side (as in Senate bill S.681), or if one participant has requirements on data retention and access whereas another party must ensure data minimization and deletion. In scenarios involving data sets that are subject to multiple legal restrictions, we are not aware of any research addressing how a system that computes over encoded data may itself provide legal compliance. If every potential adoption must "reinvent the wheel" and perform a legal analysis from scratch, this would pose a major barrier to adoption of MPC in the very applications that would would benefit the most from its use. This article aims to identify the core legal questions in that analysis and provide a framework for decision-makers and their legal advisors to use in analyzing those questions.

The aim of this work. This work seeks to promote the use of MPC in scenarios involving *multi-regulation computing*: that is, data analyses with multiple parties whose datasets might collectively be subject to several data privacy regimes that restrict the personal data they can send and receive, such as with the proposed Senate bill S.681. We remark that MPC may be necessary here because these disparate regimes may make it difficult and costly—if not impossible—for input parties to find any single organization to whom they could all legally disclose their personal data. In this article, we aim to understand whether MPC is sufficient.

We do not believe there is a simple, one-size-fits-all argument to be made about the legality of employing MPC in all circumstances. So we do the next best thing: offer a conceptual framework for attorneys and technologists to use when making this evaluation on a case-by-case basis. We identify some common features in the way that various privacy laws impose limits on data sharing, use, processing, and disclosure. Then, we explore whether and how the use of MPC comports with these restrictions.

Concretely, we contribute a three-part framework to reason about the interplay between privacy regulations and computing over data in a cryptographically protected manner.

- As a definitional matter, determine whether the encodings used within MPC constitute 'personal data' that is afforded protections under the law (§4).
- From a process perspective, consider whether the execution of an MPC protocol infringes upon any restrictions on data disclosure (§5).
- Evaluate liability risk and recourses if something goes wrong, using legal instruments like contracts to reinforce or supplement MPC's technical guarantees (§6).

Though the framework is general, applying it will require taking account of the specific data, regulatory context, analysis being performed, and MPC configuration.

MPC has the potential to enable new data analyses while preserving the underlying privacy of the regulated data and, in many cases, meeting individuals' privacy expectations. But nothing about MPC changes the fact that data analysis can be used for good and for ill, and the results of analysis may themselves be harmful or illegal (as detailed next). We emphasize that the purpose of this work is not to argue that MPC allows parties to avoid data privacy regulations—rather, we examine how MPC and privacy regulations

[2] 42 U.S.C. § 1301 et seq.
[3] See 45 C.F.R. §§ 160.102-103.
[4] 20 U.S.C. § 1231 et seq.
[5] 15 U.S.C. § 6801 et seq.
[6] 15 U.S.C. § 1681 et seq.
[7] 16 C.F.R. § 312.
[8] GDPR Art. 25.

might interact and open a discussion about whether such regulations are fit for purpose as MPC and related privacy technologies continue to mature.

Scope and related work. This project complements and builds on other research efforts at the intersection of law and technology. In particular, we highlight cross-disciplinary research into differential privacy, and the extent to which differentially private outputs comply with the protections afforded to individual input records under FERPA and the GDPR [2, 13, 35]. More broadly, there exist several works that consider reidentification risk whether a released dataset is sufficiently anonymous in order to satisfy data privacy laws (see, e.g., Rubenstein and Hartzog [42]).

For this reason, we declare out of scope for this work the question of whether the output of the data analysis is safe to reveal (legally or morally). As a starting point for this work, we presume that a lawyer has already undertaken the effort to determine that the output is acceptable.

Instead, we focus on the independent (yet complementary) task of evaluating the *process* of securely computing the desired data analysis. Our goal is to understand whether the entire computing procedure is sufficiently de-identified as to be compliant with legal limits on use and disclosure. This question is particularly important in multi-regulation computing scenarios, where having a trusted curator (as in prior work) seems unattainable due to the difficulty of co-locating input data. Put another way: in this work, we focus exclusively on the *new* legal disclosure questions introduced by the act of cryptographically secure computing itself.

Finally, in this work we explore data privacy laws in the United States that affect the private sector—including companies, non-profit organizations, and individuals. Despite our mention of S.681 above, we don't delve into laws that specifically restrict the dissemination of data to and from the government. We note that others have examined how MPC interacts with the GDPR, such as [25, 43, 46] and the works cited in the Legal chapter of a recent United Nations handbook [52].

2 SECURE MULTIPARTY COMPUTATION

In this section we delve into more details about secure multiparty computation, or MPC. This technology allows parties to perform a calculation together over data that remains siloed. At a high level, MPC accomplishes this by using a mechanism called *secret sharing*, which is a way to split a secret into multiple shares subject to two properties: (1) with enough shares the secret can be reconstructed, and conversely (2) with too few shares one cannot learn anything about the underlying secret. (Here, the word "share" should be thought of in the sense of a stock share, not in the sense of sharing data.)

Additionally and importantly, there is a way to compute over these shares. For instance, if many people hold different encoded shares of two secrets r and s, there exists a way to work together to calculate the encoding of the sum $r + s$, or the product $r \times s$, or any other function of the encoded secrets; we refer interested readers to [16, 32] for more details about this process. In this way, cryptographically secure computing techniques like MPC offer the promise to calculate socially beneficial metrics that may otherwise be impossible or near-impossible due to the challenge of complying with multiple, distinct regulatory regimes (even if all parties happen to trust each other).

Setup. We consider a set of m input parties $P_1, P_2, \ldots, P_m$ who possess input datasets $x_1, x_2, \ldots, x_m$ that may each contain personal data that is subject to one or more data privacy regulations. We presume that the input parties have already agreed upon a data analysis f to perform. The input parties generate secret shares of their data and distribute these shares to n computing parties $C_1, C_2, \ldots, C_n$.

With the secret shares in hand, the computing parties can jointly perform the data analysis and provide the result $f(x_1, \ldots, x_m)$ to a specially-appointed output party. The encoding mechanism provides the guarantee that no coalition of less than t computing parties can recover any information about any input party's data or any intermediate state, even as they jointly contribute toward the calculation of $f(x_1, \ldots, x_m)$. However, a coalition of size t or greater *can* learn all data.

Cryptographers have designed and developed a wide variety of MPC protocols and software implementations [16, 24, 32]. Options exist for different choices of t and n; for example, when $t = n$ then all computing parties are required to reconstruct the output, and if any one of them refuses to participate then all data is irrecoverable. MPC also supports different configurations of input, computing, and output parties; they can all be identical, disjoint (this case is often called *outsourced MPC*), or anything in between. Furthermore, some protocols can detect or withstand a *malicious* adversary who can deviate from the protocol arbitrarily and provide incorrect encodings to the honest parties.

Applications. While still an active area of research, MPC has also received substantial tech transition in the past decade or so. It has been deployed to protect data commercially in the healthcare [3, 21, 41], education [8, 18], finance [1, 9, 15], and technology [10, 22, 27] sectors. Additionally, MPC has been piloted and used within many public sector [23, 47] and non-profit civic benefit [31, 38, 39] applications.

