An operational architecture for privacy-by-design in public service applications

Prashant Agrawal[1], Anubhutie Singh[2], Malavika Raghavan[2], Subodh Sharma[1] and Subhashis Banerjee[1]

[1]Department of Computer Science and Engineering, IIT Delhi
[2]Dvara Research

December 17, 2020

Abstract

Governments around the world are trying to build large data registries for effective delivery of a variety of public services. However, these efforts are often undermined due to serious concerns over privacy risks associated with collection and processing of personally identifiable information. While a rich set of special-purpose privacy-preserving techniques exist in computer science, they are unable to provide end-to-end protection in alignment with legal principles in the absence of an overarching operational architecture to ensure purpose limitation and protection against insider attacks. This either leads to weak privacy protection in large designs, or adoption of overly defensive strategies to protect privacy by compromising on utility.

In this paper, we present an operational architecture for privacy-by-design based on independent regulatory oversight stipulated by most data protection regimes, regulated access control, purpose limitation and data minimisation. We briefly discuss the feasibility of implementing our architecture based on existing techniques. We also present some sample case studies of privacy-preserving design sketches of challenging public service applications.

1 Introduction

A welfare state may have legitimate interests in building large data registries with personally identifiable information (PII) for efficiency of service delivery. A state may also legitimately need to put its residents under purpose-specific surveillance. In fact, several commentators have alluded to the possibility of pervasive under-the-skin surveillance in a post-COVID world [1]. However, mandatory recordings of PII require enacting reasonable and fair laws to ensure that the processing of PII is proportionate to the stated objective, and safeguard the basic operative principles of privacy and fairness. Citizens' basic rights need to be protected even when there is a legitimate state interest in digitisation with PII [2]. The need to ensure that the information collected is not used adversely against citizens to harm them takes us into one of the hard problems of modern public policy: creating rules and technologies around information privacy to help strike this critical balance for online collection of PII at national scale.

In this paper we address the problem of operationalising the broad privacy-by-design principles outlined in [3, 4], in the context of large public service databases. We present an architecture for implementing the data protection principles after the utility and proportionality of an application have been established through an appropriate regulatory analysis [5, 6, 7].

The general principles of fair and reasonable processing, purpose, collection and storage limitation, notice and consent, data quality etc. have evolved since the 1970s, both through sector specific standards in the US such as the Social Security Number Protection Act [8] and Health Insurance Portability and Accountability Act (HIPAA) [9], or through omnibus laws in general data protection standards such as the GDPR in the European Union [6] and the Draft Data Protection Bill of India [7]. However, they have largely failed to prevent both direct harms that can occur as a result of data breaches or through unauthorised access of personal data - such as identity thefts, unethical profiling and unlawful surveillance, or secondary harms that could arise due to the use of the data to adversely affect a person - such as through discrimination or exclusion, predatory

targeting for unsuitable products, loss of employment, inaccurate credit rating etc. Dictums such as personal data shall be processed in a fair and reasonable manner are non-specific, and they do not adequately define the contours of the required regulatory actions. As episodes like Cambridge Analytica [10] demonstrate, harm is often not immediately obvious, and causal links of harm are not always easy to determine. This is compounded by the fact that data collection and use are becoming ubiquitous making it hard to trace misuse; the effects of misuse of personal data may not immediately manifest, and when they do they may not be easily quantifiable in monetary terms despite causing grave distress. Hence, ex-post accountability and punitive measures are largely ineffective, and it is imperative to operationalise ex-ante preventive principles.

As a consequence of the weak protection standards, most attempts at building large public services like national identity systems [11, 12], health registries [13, 14, 15], national population and voter registries [16, 17, 18], public credit registries [19, 20], income [21] and tax registries [22] etc. have often been questioned on privacy and fairness grounds and have been difficult to operationalise. The concerns have invariably been related to the need for protective safeguards when large national data integration projects are contemplated by governments and acknowledgment of the unprecedented surveillance power that this could create. In some situations they have even had to be abandoned altogether as they were unable to deal with these risks [13, 14, 23]. In India too, the recent momentum and concerns around informational privacy guarantees has occurred in the context of the creation of new government databases and digital infrastructures for welfare delivery [24, 25].

1.1 Requirements for privacy protection

Recording transactions with PII projects an individual into a data space, and any subsequent loss of privacy can happen only through the data pathway. Hence data protection is central to privacy protection insofar as databases are concerned. The critical challenge in design of a data protection framework is that the main uses of digitisation - long term record keeping and data analysis - are seemingly contradictory to the privacy protection requirements. The legal principles around "fair information practice" attempt to reconcile these tensions, but there are four broad areas that require careful attention for effective data protection.

1.1.1 Use cases and data minimisation

First, a data protection framework is incomplete without an investigation of the nuances of digital identity, and guidelines for the various use cases of authentication, authorisation and accounting. It is also incomplete without an analysis of the extent to which personal information needs to be revealed for each use case, for example during know-your-citizen or -customer (KYC) processes. In addition, effective protection requires an understanding of the possible pathways of information leaks; of the limits of anonymisation with provable guarantees against re-identification attacks [26]; and of the various possibilities with virtual identities [27, 28].

1.1.2 Access control and purpose limitation

Second, there have to be clear-cut guidelines for defining the requirements and standards of access control, and protection against both external and insider attacks in large data establishments, technically as well as legally. In particular, insider attacks are the biggest threat to privacy in public databases [12]. These include possible unauthorised and surreptitious examination of data, transaction records, logs and audit trails by personnel with access, leading to profiling and surveillance of targeted groups and individuals, perhaps at the behest of interested and influential parties in the state machinery itself [29]. Thus, there must be guidelines on how the data may be accessed, under what authorisation and for what purpose. In addition, post data access purpose limitation - ensuring that there is no illegal use after the data crosses the access boundaries - is also crucial for privacy protection.

1.1.3 Inferential privacy and purpose limitation

Third, a data protection framework is incomplete without guidelines for safe use of AI and data analytics. Most theories for improving state efficiency in delivery of welfare and health services using personal data will have to consider improved data processing methods for targeting, epidemiology, econometrics, tax compliance, corruption control, analytics, topic discovery, etc. This, in turn, will require digitisation, surveillance and processing of large-scale personal transactional data. This requires detailed analyses of how purpose limitation of such surveillance - targeted towards improving efficiency of the state's service delivery - may be achieved

without enabling undesirable mass surveillance that may threaten civil liberty and democracy. There must also be effective guidelines to prevent discriminatory and biased data processing [30].

1.1.4 Regulatory oversight

Finally, it is well recognised in data protection frameworks [5, 6, 7] that regulatory oversight is a necessary requirement for ensuring the above.

1.2 Our contribution

While there is a rich set of tools and techniques in computer science arising out of decades of innovative privacy research, there is no overarching general framework for a privacy preserving architecture which, in particular, allows regulatory supervision and helps deal with the above issues in effective designs. In this paper we propose such an operational architecture for implementing the data protection principles. Our immediate objective here is design space exploration and not specific implementations to evaluate performance and scalability.

We illustrate the effectiveness of our proposal through design sketches of some challenging large public service applications. In particular, we illustrate through some real world case studies how some state-of-the-art designs either fail in their data protection goals, or tend to be overly defensive at the cost of utility in the absence of such an architecture.

The rest of the paper is organized as follows. Section 2 briefly reviews the basic legal principles for data protection. Section 3 reviews concepts, tools and techniques from computer science for privacy protection. Section 4 presents our operational architecture. Section 5 discusses the feasibility and Section 6 discusses some illustrative case studies of large government applications.

2 The India context and privacy and data protection concepts in law and regulation

In what follows we briefly discuss the context of digitisation and privacy in India and the basic legal principles around privacy. We situate this analysis within the context of India's evolving regulatory and technical systems. However, many of these principles are relevant for any country seeking to align legal and technical guarantees of privacy for citizens.

2.1 Public digital infrastructures and informational privacy in India

Building public digital infrastructures has received an impetus in India in recent times [31] and privacy has been an obvious concern. India has a long-standing legal discourse on privacy as a right rooted in the country's Constitution. However, informational privacy and data protection issues have gained renewed visibility due to the recent national debate around the country's Aadhaar system [24]. Aadhaar is a unique, biometric-based identity system launched in 2009, with the ambitious aim of enrolling all Indian residents, and recording their personal information, biometric fingerprints and iris scans against a unique identity number. Aadhaar was designed as a solution for preventing leakages in government welfare delivery and targeting public services through this identity system. In addition, the "India stack" was envisioned as a set of APIs that could be used - by public and private sector entities contract - to query the Aadhaar database to provide a variety of services [32]. However, as the project was unrolled across the country, its constitutionality was challenged in the courts on many grounds including the main substantive charge that it was violative of the citizens' right to privacy. Over 30 petitions challenging the system were eventually raised to the Supreme Court of India for its final determination. In the course of the matter, a more foundational question arose, i.e., whether the Indian Constitution contemplated a fundamental right of privacy? The question was referred to a separate 9-judge bench of the Indian Supreme Court to conclusively determine the answer to this question. The answer to this question is important both for law and computer science, since the response creates deep implications for the design of technical systems in India. The Supreme Court's unanimous response to this question in Justice K.S.Puttaswamy (Retd.) vs Union of India (Puttaswamy I) [2] was to hold that privacy is a fundamental right in India guaranteed by Part III (Fundamental Rights) of the Indian Constitution. Informational privacy was noted to be an important aspect of privacy for each individual, that required protection and security. In doing so, the Court recognised the interest of an individual in controlling or limiting the access to their personal

information, especially as ubiquitous data generation and collection, combined with data processing techniques, can derive information about individuals that we may not intend to disclose.

2.2 Defining informational privacy

In addition to cementing privacy as a constitutional right for Indians, the Supreme Court in Puttaswamy I [2] also played an important role in clarifying certain definitional aspects of the concept.

First, when defining privacy, the lead judgement noted that every person's reasonable expectation of privacy has both subjective and objective elements (see page 246 of Puttaswamy I), i.e.,

1. the subjective element which is to the expectation and desire of an individual to be left alone, and
2. the objective element, which refers to objective criteria and rules (flowing from constitutional values) that create the widely agreed content of "the protected zone", where a person ought to be left alone in our society.

Second, informational privacy was also recognised (see page 201 of Puttaswamy I, from a seminal work which set out a typology of privacy) to be:

> "... an interest in preventing information about the self from being disseminated and controlling the extent of access to information."

It would be the role of a future Indian data protection law to create some objective standards for informational privacy to give all actors in society an understanding of the "ground rules" for accessing an individuals' personal information. These principles are already fairly well-developed through several decades of international experience. India is one of the few remaining countries in the world that is yet to adopt a comprehensive data protection framework. This section provides a brief overview of some of these established concepts.

2.3 Data protection principles

One of the early and most influential global frameworks on privacy protection are the 1980 OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [4]. These were formulated as a response to the advancements in technology that enabled faster processing of large amounts of data as well as their transmission across different countries. These Guidelines were updated in 2013, reflecting the multilateral consensus of the changes in the use and processing of personal data in that 30 year period. Therefore, it is a good starting point for the fundamental principles of privacy and data protection.

The key principles of the OECD Privacy Framework 2013 are:

Collection Limitation: Personal data should be collected in a fair and lawful manner and there should be limits to its collection.

