SCTE·ISBE CABLE-TEC EXPO 2019 NEW ORLEANS, LA SEPT. 30-OCT. 3

2019 Fall Technical Forum SCTE·ISBE • NCTA • CABLELABS

Opting-In: Designing Privacy Tracking for Consumer Confidentiality & Cryptographic Assurance for Enterprises

A Technical Paper prepared for SCTE•ISBE by

Brian A. Scriber
Distinguished Technologist and VP of Security Technologies
CableLabs
858 Coal Creek Circle, Louisville, CO 80027
@brianscriber
b.scriber@cablelabs.com

Table of Contents

List of Figures

List of Tables

Introduction

Privacy, particularly consumer privacy, has lived a dynamic existence over the last decade. Consumer views have changed as has the regulatory environment. With GDPR in the EU, PIPEDA in Canada, and with CPA in California, there is an immediate need for technical solutions to efficiently run our businesses in this strong regulatory environment. What if you and your department could help enable a revenue opportunity in this space as opposed to just mitigating regulatory risk? Join us to explore how self-sovereign identity and opt-in/opt-out tracking can work hand in hand using some of the nascent tools in cryptography and software development. We will explore transaction signing, distributed verification, collaborative acknowledgments, time synchronization, user-directed sharing of protected information, the Right to be Forgotten, and cryptographic key distribution systems enforced by smart contracts. A key part of this paper investigates how data analytics and privacy can coexist without one working opposite the other through the application of advanced key management techniques, zero knowledge proofs, and smart contracts.

Privacy and International Policy Initiatives

Privacy has regional meaning and with different practices, the legislative and compliance directives differ as well. There are some themes that can be extracted from the California Consumer Privacy Act (CCPA) in the United States, the General Data Protection Regulation (GDPR) in the European Union, and PIPEDA in Canada. Each is unique, but a policy that addresses several of these could be successful for companies deploying capabilities in each of these regions. Those key attributes include the following:

1. Defining Protected Data

As mentioned in the intro to this section, protected data has differing definitions based on region.

> "Personal data is any information relating to an individual, whether it relates to his or her private, professional or public life. It can be anything from a name, a home address, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer's IP address" -- GDPR

Figure 1 GDPR Definition of Personal Data

The United States has a couple differing places to look for data protection; with state-level initiatives, look for these definitions to remain a bit fluid over the near future.

> "Individually identifiable health information held or transmitted by a covered entity or its business associate, in any form or medium, whether electronic, on paper, or oral" – Privacy Rule, HIPAA

Figure 2: HIPAA Privacy Rule

In Canada, PIPEDA is the prevailing legal doctrine related to privacy and it defines the protected data

> "Data that contains any factual or subjective information, recorded or not, about an identifiable individual. This consists of not only personally identifiable information (PII) such as name, age, ID number and ethnicity or medical records, employee files, credit records and so on, but also opinions, evaluations, comments, social status and disciplinary actions." -- PIPEDA

Figure 3 Protected Data in PIPEDA

thus (it goes into further detail around sensitive data not protected explicitly by PIPEDA as well):

2. De-Identification of Data

According to the GDPR Recital 26, "Pseudonymised data is still considered personal data" and the US Department of Health and Human Services goes into great detail over de-identification of data and the complexities therein[i]. The undercurrent of all of this is that simply taking the name out of a data set, and optionally replacing it with a number/key/identifier, is insufficient for privacy protections because the collection of data held may still be enough to identify participants and to enable the ability to re-identify them.

3. Opt-In, Opt-Out, and Explicit Consent

The ability for consumers to opt out of the use of their personal data and to be most generally compliant, a policy of opting in for use of that data is a path some are taking as explicit consent, such as in a clear affirmative action, for data use is a requirement of GDPR.