That said, MPC deployments to date in the United States tend not to come into direct confrontation with a data privacy law. Instead, they involve input data that is:

- Viewed as personally sensitive but not actually covered under a privacy law, such as Project Callisto's effort to provide survivors of sexual assault with a privacy-preserving way to report their experiences and identify fellow survivors so they can act together [39],
- Deemed 'deidentified' or otherwise acceptable to use consistent with data privacy regulations, such as Apple and Google's use of MPC to calculate metrics about COVID-19 exposure rates from smartphone applications with opt-in notice and choice [22], or
- Already protected to the same extent as would be required to perform data processing in the clear, such as the pilot of VaultDB to perform healthcare analytics using MPC on HIPAA-compliant servers [41].

Focus of this work. Regarding the configuration of parties: we focus on the case where $t < n$ and the input, computing, and output

parties are all disjoint because it is most challenging setting from a privacy law standpoint. Also, we presume that the coalition of fewer than t computing parties is *malicious*; the error-correcting properties of secret shares ensure that the computation maintains privacy and integrity in the face of such an attack.

Throughout this work, we consider applications of MPC where the inputs involve data that is subject to one or more privacy laws. Conversely, we presume that the output of the data analysis is safe to reveal to the output party, perhaps because it is not subject to a data privacy law or because the output party is authorized to read it under a data privacy law. Performing a legal analysis of the safety of revealing the output is outside the scope of this work. Instead, our focus is to examine the *use and disclosure* of information during the *process* of calculating the analytic via MPC.

3 OVERVIEW OF LEGAL AND POLICY ISSUES

Intuitively, MPC offers the promise of conducting data science in a way that—so long as the output is socially desirable and legally acceptable—can meet or even exceed most people's conceptions of data privacy for protecting personal data while in use. With MPC, we can correlate individual records across multiple datasets without revealing the underlying records, we can conduct aggregate analysis across datasets which parties are otherwise unwilling to share for competitive reasons, and we can analyze aggregate statistics across datasets which no individual party may lawfully hold.

A computer scientist might reasonably think that transforming any data analysis into its privacy-protective variant using MPC would be a clear win. On the other hand, a lawyer will likely want to move much more carefully, asking several important questions about the pre-conditions that are necessary for MPC to succeed, the risk that data might inadvertently or maliciously be disclosed to someone other than the output party, and what recourses to take if this bad event occurs.

We have two goals for this work: convincing readers that the lawyer is correct to proceed cautiously, and providing a framework to help the lawyer reason about questions regarding the applicability of MPC to any use case. It is our view that, in the absence of a public document describing the legal implications of using MPC, adoption of the technology for truly sensitive data will always be limited by the fact that every lawyer will be asked to "re-invent the wheel" and perform a thorough analysis of the technology from scratch. It is our hope that this work provides a step toward a holistic framework that companies, non-profit organizations, and governments can use when contemplating the use of MPC.

The incompleteness of MPC's security guarantees. A common way that cryptographers think about MPC is that, mathematically, it is just like having a magical "black box" where everyone can provide their input and the analyst receives the desired output. From this idealized perspective, it may be tempting to think that a lawyer need only determine whether the inputs and output constitute personal data under a data privacy law. However, we argue below that this approach misses a crucial third step: analyzing the process of executing a real MPC protocol.

We provide a simple argument to demonstrate to the computer scientist that the mathematical security guarantees provided by MPC might not always lead to sufficient legal protection as required by a data privacy law. We present the argument in the setting of outsourced MPC (though it generalizes), looking from the perspective of one computing party C^*.

Let's consider what happens if the input parties disclose secret shares of their personal data to the computing parties. If no single entity can be simultaneously authorized to receive all of the input data in the clear, there must exist at least one protected data class that C^* is not authorized to hold. If $t-1$ other computing parties have had their shares breached, then knowing C^*'s state becomes equivalent to knowing the protected input. This demonstrates that the input parties' initial act of dispersing secret shares might, in some cases, implicate a legal disclosure.

The possibility of such disclosures, and their dependence on the actions of other computing parties, necessitates *a priori* consideration. The analysis proposed in this work revolves around the applicability of data privacy laws to secret shares, whether secure computation constitutes a disclosure event, and which parties are liable in the event of a breach or other error.

The challenges of mapping MPC onto the law. What went wrong here? At a high level, the issue is that the computer security guarantees that MPC technology provides are not identical to, and in some cases do not map neatly onto, the legal restrictions imposed on data use and sharing. Our core contribution is to identify three specific issues when mapping MPC technology onto the law.

First, it may be unclear whether a statute or regulation's *definition* of personal data applies to secret shares used for computation in an MPC protocol. As illustrated by our simple example above, we will argue in §4 that typically data privacy protections *should* be afforded to secret shares of personal data.

Second, one must decide whether the use of MPC constitutes a *disclosure* of personal data under the law. The answer to this question depends on the reasonableness of assuming that the computing parties won't collude and will restrict themselves only to calculating the desired function f. The upshot here is that it's not possible to make a blanket statement that "MPC ensures legal/regulatory compliance." As we discuss in §5, whether a particular system complies with a particular set of regulations will often depend on the design and implementation of the protocol, as well as the specific restrictions in the legal and regulatory framework governing the personal data used in the protocol.

Third, despite our best intentions it might be possible that adversarial actors can in some circumstances undermine some or all of the benefits which the MPC protocol provides. This can in turn lead to a legal risk: a good actor who does not use or reveal any data inappropriately may find that the bad actions of another party could impact their own compliance with the law. In some sense, this risk is inherent in the nature of a protocol that involves numerous parties. This is another reason that categorical statements suggesting that MPC can ensure legal or regulatory compliance must be qualified by context. The challenges arising from adversarial behavior—and the "toolkit" one might use to address relevant risks—are discussed in more detail in §6-7.

4 DEFINITIONAL QUESTION: ARE SECRET SHARES PERSONAL DATA?

The first question in our framework is a definitional one: do the secret shares of data used in an MPC protocol constitute personal data that is protected under the law? Focusing on the effect of secret sharing, we assume that the input parties' raw data is itself personal data. Recall that throughout this work, we use the term "personal data" as a catch-all term for the various kinds of data afforded protections under the law—such as personally identifiable information held by a financial institution, electronic personal health information held by a healthcare entity regulated by HIPAA, or educational records held by an educational institution subject to FERPA.

This may be the most fundamental question in our framework because if the answer were to be 'no,' then data privacy laws may not impose any restrictions on data use, processing, and disclosure (as with some prior deployments of MPC described in §2) and we could stop the analysis here. Looking ahead, in this section we discuss why the answer to this question should often be 'yes' but also that it is context and use case-dependent.

The primary challenge in this section is that secret shares do not clearly map onto the existing legal definitions. In many typical cases, whether you're dealing with "personal data" is a fairly straightforward question. But what happens when the input parties disperse secret shares of their data to the computing parties? For instance, suppose that a university and hospital wish to run an MPC protocol where the two parties act as both the input and computing parties—perhaps the university aims to analyze whether students' health conditions or number of health appointments are correlated with changes to individual student's grades over time. However, the university isn't a HIPAA covered entity and the hospital isn't subject to FERPA. So, when feeding personal data as input to an MPC computation, it isn't obvious whether some—or all—computing parties are receiving personal data. Are the secret shares–unintelligible to the computing party–considered personal data in the hands of that entity?