Use Limitation: Collected personal data be used or disclosed for any purposes other than those stated. If personal data must be used for purposes other that those stated, it should with the consent of the data subject or with the authority of the law.

Purpose Specification: The purpose for collection of personal data should be stated no later than the point of collection. All subsequent uses of such data must be limited to the stated purposes.

Data Quality: Collected personal data should be relevant for the stated purposes and its accuracy for such a purpose must be maintained.

Security Safeguards: Reasonable safeguards must be adopted by the data controller to protect it from risks such as unauthorised access, destruction, use, modification or disclosure of the data.

Accountability: Any entity processing personal data must be responsible and held accountable for giving effect to the principles of data protection and privacy.

Openness: Any entity processing personal data must be transparent about the developments and practices with respect to the personal data collected.

Individual Participation: Individuals should have the rights to confirm from the data controller whether they have any personal data relating to them and be able to obtain the same within a reasonable time, at a reasonable charge and in a reasonable manner. If these requests are denied, individuals must be given the reasons for such denial and have the right to challenge such denials. Individuals must also retain the right to be able to challenge personal data relating to them and able to erase, rectify, complete or amended.

These principles, and many international instruments and national laws that draw from them, set some of the basic ground rules around the need for clear and legitimate purposes to be identified prior to accessing personal information. They also stress on the need for accountable data practices including strict access controls. Many of these principles are reflected to varying degrees in India's Personal Data Protection Bill in 2019 [7] which was introduced in the Lower House of the Indian Parliament in December 2019. The Bill is currently under consideration by a Joint Select Committee of Parliamentarians following which it will enter Parliament for final passage.

The OECD Privacy Framework 2013 [4] in Article 19(g) also recognised the need for the promotion of technical measures to protect privacy in practice. There is also a growing recognition that if technical systems are not built with an appreciation of data protection and privacy principles, they can create deficits of trust and other dysfunctions. These are particularly problematic in government-led infrastructures.

2.4 The failure of privacy self-management and the need for accountability-based data protection

The need for data processing entities to adhere to objective and enforceable standards of data protection is heightened because of vulnerability of the individuals whose data they process. Although research shows that individuals value their privacy and seek to control how information about them is shared, cognitive limitations operate at the level of the individuals' decision-making about their personal data [5]. This "Privacy Paradox" signals the behavioural biases and information asymmetries that operate on people making decisions about sharing their personal information. Especially in contexts where awareness that personal data is even being collected in digital interactions is low, such as with first-time users of digital services in India, it is often unfair and meaningless to delegate the self-management of privacy to users entirely through the ineffective mechanism of "consent". The inadequacy of consent alone as a privacy protection instrument has been well established, especially given that failing to consent to data collection could result in a denial of the service being sought by the user [5].

In the context of these findings, it is crucial that digital ecosystems be designed in a manner that protects the privacy of individuals, does not erode their trust in the data collecting institution and does not make them vulnerable to different natures of harm. Therefore, mere dependence on compliance with legal frameworks by data controllers is not sufficient. Technical guarantees that the collected data will only be used for the stated purposes and in furtherance of data protection principles must become a reality, if these legal guarantees are to be meaningful. The need for early alignment of legal and technical design principles of data systems, such as access controls, purpose limitation and clear liability frameworks under appropriate regulatory jurisdictions are essential to create secure and trustworthy public data infrastructures [5, 6, 7].

3 Privacy concepts in computer science

Before we present our architectural framework, we briefly review some privacy preserving tools from computer science.

3.1 Encryption, signatures and cryptographic hashes

3.1.1 Encryption

Cryptographic encryption [33], for protecting data either in storage or transit, have often been advocated for privacy protection. The following types are of particular importance:

Symmetric encryption Symmetric encryption allows two parties to encrypt and decrypt messages using a shared secret key. Diffie-Hellman key exchange protocol [34] is commonly used by the parties to jointly establish a shared key over an insecure channel.

Asymmetric encryption Asymmetric or public key encryption [34] allows two parties to communicate without the need to exchange any keys beforehand. Each party holds a pair of *public* and *private* keys such that messages encrypted using the receiver's public key cannot be decrypted without the knowledge of the corresponding private key.

ID-based encryption ID-based encryption [35] allows the sender to encrypt the message against a textual ID instead of a public key. A trusted third party provisions decryption keys corresponding to the IDs of potential receivers after authenticating them through an out-of-band mechanism. ID-based encryption considerably simplifies the public key infrastructure: a sender can encrypt messages using the semantic identifier of the intended recipient without explicitly knowing the public keys of the particular receivers.

Encryption with strong keys is a powerful method for privacy protection provided there are no unauthorised accesses to the keys. Insider attacks, however, pose serious risks if the keys also reside with the same authority. Even when the keys are stored securely, they have to be brought into the memory for decryption during run-time, and can be leaked by a compromised privileged software, for example an operating system or a hypervisor.

3.1.2 Signatures

Digital signature A digital signature [34] $\sigma_{pk}(m)$ on a message m allows a verifier to verify using the public key pk that m was indeed signed with the corresponding the private key. Any alteration of m invalidates the signature. Signatures also provide non-repudiation.

Blind signatures Blind signatures [36] are useful to obtain a signature on a message without exposing the contents of the message to the signer. A signature $\sigma'_{pk}(b(m))$ by a signer holding public key pk allows the signer to sign a *blinded message* $b(m)$ that does not reveal anything about m. The author of the message can now use the $\sigma'_{pk}(b(m))$ to create an unblinded digital signature $\sigma_{pk}(m)$.

3.1.3 Cryptographic hash function (CHF)

CHFs are functions that are *a*) 'one-way', i.e., given hash value h, it is difficult to find an x such that $h = hash(x)$, and *b*) 'collision-resistant', i.e., finding any x_1 and x_2 such that $hash(x_1) = hash(x_2)$ is difficult. CHFs form the basis of many privacy preserving cryptographic primitives.

3.2 Data minimisation

There are several techniques from computer science that are particularly useful for data minimisation - at different levels of collection, authentication, KYC, storage and dissemination. Some of these are:

3.2.1 Zero knowledge proofs (ZKPs) and selective disclosures

ZKPs [37] are proofs that allow a party to prove to another that a statement is true, *without* leaking any information other than the statement itself. Of particular relevance are ZKPs of knowledge [38], which convince a verifier that the prover knows a secret without revealing it. ZKPs also enable *selective disclosure* [39], i.e., individuals can prove only purpose-specific attributes about their identity without revealing additional details; for example, that one is of legal drinking age without revealing the age itself.

3.2.2 Anonymity and anonymous credentials

"Anonymity refers to the state of being not identifiable within a set of individuals, the anonymity set" [40]. In the context of individuals making transactions with an organisation, the following notions of anonymity can be defined:

Unlinkable anonymity Transactions provide unlinkable anonymity (or simply *unlinkability*) if *a*) they do not reveal the true identities of the individuals to organisations, and *b*) organisations cannot identify how different transactions map to the individuals.

Linkable anonymity Transactions provide linkable anonymity if an organisation can identify whether or not two of its transactions involve the same individual, but individuals' true identities remain hidden. Linkable anonymity is useful because it allows individuals to maintain their privacy while allowing the organisation to aggregate multiple transactions from the same individual. Linkable anonymity is typically achieved by making individuals use *pseudonyms*.

Anonymous credentials Authenticating individuals online may require them to provide credentials from a credential-granting organisation A to a credential-verifying organisation B. Privacy protection using anonymous credentials [41, 28, 27] can ensure that transactions with A are unlinkable to transactions with B. Anonymous credentials allow an individual to obtain a credential from an organisation A against their pseudonym with A and transform it to an identical credential against their pseudonym with organisation B. An identity authority provisions a master identity to each individual from which all pseudonyms belonging to an individual, also known as *virtual identities*, are cryptographically derived. Anonymous credentials are typically implemented by obtaining blind signatures (see Section 3.1.2) from the issuer and using ZKPs of knowledge (see Section 3.2.1) of these signatures to authenticate with the verifier. The credential mechanism guarantees:

- *Unlinkable anonymity across organisations.* This property ensures that A cannot track the uses of the issued credentials and B cannot obtain the individual's information shared only with A even when A and B collude.
- *Unforgeability.* A credential against an individual's pseudonym cannot be generated without obtaining an identical credential against another pseudonym belonging to the same individual.
- *Linkable anonymity within an organisation.* Depending on the use case requirements, individuals may or may not use more than one pseudonym per organisation. In the latter case the transactions within an organisation also become unlinkable.

If an organisation A requires to link multiple transactions from the same individual, it can indicate this requirement to the identity authority that checks if pseudonyms used by individuals with A are unique. If A does not require linking, the identity authority merely checks if the pseudonyms are correctly derived from the individual's master identity. If the checks pass, an anonymous credential certifying this fact is issued by the identity authority. All checks by the identity authority preserve individuals' anonymity.

Accountable anonymous credentials Anonymity comes with a price in terms of accountability: individuals can misuse their credentials if they can never be identified and held responsible for their actions. Trusted third parties can revoke the anonymity of misbehaving users to initiate punitive measures against them [42, 43, 44]. *One-time credentials* and *k-times anonymous authentication schemes* [45, 46, 47] also prevent overspending of limited-use credentials by revoking individuals' anonymity if they overspend. Blacklisting misbehaving users for future access without revoking their anonymity is also feasible [48].

Linkability by a trusted authority Linking across organisations may also be required for legitimate purposes, for example for legitimate data mining. Also see examples in Section 5. Such linkability also seems to be an inevitable requirement to deter sharing of anonymous credentials among individuals [49]. Linkability by a trusted authority can be trivially achieved by individuals attaching a randomised encryption of a unique identifier against the trusted authority's public key for transactions requiring cross-linking. Of course, appropriate mechanisms must exist to ensure that the trusted authority does not violate the legitimate purpose of linking.

Note that the anonymity of credentials is preserved only under the assumption that individuals interact with organisations through anonymous channels (e.g., in [42]). In particular, neither the communication network nor the data that individuals share with organisations should be usable to link their transactions (see Section 3.2.3 and 3.2.4).

3.2.3 Anonymous networks

Anonymous networks, originally conceptualised as *mix networks* by Chaum [50], are routing protocols that make messages hard-to-trace. Mix networks consist of a series of proxy servers where each of them receives messages from multiple senders, shuffles them, and sends to the next proxy server. An onion-like encryption

scheme allows each proxy server to only see an encrypted copy of the message (and the next hop in plaintext), thus providing untraceability to the sender even if only one proxy server honestly shuffles its incoming messages.

3.2.4 Database anonymisation

Anonymisation is the process of transforming a database such that individuals' data cannot be traced back to them. However, research in de-anonymisation has shown that anonymisation does not work in practice, as small number of data points about individuals coming from various sources, none uniquely identifying, can completely identify them when combined together [26]. This is backed by theoretical results [51, 52] which show that for high-dimensional data, anonymisation is not possible unless the amount of noise introduced is so large that it renders the database useless. There are several reports in literature of de-anonymisation attacks on anonymised social-network data [53, 54], location data [55], writing style [56], web browsing data [57], etc.

3.2.5 Interactive database anonymisation

In this setting, analysts interact with a remote server only through a restricted set of queries and the server responds with possibly noisy answers to them. Dinur and Nissim [58] show that given a database with n rows, an adversary having no prior knowledge could make $O(n\ polylog(n))$ random subset-sum queries to *reconstruct* almost the entire database, unless the server perturbs its answers too much (by at least $O(\sqrt{n})$). This means that preventing inference attacks is impossible if the adversary is allowed to make arbitrary (small) number of queries. Determining whether a given set of queries preserves privacy against such attacks is in general intractable (NP-hard) [59].