4. Data Portability

Data portability is a requirement that looks for consumers to be able to extract the data related to them, including the personal information, but to do so using a "standard" mechanism. The implication of this requirement is that other systems should also be able to consume this format as well. Personal information, the definition of what constitutes Personally Identifiable Information, Protected Health Information, and Personal Data differs between regions as well. That difference makes the standardization of this data complicated. Heterogeneous semantic mapping between these, and the differences in how data is stored, highlight inconsistencies (e.g. Social Security Numbers in the USA and Social Insurance Number in Canada).

5. Right to be Forgotten and Right to Deletion

In the EU, there exists a Right to be Forgotten (RTBF) and in the CCPA there is a "right to deletion" RTD which were both created for a consumer to be able to exert a level of control over any future use of their data. When RTBF or RTD are exercised, the personal information related to that consumer is to no longer be used.

6. Data Provenance

The requirement to be able to share with a consumer where a given article of data originated, how the possessing company came to own it, and potentially whether or not it was purchased, guessed, inferred, or otherwise arrived at is a part of this privacy landscape that this paper will show may drive some of the most significant changes. The granularity of data, the mapping to origin metadata, and the combination of technology and process required to deliver that information are going to drive costs in enterprise development.

7. Privacy and the Network Operator

The GDPR states that, as shown in Figure 2., a computer's IP address is part of the Personal Data protection. As a network operator who bills for connectivity services, this is a critical aspect of the data used throughout operational systems. While this paper is not offering legal advice, Article 49 of the GDPR does have derogations and special conditions where there is a contract in place or in service of public interest.

8. Costs of Non-Compliance

While the above technical implications and process flows for potentially protected data collection are potentially costly, so too are the potential penalties for non-compliance with the laws. GDPR has potential fines of up to €20 million or 4% of annual turnover, whichever is greater. The CCPA has a $7500 cap per intentional violation ($2500 cap for unintentional violations), but the definition of a violation is likely[ii] to be interpreted to mean per incident per consumer in line with the data breech class action section of the law.

9. Revenue Opportunities

While the costs are high, there are opportunities for companies who find the path to successfully navigate the privacy concerns, that can scale their solutions, and which can offer such solutions as shields to the misuse of protected data while simultaneously providing a similar protection for enterprises with legitimate needs for use of data.

Data Modeling

10. Databases

Since we are talking about requirements on data, the important place to look inside any enterprise is at the database level. Several databases will exist within a typical enterprise, including network operators. There are customer support databases, billing databases, service databases, advertising databases, operational databases, data warehouses, reports that utilize data, online systems, back-end systems – each with their own databases to support their missions. Data is pervasive, and it is important to protect it. Current data models follow a similar format to the ERD shown in Figure 4 Conventional Database Entity Relationship Data Model. This model shows the relationships between a person, their means of contact,

and any organizations they belong to, among a few other additional relationships and details.