The answer to this question depends on several factors that are specific to the circumstances of any deployment of MPC:

- The type of data, i.e., which privacy laws are implicated. U.S. sectoral privacy laws tend to identify personal data based on specific features of the data (e.g., does it include the subject's name?). By contrast, omnibus privacy laws like the GDPR and CCPA can apply to all features within a dataset about people.
- The data recipients and the regulations that apply to them. As described previously, the easiest case to consider is when all computing parties are already subject to the same restrictions on use and disclosure, such as performing a secure computation between several healthcare entities on HIPAA-compliant servers [41].

We explore this question in more detail with respect to specific data privacy laws involving financial, healthcare, and education data in the United States and one aspect of the European Union's GDPR.

FCRA & consumer reports. The FCRA regulates consumer reporting agencies (CRAs), which are entities that are engaged in the business of providing consumer reports to third parties. The law places certain limitations on how data may be shared and enshrines certain consumer rights, such as the right to correct inaccurate information. The regulated data—consumer reports—are "any written, oral, or other communication of any information by a consumer reporting agency bearing on a consumer's credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living which is used or expected to be used or collected in whole or in part for the purpose of serving as a factor in establishing the consumer's eligibility for—(A) credit or insurance to be used primarily for personal, family, or household purposes; (B) employment purposes; or (C) any other purpose authorized under section 1681b of this title."[9]

It is not immediately obvious how this definition would apply to secret shares in an MPC protocol. For example, imagine that a security startup offers to serve as a computing party–facilitating the computation over secret shares of encoded credit reports–for an alternative credit provider who wants to combine different datasets to compute creditworthiness in a new way. The startup will not see the credit results (those will go to the alternative credit provider); it will only perform the computation. Is the startup a CRA? The protocol inputs are consumer reports. The startup is engaged in the business of communicating information from those consumer reports to a third party, the alternative credit provider. And the startup computes and sends results so that the information from the consumer report can be used in credit decisions. However, the secret shares that the startup receives are indecipherable and, because the consumer reports are split among multiple parties, it may not be linkable to an individual consumer. Do indecipherable secret shares constitute information "bearing on a consumer's creditworthiness"? If the answer is "no," then MPC would appear to provide a loophole through which information from consumer reports could be produced to credit providers without being subject to the strictures of the FCRA. But if "yes," then the startup bears the regulatory burden applicable to consumer reporting agencies even though it would be difficult if not impossible to fulfill those obligations (the startup cannot access or edit the plain text of the consumer reports).

HIPAA. Protected health information (PHI) means individually identifiable health information held by certain healthcare providers and their affiliates.[10] Health information is "individually identifiable" if it is created by a healthcare provider and includes demographic information about the relevant person and relates to that person's "past, present, or future physical or mental health or condition," health care provision, or payment for health care provision" and identifies the person or there is "a reasonable basis to believe the information can be used to identify" the person.[11] This definition raises similar questions to those mentioned above–are secret shares generated from individually identifiable health information regulated to the same extent as the source data? Does that depend on whether a computing party could actually identify an individual if it was able to decipher the secret shares in its possession? Does

[9] 15 U.S.C. §1681a(d)(1)
[10] 45 CFR §160.103
[11] 45 CFR §160.103

the "reasonable basis" test require analyzing the data in one computing party's possession, or all of the data held by all computing parties?

Under what circumstances secret shares do or don't constitute PHI will often depend on the application of HIPAA's Expert Determination Method for de-identifying PHI. It requires that a domain expert must determine that "the risk is very small that the information [under consideration] could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual ..." [54]. In the outsourced MPC setting, the secret shares given to any one compute party considered in isolation cannot "identify an individual;" but, they can when combined with $t - 1$ other parties' shares. To the latter point, the Department of Health and Human Services has provided guidance about sharing two or more datasets based on the same input data. "In such cases, the expert must take care to ensure that the data sets cannot be combined to compromise the protections set in place through the mitigation strategy. ...The expert may certify a covered entity to share both data sets after determining that the two data sets could not be merged to individually identify a patient. This certification may be based on a technical proof regarding the inability to merge such data sets. Alternatively, the expert also could require additional safeguards through a data use agreement." [54]. We emphasize that any such contracts are only one aspect of determining whether other shares are "reasonably available" and to what "anticipated recipients"; we will say more about contracts in §6.

FERPA. FERPA regulates the release of "education records," which are "records, files, documents, and other materials which— (i) contain information directly related to a student; and (ii) are maintained by an educational agency or institution or by a person acting for such agency or institution."[12] Regulations require educational agencies or institutions to get prior written consent before disclosing personally identifiable information from education records, subject to certain exceptions.[13] The regulations also provide that prior consent is not required if the agency or institution only releases "education records ...after the removal of all personally identifiable information provided that the educational agency or institution or other party has made a *reasonable determination that a student's identity is not personally identifiable*, whether through single or multiple releases, and taking into account *other reasonably available* information"[14] (emphasis added).

As a result, if an educational institution provides input to an MPC computation, the question of whether the shares are protected in the first instance might depend on the reasonableness of believing that the input data cannot be identified. We will explore this topic in more detail in §5-6.

The GDPR & pseudonymization versus anonymization. The European Union's General Data Protection Regulation (GDPR) is generally outside the scope of this paper, which focuses on U.S. privacy laws. We mention it here just to note one interesting definitional question that is similar to those above.

Recital 26 of the GDPR distinguishes between *pseudonymous* and *anonymous* data. Pseudonymous data includes any information that can be attributed to a single individual when linked to additional information using any means "reasonably likely to be used" to identify that person.[15] Pseudonymous data is still considered personal data, in contrast to *anonymous* data—data that (the GDPR helpfully clarifies) has been "rendered anonymous" such that it is not individually identifiable—which is outside the scope of the regulation.[16] Thus, whether secret shares are regulated by GDPR likely turns on what means are "reasonably likely to be used" to transform the secret shares into identifiable information. We observe that a single secret share does not enable *predicate singling-out attacks*, passing a test that prior work argues is necessary for anonymization under GDPR [2]. But that test is not a sufficient condition for anonymization and therefore does not resolve the question.

5 PROCESS QUESTION: DOES MPC CONSTITUTE DISCLOSURE?

We next move to a process question: if secret shares constitute personal data, then should the act of participating in a secure multiparty computation protocol—which involves sending shares between parties over a network—be considered as a data disclosure or processing? This may be the most complex and nuanced question in our framework, and as such it is difficult even to give a full list of considerations.

As a result, we begin by briefly listing a few of the salient features shared across various privacy laws that may be relevant to the question about whether using MPC constitutes a data disclosure. Then, we describe several factors about a deployment of MPC that may influence how privacy law apply to the operation, such as whether it is reasonable to believe that the computing servers will not collude to reconstruct the data based on organizational and contractual relationships, and the technical and social processes in place to ensure that the data processing system only performs the approved analysis. While it is difficult to provide a generic answer to this question, we conclude this section by analogizing some aspects of MPC to encryption, and explore how that analogue might help technologists and attorneys parse whether using MPC constitutes disclosure under applicable privacy statutes.

5.1 Design and intent of data privacy laws

In this subsection, we examine a few salient features of disclosure limitations that appear in various data privacy laws. We describe the tradeoffs inherent in the concept of privacy and information sharing, and how different styles of privacy laws define the privacy-relevant activity and who they empower to decide between the tradeoffs regarding data use, disclosure, and processing.