3.3 Inferential and differential privacy

3.3.1 Inferential privacy

Inferential privacy [60, 61] is the notion that no information about an individual should be learnable with access to a database that could not be learnt without any such access. In a series of important results [58, 62, 63], it was established that such an absolute privacy goal is impossible to achieve if the adversary has access to arbitrary auxiliary information. More importantly, it was observed that individuals' inferential privacy is violated even when they do not participate in the database, because information about them could be leaked by correlated information of other participating individuals.

3.3.2 Differential privacy

In the wake of the above results, the notion of *differential privacy* was developed [63] to allow analysts extract meaningful distributional information from statistical databases while minimising the *additional* privacy risk that each individual incurs by participating in the database. Note that differential privacy is a considerably weaker notion than inferential privacy as reconstruction attacks described in Section 3.2.5 or other correlation attacks can infer a lot of non-identifying information from differentially private databases too.

Mechanisms for differential privacy add noise to the answers depending on the *sensitivity* of the query. In this sense, there is an inherent utility versus privacy tradeoff. Differentially private mechanisms possess composability properties. Thus, privacy degrades gracefully when multiple queries are made to differentially private databases. However, this alone may not protect against an attacker making an arbitrary number of queries. For example, the reconstruction attacks mentioned in Section 3.2.5 prevent many differentially private algorithms from answering a linear (in the number of rows) number of queries [64]. For specific types of queries though, e.g., predicate queries, sophisticated noise-addition techniques [65] can be used to maintain differential privacy while allowing for an exponential number of queries [64, 66].

3.3.3 Group and societal privacy

Differentially private mechanisms also degrade gracefully with respect to group privacy as the group size increases. These guarantees may not be enough for policymakers who must protect the profile of specific communities constituting a sizable proportion of the population. The ability of an adversary to manipulate and influence a community even without explicitly identifying its members is deeply problematic, as demonstrated

by episodes like Cambridge Analytica [10]. Therefore, the goal of modern private data analysis should not be limited to protecting only individual privacy, but also extend to protecting sensitive aggregate information.

3.3.4 A note on non-statistical databases

Due to the inherently noisy nature of differentially private mechanisms, they are not suitable for any non-statistical uses, e.g., financial transactions, electronic health records, and password management. Privacy mechanisms for such use-cases must prevent misuse of data for malicious purposes such as illegal surveillance or manipulation, without hampering the legitimate workflows.

The difficulties with differential privacy, and the impossibility of protection against inferential privacy violations, suggest that privacy protection demands that there should be no illegal access or processing in the first place.

3.4 Purpose limitation

3.4.1 Program analysis techniques

These check whether a given code-base uses personal data in accordance with a given privacy policy [67, 68, 69]. Privacy policies are expressed in known formal languages [70, 71]. A compiler verifies, using standard information flow analysis [72] and model-checking techniques [73], if a candidate program satisfies the intended privacy policy. In order to enforce various information flow constraints these techniques rely on manual and often tedious tagging of variables, functions and users with security classes and verify if information does not flow from items with high security classes to items with low security classes.

3.4.2 Purpose-based access control

These techniques define purpose hierarchies and specify purpose-based access-control mechanisms [74, 75, 76]. However, they typically identify purpose with the role of the data requester and therefore offer weak protection from individuals claiming wrong purposes for their queries.

3.4.3 Formalisation of purpose

Jafari et al. [77] formalise purpose as a relationship between actions in an action graph. Hayati et al. [69] express purpose as a security class (termed by them as a "principal") and verify that data collected for a given purpose does not flow to functions tagged with a different purpose. Tschantz et al. [78] state that purpose violation happens if an action is *redundant* in a plan that maximises the expected satisfaction of the allowed purpose. However, enforcement of these models still relies on fine-grained tagging of code blocks, making them tedious, and either a compiler-based verification or post-facto auditing, making them susceptible to insider attacks that bypass the checks.

3.5 Secure remote execution

Secure remote execution refers to the set of techniques wherein a client can outsource a computation to a remote party such that the remote party does not learn anything about the client's inputs or intermediate results.

3.5.1 Homomorphic encryption

Homomorphic encryption (HE) schemes compute in the ciphertext space of encrypted data by relying on the additive or multiplicative homomorphism of the underlying encryption scheme [79, 80, 81]. Designing an encryption scheme that is both - which is required for universality - is challenging. Gentry [82] gave the first theoretical fully homomorphic encryption (FHE) scheme. Even though state-of-the-art FHE schemes and implementations have considerably improved upon Gentry's original scheme, the performance of these schemes is still far from any practical deployment [83]. Functional encryption (FE) [84] schemes have similar objectives, with the crucial difference that FE schemes let the remote party learn the output of the computation, whereas FHE schemes compute encrypted output, which is decrypted by the client.

3.5.2 Secure multiparty computation

Secure multiparty computation (SMC) - originally pioneered by Yao through his garbled circuits technique [85] - allows multiple parties to compute a function of their private inputs such that no party learns about others' private inputs, other than what the function's output reveals. SMC requires clients to express the function to be computed as an encrypted circuit and send it to the server alongwith encrypted inputs; the server needs to evaluate the circuit by performing repeated decryptions of the encrypted gates. As a result, SMC poses many challenges in its widespread adoption - ranging from the inefficiencies introduced by the circuit model itself to the decryption overhead for each gate evaluation, even as optimisations over the last two decades have considerably improved the performance and usability of SMC [86].

However, HE, FE and SMC based schemes involve significant application re-engineering and may offer reduced functionality in practice.

3.6 Hardware-based security

In recent times, secure remote execution is increasingly being realised not through advances in cryptography but through advances in hardware-based security. This approach commoditises privacy-preserving computation, albeit at the expense of a weakened trust model, i.e., the increased trust on the hardware manufacturer.

3.6.1 Intel SGX

Intel Software Guard Extensions (SGX) [87] implements access control in the CPU to provide confidentiality and integrity to the executing program. At the heart of the SGX architecture lies the notion of an isolated execution environment, called an *enclave*. An enclave resides in the memory space of an untrusted application process but access to the enclave memory and leakage from it are protected by the hardware. The following are the main properties of SGX:

Confidentiality Information about an enclave execution can not leak outside the enclave memory except through explicit exit points.

Integrity Information can not leak into the enclave to tamper with its execution except through explicit entry points.

Remote attestation For an enclave's execution to be trusted by a remote party, it needs to be convinced that *a*) the contents of the enclave memory at initialisation are as per its expectations, and *b*) that confidentiality and integrity guarantees will be enforced by the hardware throughout its execution. For this the hardware computes a *measurement*, essentially a hash of the contents of the enclave memory and possibly additional user data, signs it and sends it over to the remote party [88]. The remote party verifies the signature and matches the enclave measurement with the measurement of a golden enclave it considers secure. If these checks pass, the remote party trusts the enclave and sends sensitive inputs to it.

Secure provisioning of keys and data SGX enclaves have secure access to hardware random number generators. Therefore, they can generate a Diffie-Hellman public/private key pair and keep the private key secured within enclave memory. Additionally, the generated public key can be included as part of additional user data in the hardware measurement sent to a remote verifier during remote attestation. These properties allow the remote verifier to establish a secure TLS communication channel with the enclave over which any decryption keys or sensitive data can be sent. The receiving enclave can also seal the secrets once obtained for long-term use such that it can access them even across reboots, but other programs or enclaves cannot.

3.6.2 Other hardware security mechanisms

SGX has been preceded by the Trusted Platform Module (TPM) [89]. TPM defines a hardware-based root of trust, which measures and attests the entire software stack, including the BIOS, the OS and the applications, resulting in a huge trusted computing base (TCB) as compared to SGX whose TCB includes only the enclave code. ARM Trustzone [90] partitions the system into a secure and an insecure world and controls interactions between the two. In this way, Trustzone provides a single enclave, whereas SGX supports multiple enclaves.

Trustzone has penetrated the mobile world through ARM-based Android devices, whereas SGX is available for laptops, desktops and servers.

SGX is known to be susceptible to serious side-channel attacks [91, 92, 93, 94]. Sanctum [95] has been proposed as a simpler alternative that provides provable protection against memory access-pattern based software side-channel attacks. For a detailed review on hardware-based security, we refer the reader to [86].

3.7 Secure databases

Stateful secure remote execution requires a secure database and mechanisms that protect clients' privacy when they perform queries on them.

3.7.1 Querying encrypted databases

The aim of these schemes is to let clients host their data encrypted in an untrusted server and still be able to execute queries on it with minimal privacy loss and maximal query expressiveness. One approach for enabling this is searchable encryption schemes, i.e., encryption schemes that allow searching over ciphertexts [96, 97]. Another approach is to add searchable indexes along with encrypted data, or to use special property-preserving encryptions to help with searching [98, 99, 100, 101]. However, both approaches are susceptible to inference attacks [102, 103, 104, 105] (cf. Sections 3.2.4, 3.2.5 and 3.3.1). Oblivious RAM [106, 107] is a useful primitive that provides read/write access to encrypted memory while hiding all access patterns, but these schemes require polylogarithmic number of rounds (in the size of the database) per read/write request.

EnclaveDB [108] has been recently proposed as a solution based on Intel SGX. It hosts the entire database within secure enclave memory, with a secure checkpoint-based logging and recovery mechanism for durability, thus providing complete confidentiality and integrity from the untrusted server without any loss in query expressiveness.

3.7.2 Private information retrieval

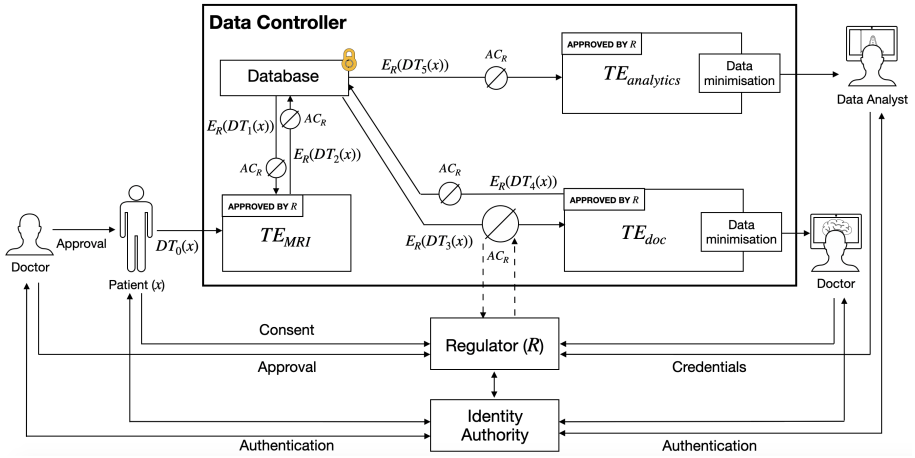
Private information retrieval (PIR) is concerned with hiding which database rows a given user query touches - thus protecting user intent - rather than encrypting the database itself. Kushilevitz and Ostrovsky [109] demonstrated a PIR scheme with communication complexity $O(n^{\epsilon})$, for any $\epsilon > 0$, using the hardness of the quadratic residuosity problem. Since then, the field has grown considerably and modern PIR schemes boast of $O(1)$ communication complexity [110]. Symmetric PIR (also known as oblivious transfer), i.e., the set of schemes where additionally users cannot learn anything beyond the row they requested, is also an active area of research.