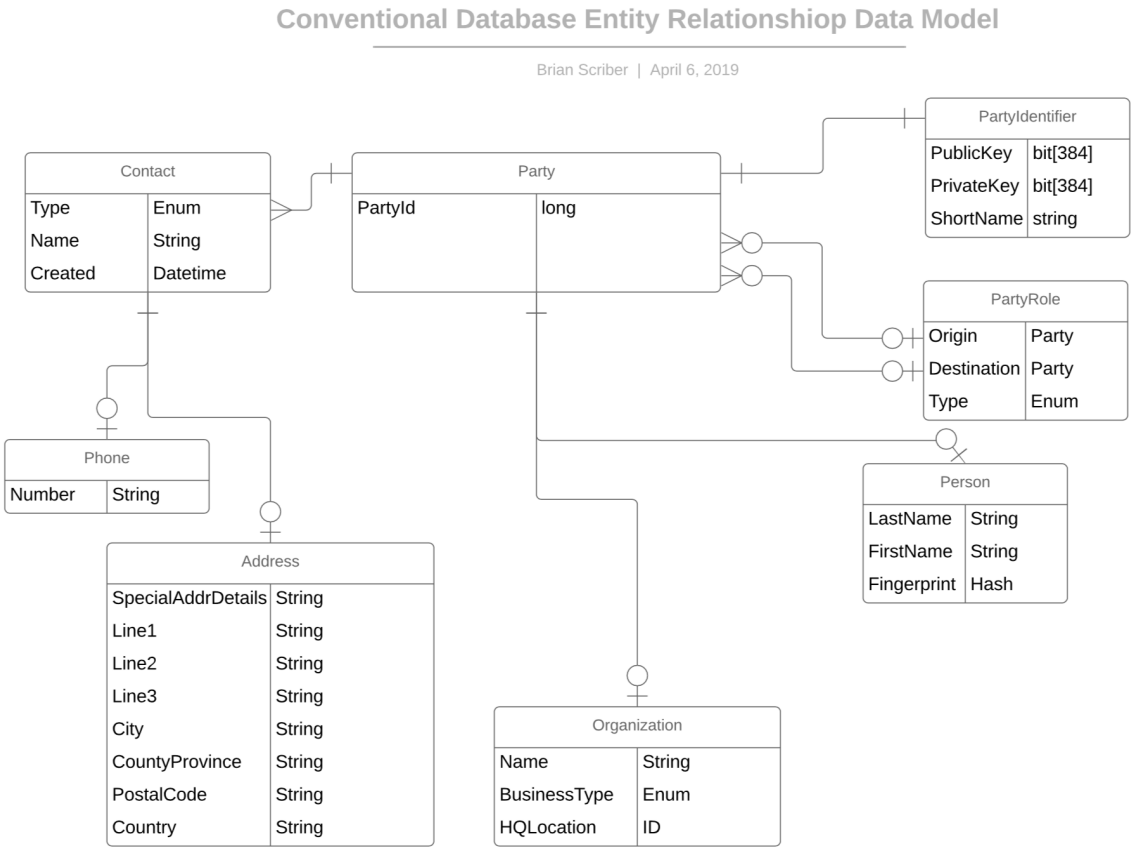

Figure 4 Conventional Database Entity Relationship Data Model

11. Adding Data Protection

Adding the elements described by the confluence of data protection regulations to the already existing data models means exploring how to add at least the following eight elements of metadata:

- Source (what is the provenance of this data?)
- Date Entered (when did it make it to this database?)
- Expiration (when can this data no longer be used?)
- Allowed Uses (for what can this data be used?)
- Sensitivity (what protections must exist and what authorization is required for access to this data?)
- Explicit Consent (was this permission for use granted explicitly? When? By whom? In what mechanism?)
- Revocation (has the right to use this data been terminated?)
- Right to be Forgotten (can this data be completely removed? Remove all null references to these rows as well)

Reconciling that with existing data models can be problematic. To further reduce scope of the above already simple data model example, this analysis will focus exclusively on the Person table.

12. Data Model Impact of Adding Protection

The Person table from the prior diagram is now shown in the upper left hand corner of Figure 5 Extending the Data Model for Sensitive Data. The expansion of two of those three fields, LastName and FirstName make up the rest of the example. The "Fingerprint" aspect of the prior table is included to show that one of the greatest steps an enterprise can take is to ensure protection of sensitive data is to not collect or store sensitive data. The collection of biometric information about consumers is dangerous because an enterprise cannot issue a new fingerprint or offer fingerprint-monitoring-services free for a year to anyone impacted by a breech. It's simply safer to not collect or store some information.

No longer can LastName simply be a string. The need now is to create a StringProtectedDataElement and reference it from the Person table. The new table needs to be able to have the Id being referenced and the actual data element, but since we want to reuse this table for multiple lengths of strings, we may have to be creative about how to not end up with a sparsely populated column/field where only a small portion of the data we set aside for storage is actually used by the data going into it (e.g. if most first names are only five characters long on average, and most last names are an average of 14 characters long, then using the same table to store only those two fields could result in an average of nine characters of unused storage capacity which can be costly in its own right).