Tradeoffs. Laws that regulate privacy and data sharing rarely prohibit information sharing outright. Instead, most privacy laws recognize that there are *tradeoffs*, and seek to strike a balance between competing goals. These goals may include: (a) empowering the person to whom the data pertains with adequate notice and

[12] 20 U.S.C. §1232g(a)(4)
[13] *See* 34 C.F.R. §§99.30, 99.31(a)
[14] 34 C.F.R. §99.31(b)(1)
[15] GDPR Art. 4(5).
[16] GDPR Art. 4(1); Recital 26(5)

choice about whether to share their personal data, and how that data may be used and shared; (b) empowering the recipient(s) of personal data to use that data consistent with (a); (c) facilitating economic or other socially beneficial activity consistent with (a) and (b); and (d) advancing other values, such as a sense of personal autonomy and privacy or other social goals. We illustrate this point with two examples from the domain of finance.

First, Congress enacted the FCRA to ensure that consumer reporting agencies can create data-driven credit reports while also providing reasonable procedures for using this information "in a manner which is fair and equitable to the consumer, with regard to the confidentiality, accuracy, relevancy, and proper utilization of such information."[17] Note the emphasis on the tradeoffs—that the procedures must be reasonable, and the context for sharing appropriate.

Second, the Financial Modernization Act of 1999, colloquially the Gramm-Leach-Blilley Act or GLBA, prescribed certain notice and privacy standards for personally identifiable financial information. Congress enacted GLBA to underscore that each financial institution has an "affirmative and continuing obligation to respect the privacy of its customers and to protect the security and confidentiality of those customers' nonpublic personal information."[18] Yet it also recognized the importance of information sharing, commissioning studies to examine the risks and the benefits of sharing financial information to consumers and financial institutions[19] and to explore when consumers may wish to direct institutions to share their private information for a variety of purposes and ends.[20]

Regulating process versus outcome. A fundamental choice for all data privacy regimes is the extent to which they regulate process or outcome. Actually there are two questions here that we tease out separately:

(1) Whether the data privacy law restricts the act of *data processing*, or simply the resulting *output* of the analysis.
(2) How the data privacy law influences the *process* of navigating tradeoffs and deciding on a privacy decision, or prescribes a desired *privacy outcome*.

The first question is easier to address. Both the law and computer science tend to agree that the concept of privacy is relevant to *how* data is used (processed) as well as the output data that *results* from a process or operation. There may be a variety of reasons for this—this might reflect how most people think about their own privacy, the process might reveal the intent for data use (which people might care about), or it might be more informative as to the tradeoffs inherent in the operation. But, data privacy laws can differ substantially in terms of the extent to which they regulate data processing.

In the remainder of this section we focus on the second question: who is empowered under the law to make decisions about privacy choices, and how they may navigate among the tradeoffs involved. When we use the words 'process' and 'outcome' in this section, we always refer to the meaning in question 2. We stress that the process of navigating between tradeoffs and selecting a privacy policy has nothing to do with the concept of data processing itself.

Privacy laws in the U.S. tend to begin by emphasizing the principle of consumer choice: most privacy laws begin by requiring the entity to provide notice to the consumer, and then set various defaults around how the entity may collect, use, and disclose personal data to third parties—including by requiring specific notice or more explicit consent before the entity may *disclose* sensitive data to third parties, or requiring entities opt-outs from certain practices. Some defaults are imposed by law, but also flexibility is provided for consumers to opt out of the defaults and make different choices with relatively little friction. Consider the GLBA for example: rather than requiring a particular *privacy outcome*—like prohibiting the sharing of an individual's financial information in all contexts—GLBA and associated rules focus on process constraints and providing consumers with information about how their data is shared. One partial exception to the above is HIPAA, where the guiding principle is to protect patient records either to a pre-specified standard or based on the determination of a privacy expert.

As a consequence, U.S. sectoral privacy laws generally tend to provide broad latitude for use by any authorized holder of data, and the restraints imposed within privacy laws typically involve data security requirements and prohibitions on the unauthorized *disclosure* or sharing of information. Some U.S. privacy laws additionally impose some restrictions on the *purposes* for which data may be used, such as the Fair Credit Reporting Act. By contrast, the EU and some state and local laws in the United States impose broader and more general constraints on the *processing* of personal data. These distinctions may be relevant to examining how MPC fits within a particular regulatory framework.

5.2 Opportunities and challenges of MPC

Used properly, a tool like MPC can help to *foster the positive effects* of data usage while *mitigating potential negative impacts* of data sharing. But just as we saw in §4, it can be a challenge to map the benefits of MPC onto the space of tradeoffs considered within a data privacy law. In this section, we describe the extent to which MPC implicates data privacy laws, questions that may influence adherence to the law, and how the complexity grows when analyzing data that spans multiple jurisdictions or regulations.

How MPC impacts the process of deciding between tradeoffs. It is easy to see why the process constraints from §5.1 are useful in the use case that the drafters likely imagined—the sharing of personal data that can be read in the clear by a third party recipient. Here the privacy and security concerns are clear.

But if consumers' private data is instead used in an MPC protocol, the privacy and security concerns may be different, and the process constraints imposed by U.S. data privacy laws may not fit the use case. For example, even though personal data may be in some sense shared with third parties, if it cannot be decoded by those third parties, it is not clear whether additional notice of the data sharing imposed by GLBA will improve the consumer experience or addresses consumer concerns. In other words, the technical requirements of the law might apply—we discuss further below—but there may be a mismatch between the controls which the law

[17] 15 U.S.C. §1681
[18] 15 U.S.C. §6801
[19] *See* 15 U.S.C. Sec. 6808(a)
[20] 15 U.S.C. §§6808(4), (5), (9)

imposes and the risks or potential harms that a consumer might face. Indeed, MPC might provide greater privacy protections than the law requires, but the law's technical notice and choice requirements might still be imposed in addition to the use of MPC.

Reasonableness of beliefs. Data privacy laws sometimes require that decisions about security practices adhere to a standard of reasonableness, or that the action can be explained and defended to a neutral observer. In the context of MPC, all participants need to justify their belief that t or more computing servers will not collude to reconstruct the data. For example, if the computing parties were subsidiaries of the same parent company or have a history of collaboration, then it may be unreasonable to believe that the parties will adhere to MPC's non-collusion requirement.

Legal instruments such as non-disclosure agreements or information fiduciary relationships [4] can foster a reasonable belief in non-collusion. We will have more to say about contracts in §6 when we discuss liability when things go wrong, but here we stress that the input parties must be able to defend *a priori* their decision that the computing parties won't collude—even if it later turns out that everyone acts honestly in the execution of the MPC protocol.

Process restrictions. There is yet another consideration: what protections exist on the analytic f to be computed and on the results of the computation? Recall that we start from the stipulation that the output (on its own) of the agreed-upon analytic f would be legal to disclose to the output party. Even so, the math of an MPC protocol does not limit how the data may be queried, and so one must consider:

- The data security protections on the output, to ensure that it is only revealed to the appropriate party.
- The procedures put in place to limit the computing parties only so that they only perform the agreed-upon analysis.

The first consideration can be addressed through standard data security protection mechanisms. The second topic merits further discussion.