4 Operationalisation using trusted executables and regulatory architecture

As is evident from the discussion in the previous section, none of the techniques by themselves are adequate for privacy protection. In particular, none are effective against determined insider attacks without regulatory oversight. Hence we need an overarching architectural framework based on regulatory control over data minimisation, authorisation, access control and purpose limitation. In addition, since the privacy and fairness impacts of modern AI techniques [30] are impossible to determine automatically, the regulatory scrutiny of data processing programs must have a best effort manual component. Once approved, the architecture must prevent any alteration or purpose extension without regulatory checks.

In what follows we present an operational architecture for privacy-by-design. We assume that all databases and the associated computing environments are under physical control of the data controllers, and the online regulator has no direct physical access to it. We also assume that the data controllers and the regulators do not collude.

We illustrate our conceptual design through an example of a privacy-preserving electronic health record (EHR) system. EHRs can improve quality of healthcare significantly by providing improved access to patient records to doctors, epidemiologists and policymakers. However, the privacy concerns with them are many, ranging from the social and psychological harms caused by unwanted exposure of individuals' sensitive medical information, to direct and indirect economic harms caused by the linkage of their medical data with data



$DT_0(x)$	$VirtualHospitalID(x)$
$DT_1(x)$	$(Age(x), Sex(x), MedHistory(x))$
$DT_2(x)$	$(MRIImage(x), RadiologyReport(x))$
$DT_3(x)$	$(Age(x), Sex(x), MedHistory(x), MRIImage(x), RadiologyReport(x))$
$DT_4(x)$	$(Diagnosis(x), DoctorReport(x))$
$DT_5(x)$	$(AgeDecile(x), Sex(x), Diagnosis(x))$

Figure 1: An illustration of the architecture of trusted executables using an example involving an EHR database, a patient, an MRI imaging station, a doctor and a data analysis station. TEs marked "**APPROVED BY** R" are pre-audited and pre-approved by the regulator R. $E_R(\cdot)$ represents a regulator-controlled encryption function and AC_R represents online access control by regulator R. $DT_i(x)$ represent various data types parametrised by the patient x (as explained in the right-hand side table). In particular, $VirtualHospitalID(x)$ represents the hospital-specific virtual identity of the patient. The regulator checks the consents, approvals and other static rules regarding data transfer at each stage of online access control.

presented to their employers, insurance companies or social security agencies. Building effective EHRs while minimising privacy risks is a long standing design challenge.

4.1 Trusted executables

We propose *trusted executables* (TE) as the fundamental building blocks for privacy-by-design. We introduce them in the abstract, and discuss some possibilities for their actual realisation in Section 5. TEs are data-processing programs, with explicit input and output channels, that are designed by the data controllers but are examined, audited, and approved by appropriate regulators. TEs execute in controlled environments on predetermined data types with prior privacy risk assessment, under online regulatory access control. The environment ensures that only approved TEs can operate on data items. In particular, all data accesses from the databases, and all data/digest outputs for human consumption, can only happen through the TEs. We prescribe the following main properties of the TEs:

1. *Runtime environment*: TEs are approved by regulators. They execute in the physical infrastructure of the data controllers but cannot be modified by them.

2. *Authentication*: A regulator can authenticate related TEs during runtime, and verify that indeed the approved versions are running.

3. *Integrity*: There is no way for a malicious human or machine agent, or even for the data controller, to tamper with the execution of a TE other than by sending data through the TE's explicit input channels.

4. *Confidentiality*: There is no way for any entity to learn anything about the execution of a TE other than by reading data written at the TE's explicit output channels. All data accesses and output can only happen through TEs.

Besides, all TEs should be publicly available for scrutiny. The above properties allow a regulator to ensure that a TE is untampered and will conform to the limited purpose identified at the approval stage.

As depicted in Figure 1, a data agent - for example, a hospital - interacts with databases or users only through pre-approved TEs, and direct accesses are prevented. All data stores and communication messages are encrypted using a regulator-controlled encryption scheme to prevent any information leakage in transit or storage. The data can be decrypted only inside the TEs under regulated access control. The regulator provisions decryption keys securely to the TE to enable decryption after access is granted. The regulator allows or denies

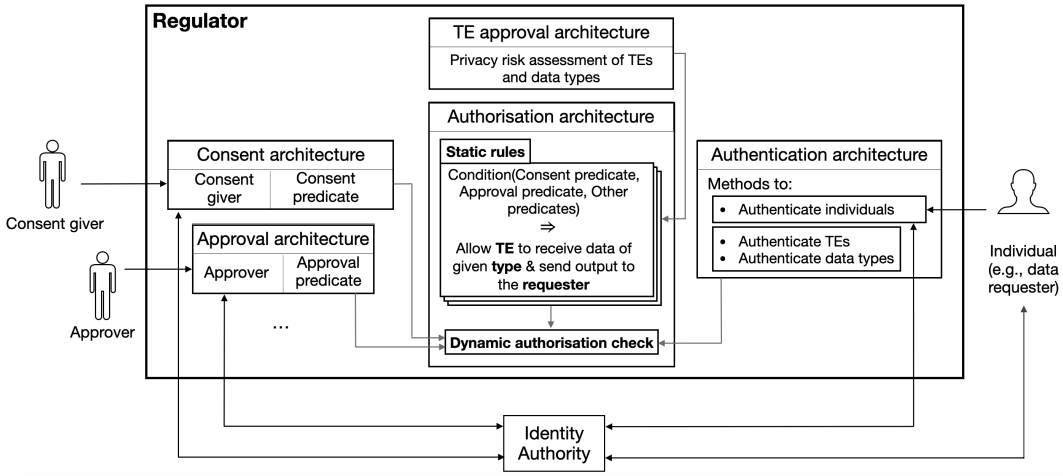


Figure 2: The regulatory architecture.

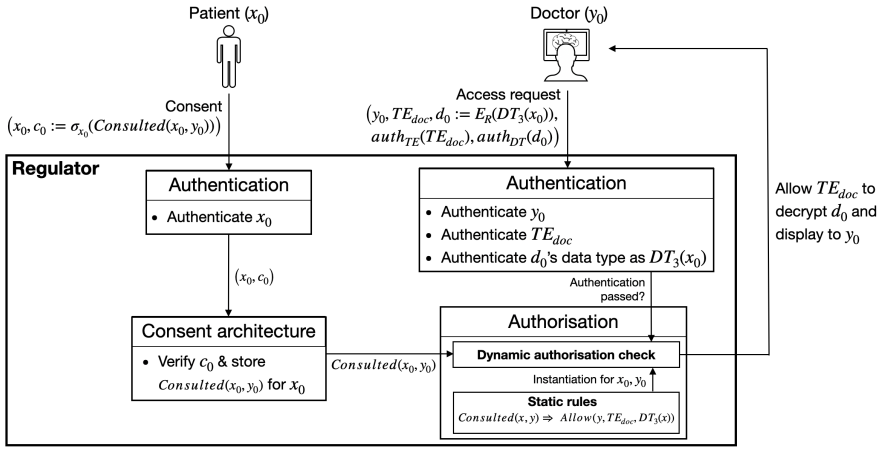


Figure 3: An example control flow diagram depicting the regulatory architecture. x_0 and y_0 represent virtual identities of the patient and the doctor, respectively. $\sigma_{x_0}(\cdot)$ represents digital signature by patient x_0. $auth_{TE}(\cdot)$ represents authentication information of the TE and $auth_{DT}(\cdot)$ represents authentication information of the supplied data's type. Individuals are authenticated by verifying their virtual identities.

access, online, based on the authentication of the TE and the incoming data type, consent and approval checks as required, and the credential authentication of any human consumers of output data (e.g., the doctor(s) and data analysts). All sink TEs - i.e., those that output data directly for consumption by a human agent - are pre-audited to employ appropriate data minimisation before sending data to their output channels. Note that extending the TE architecture to the doctors' terminals and the imaging stations ensures that the data never crosses the regulated boundary and thus enables purpose limitation.

In the above example an independent identity authority issues credentials and engages in a three-way communication to authenticate individuals who present their virtual identities to the regulator. An individual can use a master *Health id* to generate hospital-specific or doctor-specific unlinkable anonymous credentials. Only a *health authority* may be allowed to link identities across hospitals and doctors in a purpose-limited way under regulated access control.

4.2 Regulatory architecture

We depict the regulatory architecture in Figure 2. The first obligation of the regulator is to audit and approve the TEs designed by the data controllers. During this process, the regulator must assess the legality of the data access and processing requirements of each TE, along with the privacy risk assessment of its input and

output data types. In case a TE is an AI based data analytics program, it is also an obligation of the regulator to assess its fairness and the potential risks of discrimination [30]. Before approving a TE, the regulator also needs to verify that the TE invokes a callback to the regulator's online interface before accessing a data item and supplies appropriate authentication information, and that it employs appropriate encryption and data minimisation mechanisms at its output channels. Finally, the regulator needs to put in place a mechanism to be able to authenticate the TE in the data controller's runtime environment.

The second obligation of the regulator is to play an online role in authorising data accesses by the TEs. The authorisation architecture has both a static and a dynamic component. The static authorisation rules typically capture the relatively stable regulatory requirements, and the dynamic component typically captures the fast-changing online context, mainly due to consents and approvals. Specifically, each static authorisation rule takes the form of a set of pre-conditions necessary to grant access to a TE the data of a given type; and, in case of sink TEs, to output it to a requester. The design of these rules is governed by regulatory requirements and the privacy risk assessment of TEs and data types. The rules are typically parametric in nature, allowing specification of constraints that provide access to a requester only if the requester can demonstrate some specific relationship with the data individual (e.g., to express that only a doctor consulted by a patient can access her data).

The pre-conditions of the authorisation rules may be based on consent of data individuals, approvals by authorities or even other dynamic constraints (e.g., time-bound permissions). The consent architecture must be responsible for verifying signatures on standardised consent APIs from consent givers and recording them as logical consent predicates. The regulator, when designing its authorisation rules, may use a simple consent - for example, that a patient has wilfully consulted a doctor - to invoke a set of rules to protect the individual's privacy under a legal framework, rather than requiring individuals to self-manage their privacy.

Similar to the consent architecture, the approval architecture for data access must record standardised approvals from authorities as logical approval predicates. An approval from an authority may also be provided to an individual instead of directly to the regulator, as a blind signature against a virtual identity of the individual known to the approver, which should be transformed by the individual to a signature against the virtual identity known to the data controller and the regulator. This, for example, may enable a patient to present a self generated virtual identity to a doctor or a hospital instead of her universal *Health id*.

The regulator also requires an authentication architecture. First, it needs to authenticate individuals, i.e., consent givers, approvers and data requesters, by engaging in a three-way communication with an identity authority which may be external to both the data controller and the regulator. Second, it needs to authenticate TEs in order to be able identify the access requests as originating from one of the pre-approved TEs. Third, it needs to authenticate data types, i.e., identify the underlying type of the TE's encrypted input data.