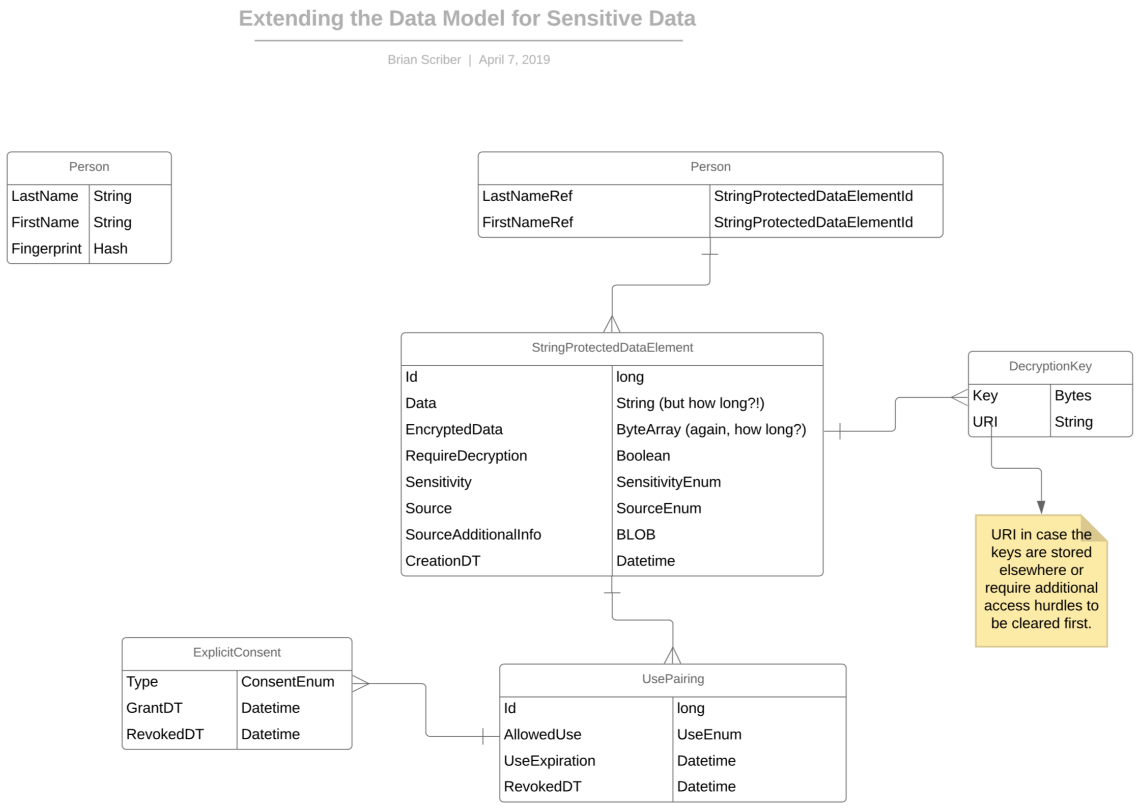

Figure 5 Extending the Data Model for Sensitive Data

This new table will also need to be able to store encrypted data if that data is designated as sensitive. Many databases have more efficient ways to deal with this than what is shown above, but the example here demonstrates that the requirement for encryption and the actual payload still need to be stored. Sensitivity may not be binary – the data may not be exclusively either sensitive or non-sensitive. There may be list of things used to qualify that sensitivity (e.g. "Top Secret", "Confidential", "Company-Only",

"Partners and Company", or even combinations like "Authorized Advertisers but Otherwise Confidential", etc.). The source for this data needs a similar enumeration (e.g. "Company A", "Employee Survey", "Data Aggregator B", etc.). It's not always clear what additional metadata might be required to help differentiate between sources, so a binary object has been created for this additional info. Each field from this string data also needs to have the date it was entered into this database, and a relationship to the allowed uses of the data, if those uses have been expired or revoked, and if there was an explicit consent provided. If there is explicit consent, there is now the capacity to know when it was granted and when it was revoked for each use (e.g. you can use my name for two years related to the bicycle contest, but you can only use my name for 3 months related to the ice cream flavor game). This data model gets complicated when one realizes that this is only for a tiny portion of only one of the databases for only the first and last name of the user.