There are a variety of technological and social methods that can proactively ensure that the computing parties only perform an agreed-upon computation, or retroactively audit that they have performed their role properly. A limitation procedure might include several of the following for defense in depth:

- Obtain written or electronic approval from all input and computing parties about the agreed-upon set of possible statistics.
- Create a log of every computation performed, either locally at each computing party or within a public ledger.
- Use a publicly verifiable MPC protocol so that everyone can check that the computing parties performed their roles correctly (e.g., [12, 19, 36, 37]).
- If multiple analytics are allowable, augment the MPC protocol itself to perform a policy check that each calculation is within the agreed-upon set (e.g., [20, 57, 58]).

Additionally, we remark that disclosure limitation techniques like differential privacy are compatible with MPC and should be considered as well as part of a holistic evaluation about data disclosure [2, 35].

The challenge of crossing jurisdictions. We have already illustrated the challenge of reaching any firm conclusion on the legal sufficiency of using MPC, even if a specific legal or regulatory context is specified. Whether a specific protocol provides protections that meet a party's legal and regulatory obligations depends not only on the relevant laws governing the underlying datasets, but also the nitty gritty details regarding what the protocol will allow the parties to compute, who may see the results, and what, if any, limits are placed on the number and type of queries allowed.

If data from multiple jurisdictions is involved, then the different legal approaches to privacy regulation (see §5.1) may mean that MPC presents different legal or policy questions in different jurisdictions. The best case scenario is that one needs only to analyze how the MPC protocol fares under each of the privacy laws involved; however, sometimes it can be much more challenging or impossible. Executing an MPC protocol across jurisdictional boundaries brings up new, complex questions about cross-border data transfers and data sovereignty that are beyond the scope of this article. Moreover, it might be impossible to comply with both regimes; for example, one regulatory regime might require data minimization and deletion where the other one requires data retention and accessibility.

5.3 Analogizing Encrypted Communications to MPC

In this subsection, we examine how MPC might be analogous to encrypted communications as a way to explore how existing practices might simplify the analysis of MPC's legality in some cases. As caveated above, there are no "one size fits all" legal theory for MPC—its legality will depend on the facts of the system and its deployment.

Concretely, we consider here the outsourced setting with $n = 2$ computing parties, such that the computing parties are distinct legal entities from both the input and output parties. Additionally, suppose that the recovery threshold of $t = 2$, meaning that both computing parties can collectively reconstruct the data but neither one individually can do so. In summary, the computing parties cannot see either the personal data that is input or the results of any analysis; they are only there to compute.

A novel analogy. Consider the similarities between MPC secret shares and encrypted communications. From the perspective of a single computing party, holding onto one secret share is akin to an Internet Service Provider who sees a ciphertext passing along its network that it doesn't know how to decrypt. Recall that the other computing party knows the other secret share, which in effect serves as a "secret key" that is needed to reveal the personal data. Conversely, in both cases the party who holds the secret shares/encrypted packet cannot, without that additional data which the computing party does not have access to, reveal personal data. (The circumstances are not perfectly analogous. By colluding with other computing parties, one computing party could reveal the input data without input from either the data originator or its intended recipient. And, disclosing the secret shares on the web might give other computing parties information about their own secret shares in a way that disclosing encrypted communications would

not reveal information except to the originator and intended recipient.)

In this analogy, encrypted data is *not* exempt from data privacy laws; several laws explicitly make this point (e.g., [55]). However, we are not aware of any regulator in the United States that has suggested that passing encrypted packets across the network constitutes disclosure, use, or processing of personal data as a legal matter. The benefit of this analogy is that it connects an unknown legal question—whether a secret share constitutes personal data and whether processing it constitutes a data disclosure event—to a better-known domain with guidance from regulators and courts.

Implications of this theory. What does this mean? We offer two possible conclusions from this theory, stressing upfront that it is untested in the legal system.

First, if regulators are sufficiently informed, and computing parties are sufficiently restricted in their ability to share with each other, there *might* be a clear path forward to treating secret shares as unencumbered data in the hands of the computing parties. The flip side of this argument is that if you pushed a regulator to think about it, perhaps they might conclude that passing packets across a network does in fact constitute disclosure, use, or processing of personal data. In either case, the regulator's decision about the treatment of encryption and MPC would be similar.

Another conclusion is that theories about what does or doesn't constitute personal data are driven more by the kind of accepted practice than by truly parsing the outer bounds of statutory interpretation. In other words, perhaps the popularity of encryption or the nature of the web either influences or creates unexamined assumptions about whether ISPs are or aren't processing personal data in this context. If MPC gains wider use and adoption without significant data leaks or security issues, then it *might* be the case that MPC technology inherits the same perceptions and assumptions. Admittedly this claim goes beyond the available evidence, so we offer a more conservative version of this conclusion. If MPC is used in a way that avoids giving personal data to parties who are not supposed to have it, then perhaps there is a low risk of legal challenges just as encryption has faced relatively few legal challenges, even if there remains some ambiguity about whether MPC fully meets all legal requirements.

Of course, analogies are imperfect. For instance, an ISP passes encrypted data between two endpoints from whom nothing about the data is kept secret—short of colluding with one of the endpoints, there is nothing the ISP can do to compromise the secrecy of the ciphertext. In contrast, each computing server in the outsourced MPC setting is communicating with other servers from whom the underlying data must be kept hidden. If these servers choose to collude, they could recover the secret data without any involvement of the originating input parties. This suggests that perhaps the specific processing involved in computing the intended analytic f under MPC should be treated differently than other processing that the parties may undertake without approval.

6 LIABILITY QUESTION: WHAT IF SOMETHING GOES WRONG?

The third and final question in our framework is about assigning liability. In this section, we catalog several ways that MPC could go wrong, so that lawyers can evaluate each of these risks in the context of an envisioned MPC deployment. Then, we describe how contracts can provide options to data holders who have been harmed by an accidental error or intentional attack, and can protect honest computing parties from being assigned blame or else they may not have an incentive to participate in an MPC data analysis in the first place.

How MPC can fail by accident. There are several ways that a secure computation could fail to provide the desired security guarantees through no fault or malice by anyone.

A particularly devastating, but relatively unlikely, concern is that the abstract mathematical algorithms used to define an MPC protocol are simply broken (as distinct from their implementation in software). Cryptographers strive to avoid this outcome by insisting upon a mathematical proof that any MPC protocol meets a formal security guarantee, but these proofs can be long and subtle, so sometimes there is a mistake that goes unnoticed for a long time. While such mistakes do occur (e.g., [5, 26, 45, 48]) and they can be incredibly damaging, they are still rare for peer-reviewed and popularly-deployed systems. An even more fundamental version of this issue is if an MPC algorithm relies on a cryptographic building block (like an encryption scheme or hash function) that is itself broken. For this reason, cryptographers tend to establish standards and best practices based on building blocks that have stood the test of time.

More commonly, sometimes the implementation of an MPC algorithm can invalidate the security guarantees that the mathematical algorithm would have provided. The likelihood of this bug depends on the creators' software development processes and the complexity of the code they produced. Relatedly, sometimes an MPC implementation adheres to a protocol specification, but it is vulnerable to a side-channel attack that was never considered in the algorithm's proof of security.

Finally, MPC algorithms can fail if the implementation is sound but the operating system, hardware, and networking infrastructure underneath it has a vulnerability. In principle this risk nearly always exists. That said, in most situations the additional risk created by using MPC (above and beyond the existing risks of breaches of an organization's enterprise network) is typically low. We observe that MPC might make existing latent vulnerabilities more exploitable because it might necessitate putting data on an Internet-connected machine that previously would have been airgapped. On the other hand, MPC adds complexity to this attack because an external hacker must subvert at least t computing parties.