The consent/approval predicates and the authentication information flow to the dynamic authorisation module, which can instantiate the static authorisation rules with the obtained contextual information to determine, in an online fashion, if access should be allowed to the requesting TE. If yes, then it must also provision decryption keys to the TE securely such that only the TE can decrypt. The keys can be securely provisioned to the TE because of the authentication, integrity and confidentiality properties, and by the fact that approved TEs must never output the obtained decryption keys.

An example control-flow diagram depicting the regulatory access control in a scenario where a doctor is trying to access the data of a patient who consulted them is shown in Figure 3.

5 Discussion on feasibility

Several existing techniques can be useful for the proposed architecture, though some techniques may need strengthening.

Trusted executables can be implemented most directly on top of trusted hardware primitives such as Intel SGX or ARM Trustzone where authentication of TEs is carried out by remote attestation. **Secure provisioning of keys and data** to TEs can be done in case of Intel SGX as per Section 3.6.1. However, since SGX includes only the CPU in its TCB, it presents challenges in porting AI applications that run on GPUs for efficiency. Graviton [111] has been recently proposed as an efficient hardware architecture for trusted execution environments on GPUs.

In our architecture, TEs fetch or update information from **encrypted databases**. This may be implemented using special indexing data structures, or may involve search over encrypted data, where the TEs act as clients and the database storage acts as the server. Accordingly, techniques from Section 3.7 can be used. Since the

TEs never output data to agents unless deemed legitimate by the regulator, the inferential attacks identified with these schemes in Section 3.7 have minimal impact. For added security, EnclaveDB [108], which keeps the entire database in secure enclave memory, can be used. EnclaveDB has been evaluated on standard database benchmarks TPC-C [112] and TATP [113] with promising results.

For **authentication of data types** messages may be encrypted using an ID-based encryption scheme, where the concrete runtime type of the message acts as the textual ID and the regulator acts as the trusted third party (see Section 3.1.1). The receiver TE can send the expected plaintext type to the regulator as part of its access request. The regulator should provision the decryption key for the ID representing the requested type only if the receiver TE is authorised to receive it as per the dynamic authorisation check. Note that authentication of the received data type is implicit here, as a TE sending a different data type in its access request can still not decrypt the incoming data.

Data minimisation for consents and approvals based on **virtual identities** is well-established from Chaum's original works [28, 41]. Individuals should use their purpose-specific virtual identities with organisations, as opposed to a unique master identity. To prevent cross-linking of identities, **anonymous credentials** may be used. In some cases, individuals' different **virtual identities may need to be linked** by a central authority to facilitate data analytics or inter-organisation transactions. This should be done under strict regulatory access control and purpose limitation.

Modern type systems can conveniently express the complex parametric constraints in the rules in the **authorisation architecture**. Efficient type-checkers and logic engines exist that could perform the dynamic authorisation checks.

Approval of TEs needs to be largely manual as the regulator needs to evaluate the legitimacy and privacy risks associated with the proposed data collection and processing activity. However, techniques from program analysis may help with specific algorithmic tasks, such as checking if the submitted programs adhere to the structural requirement of encrypting data items with the right type at their outgoing channels.

We require the **regulatory boundary to be extended** even to agent machines, which must also run TEs so that data they obtain is not repurposed for something else. However, when a TE at an authorised agent's machine outputs data, it could be intercepted by malicious programs on the agent's machine leading to purpose violation. Solutions from the DRM literature may be used to prevent this. In particular, approaches that directly encrypt data for display devices may be useful [114]. We note that this still does not protect the receiving agent from using more sophisticated mechanisms to copy data (e.g., by recording the display using an external device). However, violations of this kind are largely manual in nature and ill-suited for large-scale automated attacks.

Finally, we need **internal processes at the regulatory authority** itself to ensure that its actual operational code protects the various decryption keys and provides access to TEs as per the approved policies. To this end, the regulator code may itself be put under a TE and authenticated by the regulatory authority using remote attestation. Once authenticated, a master secret key may be provisioned to it using which the rest of the cryptosystem may bootstrap.

6 Additional case studies

In this section, we present two additional case studies to showcase the applicability of our architecture in diverse real-world scenarios.

6.1 Direct Benefit Transfer

Direct Benefit Transfer (DBT) [115] is a Government of India scheme to transfer subsidies to citizens' bank accounts under various welfare schemes. Its primary objective is to bring transparency and reduce leakages in public fund disbursal. The scheme design is based on India's online national digital identity system Aadhaar [24]. All DBT recipients have their Aadhaar IDs linked to their bank accounts to receive benefits.

Figure 4 shows a simplified schematic of the scheme that exists today [116]. A ministry official initiates payment by generating a payment file detailing the Aadhaar IDs of the DBT recipients, the welfare schemes under which payments are being made and the amounts to be transferred. The payment file is then signed and sent to a centralised platform called the Public Financial Management System (PFMS). PFMS hosts the details of various DBT schemes and is thus able to initiate an inter-bank fund transfer from the bank account of the sponsoring scheme to the bank account of the beneficiary, via the centralised payments facilitator NPCI

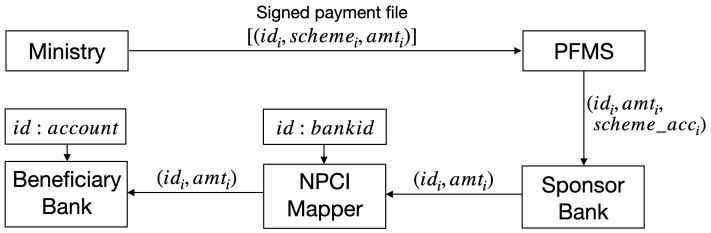


Figure 4: A simplified schematic of direct benefit transfer as it exists today.

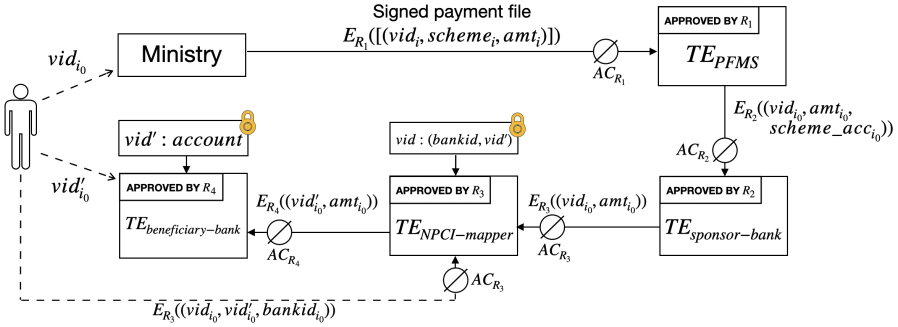


Figure 5: Our proposal for privacy-preserving direct benefit transfer. vid_{i_0} and vid'_{i_0} represent i_0's DBT-specific and Bank-specific virtual identities, respectively. Dashed arrows represent one-time onboarding steps. R_1, R_3, R_2 and R_4 represent a DBT regulator, a centralised financial regulator, and internal regulators of the sponsoring and the beneficiary banks, respectively.

(National Payments Corporation of India). NPCI maintains a mapping of citizen's Aadhaar IDs to the banks hosting their DBT accounts. This mapping allows NPCI to route the payment request for a given Aadhaar ID to the right beneficiary bank. The beneficiary bank internally maintains a mapping of its customers' Aadhaar IDs to their bank account details, and is thus able to transfer money to the right account.

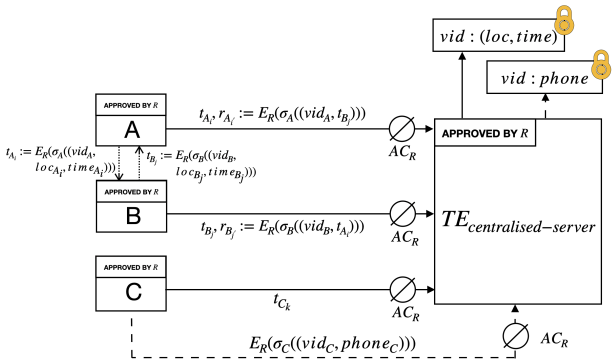
As DBT payments are primarily directed towards people who need benefits, precisely because they are structurally disadvantaged, their DBT status must be protected from future employers, law enforcement, financial providers etc., to mitigate discrimination and other socio-economic harms coming their way. Further, since DBT relies on the recipients' national Aadhaar IDs, which are typically linked with various other databases, any leakage of this information makes them directly vulnerable. Indeed, there are reports that bank and address details of millions of DBT recipients were leaked online [117]; in some cases this information was misused to even redirect DBT payments to unauthorised bank accounts [118].

We illustrate our approach for a privacy-preserving DBT in Figure 5. In our proposal, DBT recipients use a virtual identity for DBT that is completely unlinkable to the virtual identity they use for their bank account. They may generate these virtual identities - using suitably designed simple and intuitive interfaces - by an anonymous credential scheme where the credentials are issued by a centralised identity authority. Additionally, they provide the mapping of the two virtual identities, along with the bank name, to the NCPI mapper. This information is provided encrypted under the control of the financial regulator R_3 such that only the NPCI mapper TE can access it under R_3's online access control. This mechanism allows the NPCI mapper to convert payment requests against DBT-specific identities to the target bank-specific identities, while maintaining the mapping private from all agents. Regulator-controlled encryption of data in transit and storage and the properties of TEs allow for an overall privacy-preserving DBT pipeline.

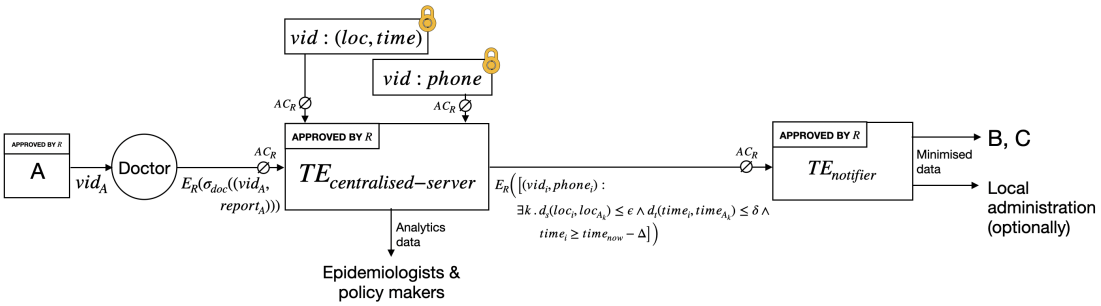
Note that data flow is controlled by different regulators along the DBT pipeline, providing a distributed approach to privacy protection. PFMS is controlled by a DBT regulator; NPCI mapper is controlled by a financial regulator, and the sponsor and beneficiary banks are controlled by their respective internal regulators.

6.2 Contact tracing

There have been a plethora of attempts recently from all over the world towards electronic app-based contact tracing for COVID-19 using a combination of GPS and Bluetooth [119, 120, 121, 122, 123, 124, 125, 126, 127].



(a) Collecting spatiotemporal information. A and B come in contact via BLE, as denoted by the dotted arrows. C does not come in contact with A or B via BLE but is spatially close within a time window, as per GPS data. vid_X represents the virtual identity of agent X; loc_{X_i} represents X's i-th recorded location; $time_{X_i}$ represents its i-th recorded time. t_{X_i} represents i-th token generated by X; r_{X_i} represents i-th receipt obtained by X; $\sigma_X()$ represents signing by X. Dashed arrows represent one-time registration steps (illustrated only for C).