Incurred Costs Related to Data Model Changes

13. Technical Transition

The technical transition will typically need to incorporate the aspects defined in Table 1 Technical Transition Costs. The right-most column in that table lists some of the typical roles that will be involved in the exercise of data management identified in the middle column. The three primary technical categories are the Data Model itself, the Software which directly accesses that data model, and then the Quality Assurance and Business Analysis required for each system. Project management expertise will likely be required at all phases of this work.

Table 1 Technical Transition Costs

Data Model		
	New Data Model	Data Modelers, Database Administrators, Application Engineers, Architects
	New Database Tables	Data Modelers, Database Administrators
	New Object-Relational Mappings	Database Administrators, Application Engineers, Architects
Software		
	New Allowed Use Checks	Software Design Engineers, Application Engineers, Architects
	Copying and Caching Validation	Application Engineers
	Integration APIs to be Updated	Systems Architects, Systems Engineers
Quality Assurance & Business Analysis For Each Group		

Review Requirements	Business Analyst, Quality Assurance (QA) Engineer, Architect, Systems Architect
Test/Confirm Privacy Requirements Met	Business Analyst, QA Engineer
New Bugs and Bug Fixes	Application Engineer, QA Engineer, Configuration Management Engineer

14. Operational/Procedural Transition

The technical transition is not the only part of a project like this, the impact to the operational, procedural and legal aspects of the enterprise should not be underestimated. Again, project management involvement as well as the office of the Chief Privacy Officer (a requirement in the EU, but optional elsewhere) are presumed for all of the following activities laid out in Table 2 Operational/Procedural Transition Costs.

Table 2 Operational/Procedural Transition Costs

Procedures		
	Notify Downstream Users of New Data Contracts	Legal, Marketing, Application Engineers, Business Analysts
	Data Warehousing and Reporting Data Curation Requirements	Data Warehousing Architects, Data Reporting Engineers
	Limiting Access to DBAs, Operational Teams, Strict Access Controls	Security Engineers, Business Analysts, Legal
Legal		
	Revocation (primary versus secondary/tertiary users of data)	Legal, Business Analysts, Application Engineers, QA Engineers
	Audit (data model, access control, even with decryption keys requires trust that regulators may still question)	Legal, Compliance/Auditor, Audit Committee, Business Analyst
New Customer-Facing Obligations		
	Exporting Data	Data Engineers, Legal, Standards Compliance Engineers
	Right to be Forgotten	Business Analyst, Data Engineers, Legal, Standards Compliance Engineers

Data Provenance Reporting	Systems Architects, Data Engineers, Legal, Standards Compliance Engineers

15. Privacy Impact Areas

Across the enterprise, the privacy and compliance changes will touch many departments if not all of them. The following list identifies several that will need to be kept at the front of mind while engaging in a holistic compliance effort:

- Online Account and Support Systems
- Production Applications
- Network/Abuse Monitoring
- Legal and Risk Management
- Billing and Accounting
- Purchasing (Contracts)
- Data Warehousing
- Marketing and Sales
- 3rd Party Data Clients & Partnerships

Solution Options

There are several options available which can each address some of the concerns raised above. This paper presumes outright that a complete green-field implementation of all systems and integrations is cost-prohibitive. Given that, the following subsections each explore some of the pros and cons of different and progressive concepts. There are other options and other combinations that may make sense, but

16. Don't Collect Private Info in the First Place

This is likely the immediate direction many enterprises will wish to explore. How to minimize data collection and use/store only the bare minimum. In this model they will still need to address the privacy implications of the data they do collect.

17. Modify the Exiting Data Models

The data model sections above go into the impact of what is likely to feel is the compromise soluton of updating the data models of a few key systems, but the costs associated with this were explored in depth, above. This noted, there may not be easy shortcuts for some of the complications imposed by the privacy landscape.