How adversaries can attack MPC systems. Additionally, there are several ways that one or more MPC protocol participants can intentionally subvert MPC's security guarantees with a conscious attack. First, an internal party can exploit any of the accidental errors listed above. In this way, it can appear to follow the MPC protocol as designed and yet still learn or tamper with data.

Second, t or more parties might collude to reconstruct data. This is greater than the threshold that the MPC scheme was designed to withstand, and the error-correcting properties of the secret sharing scheme will allow this adversarial set to decode the data. Note that collusion by less than the threshold is already accounted for by

design in MPC and thus shouldn't be considered as a way that MPC can 'go wrong,' even accidentally.

Third, the creators of the MPC software—whether a whole firm or a rogue employee—might intentionally insert a vulnerability into the software that they know how to exploit. We remark that such supply-chain attacks tend to be a single point of failure (i.e., in practice all computing parties tend to run the same software), and they can be hard to distinguish from an accidental attack.

The Role of Contracts. Contract law can play a role here in establishing baselines for expected and disallowed behavior and indemnifying other parties in case of a breach. In more detail, contracts can:

- Reinforce technical guarantees by making certain requirements of the technology explicit or by requiring parties to behave in a particular way. Stating technical guarantees and requirements explicitly in a contract also provides a straightforward path to recovery if the guarantees or requirements are not met. The wronged party can simply sue for breach of contract, and need not identify other legal rights to exercise.
- Create incentives for good behavior or disincentivize poor behavior. Most obviously, such a contract could require restitution (redress losses) and indemnification (pay legal fees) for any losses caused by a party who failed to comply with their obligations or whose secret shares were revealed.
- Delineate a process by which any alleged bad behavior must be investigated. If all parties agree on an investigation process, the process is more likely to be implemented if something goes wrong, and the results are more likely to be accepted by the parties to the protocol.
- Provide for third-party auditing of certain aspects of the protocol, if helpful.

We provide an example that showcases the power of contracts in the next section.

7 ENCOURAGING ADOPTION OF MPC

With our framework complete, in this section we take a step back and consider how questions at the interface of MPC and the law might evolve over time. We look at actions that a regulatory agency or legislature might take to incentivize (or disincenivize) adoption of MPC, and we consider how awareness of the law might influence the design of future deployments of MPC.

The Regulator's Perspective. As MPC algorithms and software implementations mature and become more widely considered for adoption by governments, companies, and non-profit organizations, one important question will be whether and to what extent regulators conclude that the guarantees that MPC can provide may (in some cases) or do (categorically) fulfill the legal obligations of the parties to a computation. Intuitively, one might expect regulators to take a risk-averse or conservative position on the technology. Regulators may be particularly attuned to the risks posed by new technologies or acutely aware of the many ways human error can manifest in the technology and its implementation. They might also be skeptical that promises not to collude or behave in an adversarial manner will be respected. But lumping regulators together as a monolithic entity is likely a misnomer.

Some financial regulators and government agencies are taking a hands-on approach.

- A consortium of financial regulators in the US, including the Treasury Department, Federal Reserve, FDIC, FinCEN and others have created incentives for financial institutions to use existing tools or adopt new technologies to identify and report money laundering, terrorist financing, and other illicit financial activity [6].
- In 2019, The UK's Financial Conduct Authority brought together key stakeholders—including financial institutions, privacy regulators, technology companies, and law enforcement officials—to explore whether technologies including MPC may be adapted to improve anti-money laundering and to counter terrorist financing efforts [49].
- The JASON group wrote a report recommending that the U.S. Census Bureau "engage in a series of pilot projects to fully evaluate the potential of multiparty computation in Census Bureau surveys" involving business data like tax information consistent with its Title 13 and 26 obligations [28].
- At the multinational level, the UN Big Data Global Working Group has been working to improve understanding and adoption of privacy enhancing technologies like MPC [52].
- The U.S. federal government recently published a request for "public comments to help inform development of a national strategy on privacy-preserving data sharing and analytics, along with associated policy initiatives" [17]. Also, government agencies in the U.K. and U.S. announced a prize challenge competition involving privacy technology [50].

Conversely, there are also reasons why privacy enforcers in the U.S. might opt to take a more cautious or restrictive approach.

- Regulators may not want to opine on real edge cases, where the applicability of the law is unclear. The FTC, for example, tends to highlight "best practices" in its guidance documents, rather than focusing on the minimum requirements that an individual can fulfill without breaking the law.
- Given limited resources, regulators may prefer to focus on entities that they consider to be egregious wrongdoers, rather than entities whose use or sharing of data is closer to the line of legality. This behavior might be observed both in the issues they chose to opine on and in the cases they choose to bring.
- Related to both of the points above, regulators may be reticent to actually litigate edge cases like this unless they can identify concrete customer harm that resulted from the data sharing.
- Regulators may also simply prefer to take a wait-and-see approach, observing the development, implementation, and level of adoption of the technology before establishing a precedent or opinion on the matter one way or the other.

Overall though, signs are trending positive toward increased openness of legislatures and regulators toward the possibility of

using MPC. As discussed in §2, there are several tech transition successes that realize the benefits of MPC without the need for regulators to fully embrace the technology in the first instance. Implementing MPC in these contexts may provide evidence that these guarantees work and promote better understanding regarding the benefits and limits of the technology.

How Might Privacy Law Changes Impact MPC?. One way to think about how MPC fits within existing privacy frameworks is to consider how changes to existing legal structures might affect the adoption of MPC—either by making the tool more attractive, or by directly regulating its implementation. The discussion in this section is necessarily at a high level, since any consequences will depend entirely on the details of the legislation.

- Increased penalties for data breaches: Assuming an MPC system is properly configured, data breach rules are unlikely to directly regulate MPC. However, stricter breach penalties may encourage companies to adopt MPC in data licensing agreements, for example, to limit the possibility for a breach.
- Additional limits on sharing personal data with third parties: could incentivize adoption if MPC does not implicate any new restrictions on disclosure, as described in §5.
- New requirements to produce, revise, or delete consumer data upon request: these measures would not regulate MPC—the data originator would remain responsible for implementing relevant tools. MPC is, in principle, compatible with providing all of these rights.
- Certain restrictions on automated decision-making: some such restriction might regulate MPC directly because auditing or reviewing multi-party computations and outputs requires technical tools, and the system cannot be easily reviewed by an individual without technical know-how.
- Fiduciary duties: these measures are unlikely to apply directly to MPC, but may make MPC an attractive part of a privacy-protective toolkit. Additionally, as noted in §3, if MPC is used in the outsourced setting and a computing party dumps its share on the internet, then another party might be prohibited from reviewing those shares by its fiduciary duties if those duties prohibit the purposeful reidentification of data.

8 CONCLUSION

This work contributes a framework to reason about the legal implications of using MPC when the underlying personal data is subject to one or multiple data privacy laws. We provide three questions whose investigation will inform a legal analysis, and we provide guidance about the factors of any specific deployment scenario that may influence the requirements in sectoral privacy laws in the United States. We believe that this framework can increase adoption of MPC by lowering the cost of evaluating this new technology, and by showcasing opportunities where the use of MPC can lead to data analyses that would have been more costly or even prohibited by data sharing in the clear. It is also our hope that our work can spur future research into the co-design of technology, contracts, and regulations in order to allow for new ways to conduct privacy-respecting data analysis.