(b) Tracing the contacts of infected individuals. A gets infected, as certified by the doctor's signature σ_{doc} on A's virtual identity vid_A and medical report $report_A$. d_s and d_t respectively represent chosen spatial and temporal distance functions and ϵ and δ the corresponding thresholds, as per the disease characteristics. Δ represents the infection window, the time during which A might have remained infectious. $time_{now}$ represents the time when the query was executed.

Figure 6: Contact tracing

Even keeping aside the issue of their effectiveness, some serious privacy concerns have been raised about such apps.

In most of these apps the smartphones exchange anonymous tokens when they are in proximity, and each phone keeps a record of the sent and received tokens. When an individual is infected - signalled either through a self declaration or a testing process - the tokens are uploaded to a central service.

There are broadly two approaches to contact tracing:

1. those involving a trusted central authority that can decrypt the tokens and, in turn, alert individuals and other authorities about potential infection risks [119, 120, 121, 122]. Some of these apps take special care to not upload any information about individuals who are not infected.

2. those that assume that the central authority is untrusted and use privacy preserving computations on user phones to alert individuals about their potential risks of infection [123, 124, 125, 126, 127]. The central service just facilitates access to anonymised sent tokens of infected individuals and cannot itself determine the infection status of anybody.

The following are the main privacy attacks on contact tracing apps: 1) individuals learning about other individuals as high-risk spreaders, 2) insiders at the central service learning about individuals classified as high risk, 3) exposure of social graphs of individuals, and 4) malicious claims by individuals forcing quarantine on others. See [125] for a vulnerability analysis of some popular approaches.

The centralised approaches clearly suffer from many of the above privacy risks. While alerting local authorities about infection risks is clearly more effective from a public health perspective, to enable them to identify hotspots and make crucial policy decisions, it is mainly the privacy concerns that sometimes motivate the second approach. Also, it is well known that location data of individuals can be used to orchestrate de-anonymisation attacks [55], and hence many of the above approaches adopt the policy of not using geolocation data for contact tracing despite their obvious usefulness. In addition, Bluetooth based proximity sensing - which are isolated communication events over narrow temporal windows between two smartphones - is ineffective for risk assessment of indirect transmission through contaminated surfaces, where the virus can survive for long hours or even days. Such risk assessment will require computation of intersection of space-time volumes of trajectories which will be difficult in a decentralised approach. It appears that the privacy considerations have forced many of these approaches to adopt overly defensive decentralised designs at the cost of effectiveness.

In contrast, we propose an architecture where governments can collect fine-grained location and proximity data of citizens, but under regulated access control and purpose limitation. Such a design can support both short-range peer-to-peer communication technologies such as BLE and GPS based location tracking. Additionally, centralised computing can support space-time intersections.

In Figure 6, we show the design of a *state-mandated* contact-tracing app that, in addition to protecting against the privacy attacks outlined earlier, can also protect against attacks by individuals who may maliciously try to pose as low-risk on the app, for example to get around restrictions (attack 5).

As before, we require all storage and transit data to be encrypted under a regulator-controlled encryption scheme, and that they be accessible only to pre-approved TEs. We also require the app to be running as a TE on the users' phones (e.g., within a trusted zone on the phone).

We assume that everyone registers with the app using a phone number and a virtual identity unlinkable to their other identities. Periodically, say after every few minutes, each device records its current GPS location and time. The tuple made up of the registered virtual identity and the recorded location and time is signed by the device and encrypted controlled by the regulator, thus creating an ephemeral "token" to be shared with other nearby devices over BLE. When a token is received from another device, a tuple containing the virtual identity of self and the incoming token is created, signed and stored in a regulator-controlled encrypted form, thus creating a "receipt". Periodically, once every few hours, all locally stored tokens and receipts are uploaded to a centralised server TE, which stores them under regulated access control as a mapping between registered virtual identities and all their spatiotemporal coordinates. For all the receipts, the centralised server TE stores the same location and time for the receiving virtual identity as in the token it received, thus modelling the close proximity of BLE contacts.

When a person tests positive, they present their virtual identity to a medical personnel who uploads a signed report certifying the person's infection status to the centralised server TE. This event allows the centralised server TE to fetch all the virtual identities whose recorded spatiotemporal coordinates intersects within a certain threshold, as determined by the disease parameters, with the infected person's coordinates. As the recorded (location, time) tuples of any two individuals who come in contact via BLE necessarily collide in our approach, the virtual identities of all BLE contacts can be identified with high precision. Moreover, virtual identities of individuals who did not come under contact via BLE but were spatially nearby in a time window as per GPS data are also identified.

A notifier TE securely obtains the registered phone numbers corresponding to these virtual identities from the centralised server TE and sends suitably minimised notifications to them, and also perhaps to the local administration according to local regulations. The collected location data can also be used independently by epidemiologists and policy makers in aggregate form to help them understand the infection pathways and identify areas which need more resources.

Note that attack 1 is protected by the encryption of all sent tokens; attacks 2 and 3 are protected by the properties of TEs and regulatory access control; attack 4 is protected by devices signing their correct spatiotemporal coordinates against their virtual identity before sending tokens or receipts. Attack 5 is mitigated by requiring the app to run within a trusted zone on users' devices, to prevent individuals from not sending tokens and receipts periodically or sending junk data.

7 Conclusion

We have presented the design sketch of an operational architecture for privacy-by-design [3] based on regulatory oversight, regulated access control, purpose limitation and data minimisation. We have established the need

for such an architecture by highlighting limitations in existing approaches and some public service application designs. We have demonstrated its usefulness with some case studies.

While we have explored the feasibility of our architecture based on existing techniques in computer science, some of them will definitely require further strengthening. There also needs to be detailed performance and usability evaluations, especially in the context of large-scale database and AI applications. Techniques to help a regulator assess the privacy risks of TEs also need to be investigated. These are interesting open problems that need to be solved to create practical systems for the future with built-in end-to-end privacy protection.

References

[1] Y. N. Harari, "Yuval Noah Harari: the world after coronavirus." https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75, 2020. [Online; posted 20-March-2020].

[2] "K S Puttaswamy v Union of India (2017): Writ Petition (Civil) No 494 of 2012, Supreme Court judgment dated 24 August." https://www.scobserver.in/court-case/fundamental-right-to-privacy, 2017. [Accessed January 9, 2018].

[3] A. Cavoukian, "Privacy by Design: The 7 Foundational Principles." https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf, August 2009. Revised: January, 2011.

[4] OECD, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data." https://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm, 1980. [Online accessed 12-May-2020].

[5] D. J. Solove, *The Digital Person: Technology And Privacy In The Information Age*. New York, NY, USA: New York University Press, 2004.

[6] The European Parliament and the Council of European Union, "Regulation (EU) no 2016/679," 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679.

[7] Ministry of Law and Justice, Government of India, "The Personal Data Protection Bill, 2019." https://www.prsindia.org/billtrack/personal-data-protection-bill-2019, 2019. [Online accessed 12-May-2020].

[8] 111th Congress (2009-2010), "Social Security Number Protection Act of 2010." https://www.congress.gov/bill/111th-congress/senate-bill/3789, 2010. [Accessed Novenmber 3, 2019].

[9] Centers for Medicare & Medicaid Services, "The Health Insurance Portability and Accountability Act of 1996 (HIPAA)." http://www.cms.hhs.gov/hipaa/, 1996.

[10] The Guardian, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach." https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

[11] The London School of Economics and Political Science, "The Identity Project: An assessment of the UK Identity Cards Bill and its implications." http://www.lse.ac.uk/management/research/identityproject/identityreport.pdf, June 2005.

[12] S. Agrawal, S. Banerjee, and S. Sharma, "Privacy and Security of Aadhaar: A Computer Science Perspective," *Economic and Political Weekly*, vol. Vol. 52, 16 2017.

[13] J. Temperton, "NHS care.data scheme closed after years of controversy." https://www.wired.co.uk/article/care-data-nhs-england-closed, 2016. [Online July 6, 2016].

[14] R. N. Charette, "Australians Say No Thanks to Electronic Health Records." https://spectrum.ieee.org/riskfactor/computing/it/australians-choosing-to-optout-of-controversial-my-health-record-system 2018. [Online July 27, 2018].

[15] S. Shrikanth and B. Parkin, "India plan to merge ID with health records raises privacy worries." https://www.ft.com/content/4fbb2334-a864-11e9-984c-fac8325aaa04, July 2019. [Online; posted 17-July-2019].

[16] K. Zetter, "Voter Privacy Is Gone – Get Over It." https://www.wired.com/2008/01/voter-privacy-i/, 2008. [Online January 31, 2008].

[17] M. Pal, "Are citizens compromising their privacy when registering to vote?." https://gcn.com/articles/2018/12/11/voting-data-privacy.aspx, 2018. [Online; posted December 11, 2018].

[18] P. S. Tripathi, "Concerns over linking Aadhaar to voter ID and social media accounts." https://frontline.thehindu.com/the-nation/article29407553.ece, September 2019. [Online; posted 27-Septeber-2019].

[19] epic.org, "Equifax Data Breach." https://epic.org/privacy/data-breach/equifax/, 2019. [Online accessed 3-November-2019].

[20] B. Chugh and M. Raghavan, "The RBI's proposed Public Credit Registry and its implications for the credit reporting system in India." https://www.dvara.com/blog/2019/06/18/the-rbis-proposed-public-credit-registry-and-its-implications-for-the-credit-reporting-system-in-i 2019. [Online posted 18-June-2019].

[21] Yle Uutiset, "Launch of Incomes Register dogged by data security concerns." https://yle.fi/uutiset/osasto/news/launch_of_incomes_register_dogged_by_data_security_concerns/10576057, 2018. [Online posted 30-December-2018].

[22] K. Houser and D. Sanders, "The Use of Big Data Analytics by the IRS: Efficient Solution or the End of Privacy as We Know it?," *Vanderbilt Journal of Entertainment & Technology Law*, vol. 19, April 2017.

[23] GOV.UK Press Release, "National identity register destroyed as government consigns ID card scheme to history." https://www.gov.uk/government/news/national-identity-register-destroyed-as-government-consigns-id-card-scheme-to-history, 2011. [Online posted 10-February-2011].

[24] Unique Identification Authority of India, "Aadhaar." https://uidai.gov.in, 2020. [Accessed May 31, 2020].

[25] R. Khera, *Dissent on Aadhaar: Big Data Meets Big Brother*. Orient BlackSwan, 2019. Edited volume.

[26] A. Narayanan and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets," in *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, SP '08, (Washington, DC, USA), pp. 111–125, IEEE Computer Society, 2008.

[27] D. Chaum, "Security Without Identification: Transaction Systems to Make Big Brother Obsolete," *Commun. ACM*, vol. 28, pp. 1030–1044, Oct. 1985.

[28] D. Chaum, "Showing Credentials Without Identification," in *Advances in Cryptology — EUROCRYPT' 85* (F. Pichler, ed.), (Berlin, Heidelberg), pp. 241–244, Springer Berlin Heidelberg, 1986.

[29] K. Zetter, "NSA Whistleblower: The Ultimate Insider Attack." https://www.wired.com/2013/06/nsa-leaker-ultimate-insider/, September 2009.

[30] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*. fairmlbook.org, 2019. http://www.fairmlbook.org.

[31] N. Verma and S. Dawar, "Digital transformation in the indian government," *Commun. ACM*, vol. 62, p. 50–53, Oct. 2019.