18. Add Separate Data Model for Privacy

Looking at the data, it might appear that some of the data can be extracted into a privacy model, but even doing this would require the integration points that were highlighted in the "Modify the Existing Data Models" approach defined above.

19. Restrict Access to a Centralized Data Store

This approach puts all the privacy-related data in a single place, but the issue here is that now there would exixt a single point of failure for the entire enterprise and the risk of shutting down all operations while any delay or issue with this core system was fixed is likely prohibitive.

20. Restrict Access but in a Decentralized Store with User Data Self-Sovereignty

User data self-sovereignty is a relative new concept in the Privacy Enhancing Technologies (PET) space. The idea is that the owner of the data controls who has access to their information. All data for all systems is encrypted and specific grants are given to different systems who can then access the same decentralized network to request access. The result of the requests are pointers to where the data resides, and the request also enlists aid from the network in decrypting that data pending the requestor authorization. Participant nodes are all independent, but the data stores may have single points of failure for any single element of data.

The data self-sovereignty options are built around the concepts of using some of the tools available to us from the security space and apply those to the privacy domain. These tools include:

- Encryption
- Databases
- Protected File Systems
- Audit Logging
- Indelible Ledgers
- Hashing
- Smart Contracts
- Trusted Execution Environments (TEE)
- Cryptographic Grants for Information Use

The TEEs and Smart Contracts open the door for:

- Mutual Trust Environments
- Audited Transactions *
- Authenticated Access: Read/Write/Create/Delete/Update/Notify
- Grants for Information Use (time-bounded, use restricted, access controlled)

Distributed Ledgers Add:

- Indelible Transactions
- Distributed Access
- Potential to Store Critical Protected Data in a Manner Accessible by
- Byzantine Fault Tolerance

Self Sovereignty Provides:

- Putting Protected Information into the Hands of the Owners of that Data
- Enables Explicit Consent, Access Grants, and Updates

Conclusion

The costs of compliance with global privacy initiatives can be far greater than some of the assumptions that are being made about how to achieve this compliance and about the penalties that are likely to be levied. The future is likely going to be complex during the transition period from our legacy data models to those which protect private information, but the long-term future has options like that at of data self-sovereignty and individuals owning their own data and using cryptography to protect it and grant appropriate access.

Abbreviations

CCPA	California Consumer Privacy Act (California, USA)
ERD	Entity Relationship Diagram
EU	European Union
GDPR	General Data Protection Regulation (European Union)
PIPEDA	Personal Information Protection and Electronic Documents Act (Canada)
QA	Quality Assurance
RTBF	Right to be Forgotten
RTD	Right to Deletion
TEE	Trusted Execution Environment

Bibliography & References

EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

EU Directive 95/46: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ 1995 L 281/31.

Papers of the Article 29 Working Party (example): Article 29 Working Party, 'Opinion 2/2003 on the application of the data protection principles to the WHOIS directories' (WP 76, 13 June 2003), at 4

Personal Information Protection and Electronics Document Act: Amended April 1, 2019 https://laws-lois.justice.gc.ca/PDF/P-8.6.pdf

EU Charter of Fundamental Rights: Charter of Fundamental Rights of the European Union, OJ 2010 C 83/389.

European Convention on Human Rights: European Convention for the Protection of Human Rights and Fundamental Freedoms, Sept. 3, 1953, ETS 5, 213 UNTS 221.

Judgments of the Court of Justice of the EU (example): Bodil Lindqvist, Case C-101/01, [2003] ECR I-12971 (ECLI:EU:C:2003:596), at para. 74.

Council of Europe Convention 108: Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data 1981, ETS 108

NIST Privacy Framework: An Enterprise Risk management Tool: National Institute of Standards and Technology, US Departent of Commerce, April 30, 2019, https://www.nist.gov/sites/default/files/documents/2019/04/30/nist-privacy-framework-discussion-draft.pdf

[i] https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

[ii] https://www.clarip.com/data-privacy/california-consumer-privacy-act-fines/