ACKNOWLEDGMENTS

The authors are grateful to Alexandra Wood, David O'Brien, and Patrick Baier for their helpful conversations about the ideas presented in this work. This material is based upon work supported by the National Science Foundation under Grants No. 1718135, 1801564, 1915763, and 1931714, by the DARPA SIEVE program under Agreement No. HR00112020021, and by DARPA and the Naval Information Warfare Center (NIWC) under Contract No. N66001-15-C-4071. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF, DARPA, or NIWC.

REFERENCES

[1] Aysajan Abidin, Abdelrahaman Aly, Sara Cleemput, and Mustafa A. Mustafa. An mpc-based privacy-preserving protocol for a local electricity trading market. In *CANS*, volume 10052 of *Lecture Notes in Computer Science*, pages 615–625, 2016.

[2] Micah Altman, Aloni Cohen, Kobbi Nissim, and Alexandra Wood. What a hybrid legal-technical analysis teaches us about privacy regulation: The case of singling out. *BU J Sci. & Tech. L.*, 27:1, 2021.

[3] David W. Archer, Dan Bogdanov, Yehuda Lindell, Liina Kamm, Kurt Nielsen, Jakob Illeborg Pagter, Nigel P. Smart, and Rebecca N. Wright. From keys to databases - real-world applications of secure multi-party computation. *Comput. J.*, 61(12):1749–1771, 2018.

[4] Jack M Balkin. The fiduciary model of privacy. *Harv. L. Rev. F.*, 134:11, 2020.

[5] Mihir Bellare and Phillip Rogaway. Entity authentication and key distribution. In *13th Conference on Advances in Cryptology (CRYPTO)*, pages 232–249, 1993.

[6] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Financial Crimes Enforcement Network, National Credit Union Administration, and Office of the Comptroller of the Currency. Joint statement on innovative efforts to combat money laundering and terrorist financing. https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20181203a1.pdf, December 2018.

[7] Dan Bogdanov, Sven Laur, and Jan Willemson. Sharemind: A Framework for Fast Privacy-Preserving Computations. In Sushil Jajodia and Javier Lopez, editors, *Proceedings of the 13th European Symposium on Research in Computer Security - ESORICS'08*, volume 5283 of *Lecture Notes in Computer Science*, pages 192–206. Springer Berlin / Heidelberg, 2008. ISBN 978-3-540-88312-8.

[8] Dan Bogdanov, Liina Kamm, Baldur Kubo, Reimo Rebane, Ville Sokk, and Riivo Talviste. Students and taxes: a privacy-preserving study using secure computation. *Proc. Priv. Enhancing Technol.*, 2016(3):117–135, 2016.

[9] Peter Bogetoft, Dan Lund Christensen, Ivan Damgård, Martin Geisler, Thomas P. Jakobsen, Mikkel Krøigaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, Michael I. Schwartzbach, and Tomas Toft. Secure multiparty computation goes live. In *Financial Cryptography*, volume 5628 of *Lecture Notes in Computer Science*, pages 325–343. Springer, 2009.

[10] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *ACM Conference on Computer and Communications Security*, pages 1175–1191. ACM, 2017.

[11] Boston University. JIFF: Client side library for performing MPC in JavaScript. https://github.com/multiparty/jiff-client, 2022.

[12] Ran Canetti, Ben Riva, and Guy N. Rothblum. Two 1-round protocols for delegation of computation. *IACR Cryptol. ePrint Arch.*, page 518, 2011.

[13] Aloni Cohen and Kobbi Nissim. Towards modeling singling out. In *Theory and Practice of Differential Privacy*, 2018.

[14] Henry Corrigan-Gibbs and Dan Boneh. Prio: Private, robust, and scalable computation of aggregate statistics. In *Proceedings of the 14th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 259–282, Boston, Massachusetts, USA, 2017. USENIX Association. ISBN 978-1-931971-37-9. URL https://www.usenix.org/conference/nsdi17/technical-sessions/presentation/corrigan-gibbs.

[15] Ivan Damgård, Kasper Damgård, Kurt Nielsen, Peter Sebastian Nordholt, and Tomas Toft. Confidential benchmarking based on multiparty computation. In *Financial Cryptography*, volume 9603 of *Lecture Notes in Computer Science*, pages 169–187. Springer, 2016.

[16] David Evans, Vladimir Kolesnikov, and Mike Rosulek. A pragmatic introduction to secure multi-party computation. *Foundations and Trends® in Privacy and Security*, 2(2-3):70–246, 2018. ISSN 2474-1558. doi: 10.1561/3300000019. URL http://dx.doi.org/10.1561/3300000019.

[17] Federal Register: The Daily Journal of the United States Government. Request for information on advancing privacy-enhancing technologies. https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-