[32] V. Raghavan, S. Jain, and P. Varma, "India stack—digital infrastructure as public good," *Commun. ACM*, vol. 62, p. 76–81, Oct. 2019.

[33] G. J. Simmons, "Symmetric and asymmetric encryption," *ACM Comput. Surv.*, vol. 11, pp. 305–330, Dec. 1979.

[34] W. Diffie and M. E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, vol. 22, pp. 644–654, November 1976.

[35] D. Boneh and M. K. Franklin, "Identity-Based Encryption from the Weil Pairing," in *Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '01, (Berlin, Heidelberg), p. 213–229, Springer-Verlag, 2001.

[36] D. Chaum, "Blind Signatures for Untraceable Payments," in *Advances in Cryptology* (D. Chaum, R. L. Rivest, and A. T. Sherman, eds.), (Boston, MA), pp. 199–203, Springer US, 1983.

[37] O. Goldreich, S. Micali, and A. Wigderson, "Proofs That Yield Nothing but Their Validity or All Languages in NP Have Zero-knowledge Proof Systems," *J. ACM*, vol. 38, pp. 690–728, July 1991.

[38] U. Feige and A. Shamir, "Zero Knowledge Proofs of Knowledge in Two Rounds," in *Proceedings of the 9th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '89, (Berlin, Heidelberg), pp. 526–544, Springer-Verlag, 1990.

[39] P. Vullers and G. Alpár, "Efficient Selective Disclosure on Smart Cards Using Idemix," in *Policies and Research in Identity Management* (S. Fischer-Hübner, E. de Leeuw, and C. Mitchell, eds.), (Berlin, Heidelberg), pp. 53–67, Springer Berlin Heidelberg, 2013.

[40] A. Pfitzmann and M. Hansen, "Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology," February 2008.

[41] D. Chaum and J.-H. Evertse, "A Secure and Privacy-Protecting Protocol for Transmitting Personal Information Between Organizations," in *Advances in Cryptology — CRYPTO' 86* (A. M. Odlyzko, ed.), (Berlin, Heidelberg), pp. 118–167, Springer Berlin Heidelberg, 1987.

[42] J. Camenisch and A. Lysyanskaya, "An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation," in *Advances in Cryptology — EUROCRYPT 2001* (B. Pfitzmann, ed.), (Berlin, Heidelberg), pp. 93–118, Springer Berlin Heidelberg, 2001.

[43] J. Camenisch and A. Lysyanskaya, "Dynamic Accumulators and Application to Efficient Revocation of Anonymous Credentials," in *Proceedings of the 22nd Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '02, (Berlin, Heidelberg), p. 61–76, Springer-Verlag, 2002.

[44] D. Chaum and E. van Heyst, "Group Signatures," in *Advances in Cryptology — EUROCRYPT '91* (D. W. Davies, ed.), (Berlin, Heidelberg), pp. 257–265, Springer Berlin Heidelberg, 1991.

[45] I. Teranishi, J. Furukawa, and K. Sako, "k-Times Anonymous Authentication (Extended Abstract)," in *Advances in Cryptology - ASIACRYPT 2004* (P. J. Lee, ed.), (Berlin, Heidelberg), pp. 308–322, Springer Berlin Heidelberg, 2004.

[46] L. Nguyen and R. Safavi-Naini, "Dynamic k-Times Anonymous Authentication," in *Applied Cryptography and Network Security* (J. Ioannidis, A. Keromytis, and M. Yung, eds.), (Berlin, Heidelberg), pp. 318–333, Springer Berlin Heidelberg, 2005.

[47] I. Teranishi and K. Sako, "k-Times Anonymous Authentication with a Constant Proving Cost," in *Proceedings of the 9th International Conference on Theory and Practice of Public-Key Cryptography*, PKC'06, (Berlin, Heidelberg), p. 525–542, Springer-Verlag, 2006.

[48] P. P. Tsang, M. H. Au, A. Kapadia, and S. W. Smith, "Blacklistable Anonymous Credentials: Blocking Misbehaving Users without TTPs," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, CCS '07, (New York, NY, USA), p. 72–81, Association for Computing Machinery, 2007.

[49] F. Corella, "Credential Sharing: A Pitfall of Anonymous Credentials." `https://pomcor.com/2011/12/19/credential-sharing-a-pitfall-of-anonymous-credentials/#1`. [Blog Post; Online; Accessed May 24, 2020].

[50] D. L. Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Commun. ACM*, vol. 24, p. 84–90, Feb. 1981.

[51] C. C. Aggarwal, "On k-anonymity and the Curse of Dimensionality," in *Proceedings of the 31st International Conference on Very Large Data Bases*, VLDB '05, pp. 901–909, VLDB Endowment, 2005.

[52] A. Datta, D. Sharma, and A. Sinha, "Provable De-anonymization of Large Datasets with Sparse Dimensions," in *Principles of Security and Trust* (P. Degano and J. D. Guttman, eds.), (Berlin, Heidelberg), pp. 229–248, Springer Berlin Heidelberg, 2012.

[53] A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," in *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, SP '09, (Washington, DC, USA), pp. 173–187, IEEE Computer Society, 2009.

[54] A. Narayanan, E. Shi, and B. I. P. Rubinstein, "Link Prediction by De-anonymization: How We Won the Kaggle Social Network Challenge." `https://arxiv.org/pdf/1102.4374.pdf`, 2011.

[55] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility," *Scientific Reports*, vol. 3, 2013.

[56] A. Narayanan, H. Paskov, N. Z. Gong, J. Bethencourt, E. Stefanov, E. C. R. Shin, and D. Song, "On the Feasibility of Internet-Scale Author Identification," in *2012 IEEE Symposium on Security and Privacy*, pp. 300–314, May 2012.

[57] J. Su, A. Shukla, S. Goel, and A. Narayanan, "De-anonymizing Web Browsing Data with Social Networks," in *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, (Republic and Canton of Geneva, Switzerland), pp. 1261–1269, International World Wide Web Conferences Steering Committee, 2017.

[58] I. Dinur and K. Nissim, "Revealing Information While Preserving Privacy," in *Proceedings of the Twenty-second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '03, (New York, NY, USA), pp. 202–210, ACM, 2003.

[59] J. Kleinberg, C. Papadimitriou, and P. Raghavan, "Auditing Boolean Attributes," in *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '00, (New York, NY, USA), pp. 86–91, ACM, 2000.

[60] T. Dalenius, "Towards a methodology for statistical disclosure control," *Statistik Tidskrift*, vol. 15, no. 429-444, pp. 2–1, 1977.

[61] A. Ghosh and R. Kleinberg, "Inferential Privacy Guarantees for Differentially Private Mechanisms," in *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, pp. 9:1–9:3, 2017.

[62] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," in *Theory of Cryptography* (S. Halevi and T. Rabin, eds.), (Berlin, Heidelberg), pp. 265–284, Springer Berlin Heidelberg, 2006.

[63] C. Dwork, "Differential Privacy," in *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II*, ICALP'06, (Berlin, Heidelberg), pp. 1–12, Springer-Verlag, 2006.

[64] A. Blum, K. Ligett, and A. Roth, "A Learning Theory Approach to Non-Interactive Database Privacy," in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, (New York, NY, USA), p. 609–618, Association for Computing Machinery, 2008.

[65] F. McSherry and K. Talwar, "Mechanism Design via Differential Privacy," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '07, (USA), p. 94–103, IEEE Computer Society, 2007.

[66] M. Hardt, K. Ligett, and F. McSherry, "A Simple and Practical Algorithm for Differentially Private Data Release," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'12, (Red Hook, NY, USA), p. 2339–2347, Curran Associates Inc., 2012.

[67] S. Sen, S. Guha, A. Datta, S. K. Rajamani, J. Tsai, and J. M. Wing, "Bootstrapping Privacy Compliance in Big Data Systems," in *Proceedings of the 2014 IEEE Symposium on Security and Privacy*, SP '14, (Washington, DC, USA), pp. 327–342, IEEE Computer Society, 2014.

[68] L. Wang, J. P. Near, N. Somani, P. Gao, A. Low, D. Dao, and D. Song, "Data Capsule: A New Paradigm for Automatic Compliance with Data Privacy Regulations." `https://arxiv.org/pdf/1909.00077.pdf`, 2019.

[69] K. Hayati and M. Abadi, "Language-Based Enforcement of Privacy Policies," in *Privacy Enhancing Technologies* (D. Martin and A. Serjantov, eds.), (Berlin, Heidelberg), pp. 302–313, Springer Berlin Heidelberg, 2005.

[70] Paul Ashley and Satoshi Hada and Günter Karjoth and Calvin Powers and Matthias Schunter, "The Enterprise Privacy Authorization Language (EPAL) - How to Enforce Privacy throughout an Enterprise." `https://www.w3.org/2003/p3p-ws/pp/ibm3.html`, 2003. [Accessed May 31, 2020].

[71] L. F. Cranor, "P3P: Making Privacy Policies More Useful," *IEEE Security and Privacy*, vol. 1, p. 50–55, Nov. 2003.

[72] D. E. Denning, "A Lattice Model of Secure Information Flow," *Commun. ACM*, vol. 19, p. 236–243, May 1976.

[73] C. Baier and J.-P. Katoen, *Principles of Model Checking (Representation and Mind Series)*. The MIT Press, 2008.

[74] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Hippocratic Databases," in *Proceedings of the 28th International Conference on Very Large Data Bases*, VLDB '02, p. 143–154, VLDB Endowment, 2002.

[75] A. Masoumzadeh and J. B. D. Joshi, "PuRBAC: Purpose-Aware Role-Based Access Control," in *On the Move to Meaningful Internet Systems: OTM 2008* (R. Meersman and Z. Tari, eds.), (Berlin, Heidelberg), pp. 1104–1121, Springer Berlin Heidelberg, 2008.

[76] J.-W. Byun and N. Li, "Purpose based access control for privacy protection in relational database systems," *The VLDB Journal*, vol. 17, pp. 603–619, Jul 2008.

[77] M. Jafari, P. W. Fong, R. Safavi-Naini, K. Barker, and N. P. Sheppard, "Towards Defining Semantic Foundations for Purpose-Based Privacy Policies," in *Proceedings of the First ACM Conference on Data and Application Security and Privacy*, CODASPY '11, (New York, NY, USA), p. 213–224, Association for Computing Machinery, 2011.

[78] M. C. Tschantz, A. Datta, and J. M. Wing, "Formalizing and Enforcing Purpose Restrictions in Privacy Policies," in *2012 IEEE Symposium on Security and Privacy*, pp. 176–190, May 2012.

[79] R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Commun. ACM*, vol. 21, p. 120–126, Feb. 1978.

[80] T. El Gamal, "A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms," in *Proceedings of CRYPTO 84 on Advances in Cryptology*, (Berlin, Heidelberg), p. 10–18, Springer-Verlag, 1985.

[81] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," in *Proceedings of the 17th International Conference on Theory and Application of Cryptographic Techniques*, EUROCRYPT'99, (Berlin, Heidelberg), p. 223–238, Springer-Verlag, 1999.

[82] C. Gentry, "Fully Homomorphic Encryption Using Ideal Lattices," in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, (New York, NY, USA), p. 169–178, Association for Computing Machinery, 2009.

[83] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A Survey on Homomorphic Encryption Schemes: Theory and Implementation," *ACM Comput. Surv.*, vol. 51, July 2018.

[84] D. Boneh, A. Sahai, and B. Waters, "Functional Encryption: Definitions and Challenges," in *Theory of Cryptography* (Y. Ishai, ed.), (Berlin, Heidelberg), pp. 253–273, Springer Berlin Heidelberg, 2011.

[85] A. C. Yao, "Protocols for Secure Computations," in *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, SFCS '82, (USA), p. 160–164, IEEE Computer Society, 1982.

[86] J. I. Choi and K. R. B. Butler, "Secure Multiparty Computation and Trusted Hardware: Examining Adoption Challenges and Opportunities," *Security and Communication Networks*, vol. 2019, p. 1368905, Apr 2019.

[87] F. McKeen, I. Alexandrovich, A. Berenzon, C. V. Rozas, H. Shafi, V. Shanbhogue, and U. R. Savagaonkar, "Innovative Instructions and Software Model for Isolated Execution," in *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy*, HASP '13, (New York, NY, USA), Association for Computing Machinery, 2013.

[88] I. Anati, S. Gueron, S. P. Johnson, and V. R. Scarlata, "Innovative Technology for CPU Based Attestation and Sealing." `https://software.intel.com/content/www/us/en/develop/articles/innovative-technology-for-cpu-based-attestation-and-sealing.html`, 2013. [Online; accessed 4-June-2020].

[89] Trusted Computing Group, "TPM main specification version 1.2: part 1 design principles." `https://trustedcomputinggroup.org/wp-content/uploads/TPM-Main-Part-1-Design-Principles_v1.2_rev116_01032011.pdf`, 2011. [Online; accessed 31-May-2020].

[90] ARM Ltd., "Building a Secure System using TrustZone® Technology." `http://infocenter.arm.com/help/topic/com.arm.doc.prd29-genc-009492c/PRD29-GENC-009492C_trustzone_security_whitepaper.pdf`, 2005-2009. [Online; accessed 31-May-2020].

[91] Y. Xu, W. Cui, and M. Peinado, "Controlled-channel attacks: Deterministic side channels for untrusted operating systems," in *Proceedings of the 2015 IEEE Symposium on Security and Privacy*, SP '15, (USA), p. 640–656, IEEE Computer Society, 2015.

[92] M. Schwarz, S. Weiser, D. Gruss, C. Maurice, and S. Mangard, "Malware Guard Extension: Using SGX to Conceal Cache Attacks." `https://arxiv.org/pdf/1702.08719.pdf`, 2017.

[93] W. Wang, G. Chen, X. Pan, Y. Zhang, X. Wang, V. Bindschaedler, H. Tang, and C. A. Gunter, "Leaky Cauldron on the Dark Land: Understanding Memory Side-Channel Hazards in SGX," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, (New York, NY, USA), p. 2421–2434, Association for Computing Machinery, 2017.

[94] J. Van Bulck, M. Minkin, O. Weisse, D. Genkin, B. Kasikci, F. Piessens, M. Silberstein, T. F. Wenisch, Y. Yarom, and R. Strackx, "Foreshadow: Extracting the Keys to the Intel SGX Kingdom with Transient out-of-Order Execution," in *Proceedings of the 27th USENIX Conference on Security Symposium*, SEC'18, (USA), p. 991–1008, USENIX Association, 2018.

[95] V. Costan, I. Lebedev, and S. Devadas, "Sanctum: Minimal Hardware Extensions for Strong Software Isolation," in *Proceedings of the 25th USENIX Conference on Security Symposium*, SEC'16, (USA), p. 857–874, USENIX Association, 2016.

[96] D. X. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," in *Proceedings of the 2000 IEEE Symposium on Security and Privacy*, SP '00, (USA), p. 44, IEEE Computer Society, 2000.

[97] J. Baek, R. Safavi-Naini, and W. Susilo, "Public Key Encryption with Keyword Search Revisited," in *Proceeding Sof the International Conference on Computational Science and Its Applications, Part I*, ICCSA '08, (Berlin, Heidelberg), p. 1249–1259, Springer-Verlag, 2008.

[98] E.-J. Goh, "Secure indexes." `https://eprint.iacr.org/2003/216.pdf`, 2004. [Online; accessed 4-June-2020].

[99] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," *IACR Cryptology ePrint Archive*, vol. 2006, p. 210, 2006.

[100] H. Hacigümüş, B. Iyer, C. Li, and S. Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model," in *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*, SIGMOD '02, (New York, NY, USA), p. 216–227, Association for Computing Machinery, 2002.

[101] R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, "CryptDB: Protecting Confidentiality with Encrypted Query Processing," in *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles*, SOSP '11, (New York, NY, USA), p. 85–100, Association for Computing Machinery, 2011.

[102] M. S. Islam, M. Kuzu, and M. Kantarcioglu, "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation," in *in Network and Distributed System Security Symposium (NDSS*, 2012.

[103] D. Cash, P. Grubbs, J. Perry, and T. Ristenpart, "Leakage-Abuse Attacks Against Searchable Encryption," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, (New York, NY, USA), p. 668–679, Association for Computing Machinery, 2015.

[104] Y. Zhang, J. Katz, and C. Papamanthou, "All Your Queries Are Belong to Us: The Power of File-Injection Attacks on Searchable Encryption," in *Proceedings of the 25th USENIX Conference on Security Symposium*, SEC'16, (USA), p. 707–720, USENIX Association, 2016.

[105] M. Naveed, S. Kamara, and C. V. Wright, "Inference Attacks on Property-Preserving Encrypted Databases," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, (New York, NY, USA), p. 644–655, Association for Computing Machinery, 2015.

[106] R. Ostrovsky, "Efficient Computation on Oblivious RAMs," in *Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing*, STOC '90, (New York, NY, USA), p. 514–523, Association for Computing Machinery, 1990.

[107] O. Goldreich and R. Ostrovsky, "Software Protection and Simulation on Oblivious RAMs," *J. ACM*, vol. 43, p. 431–473, May 1996.

[108] C. Priebe, K. Vaswani, and M. Costa, "EnclaveDB – A Secure Database using SGX," in *Proceedings of the IEEE Symposium on Security and Privacy, May 2018*, IEEE, May 2018.

[109] E. Kushilevitz and R. Ostrovsky, "Replication is Not Needed: Single Database, Computationally-Private Information Retrieval," in *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, FOCS '97, (USA), p. 364, IEEE Computer Society, 1997.

[110] A. Kiayias, N. Leonardos, H. Lipmaa, K. Pavlyk, and Q. Tang, "Optimal Rate Private Information Retrieval from Homomorphic Encryption," *Proceedings on Privacy Enhancing Technologies*, vol. 2015, 06 2015.

[111] S. Volos, K. Vaswani, and R. Bruno, "Graviton: Trusted Execution Environments on GPUs," in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'18, (USA), p. 681–696, USENIX Association, 2018.

[112] TPC, "TPC-C Homepage." `http://www.tpc.org/tpcc/`, 2017. [Online; Accessed June 4, 2020].

[113] S. Neuvonen, A. Wolski, M. manner, and V. Raatikka, "Telecom Application Transaction Processing Benchmark." `http://tatpbenchmark.sourceforge.net`, 2017. [Online; Accessed June 4, 2020].

[114] Digital Content Protection LLC, "HDCP Interface Independent Adaptation Specification." `https://www.digital-cp.com/sites/default/files/HDCP%20Interface%20Independent%20Adaptation%20Specification%20Rev2_3.pdf`, March 2018. Revision: 2.3.

[115] Government of India, "Direct Benefit Transfer, Government of India." `https://dbtbharat.gov.in`, 2020. [Accessed May 31, 2020].

[116] DBT Mission, Cabinet Secretariat, Government of India, "Standard Operating Procedure (SOP) Modules for Direct Benefit Transfer (DBT)." `https://dbtbharat.gov.in/data/documents/Standard-Operating-Procedures.pdf`. [Online; Accessed April 14, 2020].

[117] A. Sinha and S. Kodali, "Information Security Practices of Aadhaar (or lack thereof): A documentation of public availability of Aadhaar Numbers with sensitive personal financial information." https://cis-india.org/internet-governance/information-security-practices-of-aadhaar-or-lack-thereof-a-documentation-of-public-availability-of [Online; Accessed May 19, 2020].

[118] A. Venkatanarayanan and S. Lakshmanan, "Aadhaar Mess: How Airtel Pulled Off Its Rs 190 Crore Magic Trick." https://thewire.in/banking/airtel-aadhaar-uidai. [Online; Accessed April 14, 2020].

[119] A. Krishnan, "China's high-tech battle against COVID-19." https://www.thehindu.com/opinion/lead/chinas-high-tech-battle-against-covid-19/article30993814.ece, 2020. [Online; Accessed April 14, 2020].

[120] J. W. Sonn, "Coronavirus: South Korea's success in controlling disease is due to its acceptance of surveillance." https://theconversation.com/coronavirus-south-koreas-success-in-controlling-disease-is-due-to-its-acceptance-of-surveillance-13 2020. [Online; Accessed April 14, 2020].

[121] Ministry of Health, Singapore Government, "Tracetogether app." https://www.tracetogether.gov.sg, March 2020. [Accessed June 4, 2020].

[122] National Informatics Centre, Ministry of Electronics and Information Technology, Government of India, "Aarogya Setu Mobile App." https://www.mygov.in/aarogya-setu-app/, 2020. [Accessed May 31, 2020].

[123] R. Raskar, I. Schunemann, R. Barbar, K. Vilcans, J. Gray, P. Vepakomma, S. Kapa, A. Nuzzo, R. Gupta, A. Berke, D. Greenwood, C. Keegan, S. Kanaparti, R. Beaudry, D. Stansbury, B. B. Arcila, R. Kanaparti, V. Pamplona, F. M. Benedetti, A. Clough, R. Das, K. Jain, K. Louisy, G. Nadeau, V. Pamplona, S. Penrod, Y. Rajaee, A. Singh, G. Storm, and J. Werner, "Apps Gone Rogue: Maintaining Personal Privacy in an Epidemic." https://arxiv.org/pdf/2003.08567.pdf, 2020. [Online; accessed June 4, 2020].

[124] R. Canetti, A. Trachtenberg, and M. Varia, "Anonymous Collocation Discovery: Harnessing Privacy to Tame the Coronavirus." https://arxiv.org/pdf/2003.13670.pdf, 2020. [Online; accessed June 4, 2020].

[125] N. Trieu, K. Shehata, P. Saxena, R. Shokri, and D. Song, "Epione: Lightweight Contact Tracing with Strong Privacy." https://arxiv.org/pdf/2004.13293.pdf, 2020. [Online; accessed June 4, 2020].

[126] Apple Inc. and Google LLC, "Privacy-Preserving Contact Tracing." https://www.apple.com/covid19/contacttracing, 2020. [Accessed June 4, 2020].

[127] R. Rivest, J. Callas, R. Canetti, K. Esvelt, D. K. Gillmor, Y. T. Kalai, A. Lysyanskaya, A. Norige, R. Raskar, A. Shamir, E. Shen, I. Soibelman, M. Specter, V. Teague, A. Trachtenberg, M. Varia, M. Viera, D. Weitzner, J. Wilkinson, and M. Zissman, "The PACT protocol specification." https://pact.mit.edu/wp-content/uploads/2020/04/The-PACT-protocol-specification-ver-0.1.pdf, April 2020. Version: 0.1.