for-information-on-advancing-privacy-enhancing-technologies, 2022.
[18] Joan Feigenbaum, Benny Pinkas, Raphael Ryger, and Felipe Saint-Jean. Secure computation of surveys. In *EU Workshop on Secure Multiparty Protocols*, pages 2–14, 2004. URL https://www.cs.yale.edu/homes/jf/SMP2004.pdf.
[19] Dario Fiore and Rosario Gennaro. Publicly verifiable delegation of large polynomials and matrix computations, with applications. In *CCS*, pages 501–512. ACM, 2012.
[20] Ben A. Fisch, Binh Vo, Fernando Krell, Abishek Kumarasubramanian, Vladimir Kolesnikov, Tal Malkin, and Steven M. Bellovin. Malicious-client security in blind seer: A scalable private DBMS. In *IEEE Symposium on Security and Privacy*, pages 395–410. IEEE Computer Society, 2015.
[21] Thanos Giannopoulos and Dimitris Mouris. *Privacy Preserving Medical Data Analytics using Secure Multi Party Computation. An End-To-End Use Case.* PhD thesis, National and Kapodistrian University of Athens, 09 2018.
[22] Google, Inc. Analytics in exposure notifications express: FAQ. https://github.com/google/exposure-notifications-android/blob/master/doc/enexpress-analytics-faq.md, 2021.
[23] Nick Hart, David Archer, and Erin Dalton. Privacy-preserved data sharing for evidence-based policy decisions: A demonstration project using human services administrative records for evidence-building activities. https://ssrn.com/abstract=3808054, 2019.
[24] Marcella Hastings, Brett Hemenway, Daniel Noble, and Steve Zdancewic. SoK: general-purpose compilers for secure multi-party computation. In *2019 IEEE Symposium on Security and Privacy (SP)*, 2019.
[25] Lukas Helminger and Christian Rechberger. Multi-party computation in the gdpr. In *Privacy Symposium 2022 - Data Protection Law International Convergence and Compliance with Innovative Technologies (DPLICIT)*, 2022.
[26] Dennis Hofheinz and Victor Shoup. GNUC: A new universal composability framework. *J. Cryptology*, 28(3):423–508, 2015.
[27] Mihaela Ion, Ben Kreuter, Ahmet Erhan Nergiz, Sarvar Patel, Mariana Raykova, Shobhit Saxena, Karn Seth, David Shanahan, and Moti Yung. On deploying secure computing commercially: Private intersection-sum protocols and their business applications. *IACR Cryptology ePrint Archive*, 2019:723, 2019.
[28] JASON Program Office. Secure computation for business data, November 2020.
[29] Marcel Keller. MP-SPDZ: A versatile framework for multi-party computation. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 2020. doi: 10.1145/3372297.3417872. URL https://doi.org/10.1145/3372297.3417872.
[30] KU Leuven. SCALE-MAMBA Software. https://homes.esat.kuleuven.be/~nsmart/SCALE/, 2022.
[31] Andrei Lapets, Frederick Jansen, Kinan Dak Albab, Rawane Issa, Lucy Qin, Mayank Varia, and Azer Bestavros. Accessible privacy-preserving web-based data analysis for assessing and addressing economic inequalities. In Zegura [60], pages 48:1–48:5. doi: 10.1145/3209811.3212701. URL http://doi.acm.org/10.1145/3209811.3212701.
[32] Yehuda Lindell. Secure multiparty computation. *Commun. ACM*, 64(1):86–96, 2021.
[33] Chang Liu, Xiao Shaun Wang, Kartik Nayak, Yan Huang, and Elaine Shi. Oblivm: A programming framework for secure computation. In *IEEE S & P*, 2015.
[34] Gerome Miklau. How Tumult Labs helped the IRS support educational accountability with differential privacy. https://www.tmlt.io/case-studies/how-tumult-labs-helped-irs-support-educational-accountability-with-differential-privacy, 2022. Accessed: 2022-08-15.
[35] Kobbi Nissim, Aaron Bembenek, Alexandra B. Wood, Mark Mar Bun, Marco Gaboardi, Urs Gasser, David O'Brien, and Salil P. Vadhan. Bridging the gap between computer science and legal approaches to privacy. *Harvard Journal of Law and Technology*, 2(31):687–780, 2018.
[36] Charalampos Papamanthou, Elaine Shi, and Roberto Tamassia. Signatures of correct computation. In *TCC*, volume 7785 of *Lecture Notes in Computer Science*, pages 222–242. Springer, 2013.
[37] Bryan Parno, Mariana Raykova, and Vinod Vaikuntanathan. How to delegate and verify in public: Verifiable computation from attribute-based encryption. In *TCC*, volume 7194 of *Lecture Notes in Computer Science*, pages 422–439. Springer, 2012.
[38] Lucy Qin, Andrei Lapets, Frederick Jansen, Peter Flockhart, Kinan Dak Albab, Ira Globus-Harris, Shannon Roberts, and Mayank Varia. From usability to secure computing and back again. In *SOUPS @ USENIX Security Symposium*. USENIX Association, 2019.
[39] Anjana Rajan, Lucy Qin, David W. Archer, Dan Boneh, Tancrède Lepoint, and Mayank Varia. Callisto: A cryptographic approach to detecting serial perpetrators of sexual misconduct. In Zegura [60], pages 49:1–49:4. doi: 10.1145/3209811.3212699. URL http://doi.acm.org/10.1145/3209811.3212699.
[40] Robert Bosch GmbH. https://carbynestack.io/, The Carbyne Stack: Cloud Native Secure Multiparty Computation. Last access: January 2022.
[41] Jennie Rogers, Elizabeth Adetoro, Johes Bater, Talia Canter, Dong Fu, Andrew Hamilton, Amro Hassan, Ashley Martinez, Erick Michalski, Vesna Mitrovic, Fred D. Rachman, Raj C. Shah, Matt Sterling, Kyra VanDoren, Theresa L. Walunas, Xiao Wang, and Abel N. Kho. Vaultdb: A real-world pilot of secure multiparty computation within a clinical research network. *CoRR*, abs/2203.00146, 2022.
[42] Ira S. Rubinstein and Woodrow Hartzog. Anonymization and risk. *Washington Law Review*, 91(703), 2016.
[43] James Scheibner, Jean Louis Raisaro, Juan Ramón Troncoso-Pastoriza, Marcello Ienca, Jacques Fellay, Effy Vayena, Jean-Pierre Hubaux, et al. Revolutionizing medical data sharing using advanced privacy-enhancing technologies: technical, legal, and ethical synthesis. *Journal of medical Internet research*, 23(2):e25120, 2021.
[44] Sepior. Advanced mpc for superior key management & protection. https://sepior.com/, 2022.
[45] Victor Shoup. OAEP reconsidered. *J. Cryptology*, 15(4):223–249, 2002.
[46] Gerald Spindler and Philipp Schmechel. Personal data and encryption in the european general data protection regulation. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, 7:163, 2016.
[47] Stephanie Strauss. A federal government privacy-preserving technology demonstration. https://medium.com/georgetown-massive-data-institute/a-federal-government-privacy-preserving-technology-demonstration-27415784fcda, 2021.
[48] Josh Swihart, Benjamin Winston, and Sean Bowe. Zcash counterfeiting vulnerability successfully remediated. https://electriccoin.co/blog/zcash-counterfeiting-vulnerability-successfully-remediated/, February 2019.
[49] UK Financial Conduct Authority. 2019 global AML and financial crime TechSprint. https://www.fca.org.uk/events/techsprints/2019-global-aml-and-financial-crime-techsprint, 2019.
[50] U.K.-U.S. prize challenges. Accelerating the adoption and development of privacy-enhancing technologies (pets). https://petsprizechallenges.com/, 2022.
[51] Unbound Security. https://www.unboundsecurity.com/, 2022.
[52] United Nations Global Working Group on Big Data. Privacy preserving techniques: Task team of the un committee of experts on big data and data science for official statistics. https://unstats.un.org/bigdata/task-teams/privacy/index.cshtml, 2022.
[53] US Department of Education. College scorecard glossary. https://collegescorecard.ed.gov/data/glossary/#fos-median-earnings, 2016. Accessed: 2022-08-15.
[54] US Department of Health & Human Services . Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html, 2022. Accessed: 2022-08-15.
[55] US Department of Health & Human Services. If a CSP stores only encrypted ePHI and does not have a decryption key, is it a HIPAA business associate? https://www.hhs.gov/hipaa/for-professionals/faq/2076/if-a-csp-stores-only-encrypted-ephi-and-does-not-have-a-decryption-key-is-it-a-hipaa-business-associate/index.html, 2016. Accessed: 2022-08-15.
[56] Ron Wyden. Student right to know before you go act of 2019. https://www.congress.gov/bill/116th-congress/senate-bill/681/all-info, 2019.
[57] Jean Yang, Kuat Yessenov, and Armando Solar-Lezama. A language for automatically enforcing privacy policies. *ACM SIGPLAN Notices*, 47(1):85–96, 2012.
[58] Jean Yang, Travis Hance, Thomas H Austin, Armando Solar-Lezama, Cormac Flanagan, and Stephen Chong. End-to-end policy-agnostic security for database-backed applications. *CoRR, abs/1507.03513*, 2015.
[59] Samee Zahur and David Evans. Obliv-c: A language for extensible data-oblivious computation. *IACR Cryptol. ePrint Arch.*, page 1153, 2015.
[60] Ellen W. Zegura, editor. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018, Menlo Park and San Jose, CA, USA, June 20-22, 2018*, 2018. ACM. doi: 10.1145/3209811. URL http://doi.acm.org/10.1145/3209811.