Privacy, consent, trust and security within modern applications
Open Source Summit North America – August 2019

Tim Mackey – Principal Security Strategist – Synopsys Cybersecurity Research Center
Twitter: @TimInTech
Disclaimer

IMPORTANT NOTICE
The information provided in this presentation should not be considered legal advice. Navigating the complexities and nuance of global regulations requires specialized knowledge which can not be fully imparted in a single presentation or setting.

This presentation seeks to provide a set of criteria which software development teams can use to more readily identify areas of data management which could create privacy issues. Once those areas are identified, a discussion should occur between the software development team, corporate legal, and the production operations team on how best to protect the identified data consistent with the corporate privacy policy.
The landscape
Being a security target is costly

Average total cost of data breach: $8.19 Million
Customer impact: 3.6% abnormal turnover
Average time to identify and contain a breach: 245 days

Source: 2019 Cost of Data Breach Study (US Data) – Ponemon Institute
Current crop of major data privacy regulations

**** AG Investigatiins *****

• EU - General Data Protection Regulation (GDPR)
  – Enacted in 2012, EU enforcement on May 25, 2018 and EEA enforcement July 20, 2018
  – Addresses data collection, processing and privacy implications for data on EU residents
• Canada - Personal Information Protection and Electronic Documents Act (PIPEDA)
  – Enacted in 2001 to regulate electronic commerce and align Canadian Privacy legislation with that of the EU
  – Amended June 2015 to address how private sector organizations collect, use or disclose personal information – unless all data operations are on subjects with a province with “substantially similar” privacy legislation
• Australia - Notifiable Data Breaches (NDB)
  – Enacted as 2017 Amendment to Privacy Act of 1988 to address breach disclosure requirements, effective Feb 2018
• India – Personal Data Protection Bill 2018
  – Draft privacy legislation designed to address global trade tensions
  – Contains data sovereignty provisions
• California – California Consumer Privacy Act (CCPA)
  – Enacted September 2018 with a nominal enforcement date of Jan 1, 2020 (likely mid-2020 subject to amendments)
  – Goal to provide residents with rights of disclosure around data operations
Equifax breach focused attention on open source risk
Truism #1

“

You can’t secure data you don’t know you’re processing

”
So what is “data” anyway?

Note – Lots of legal nuance here so don’t take this as legal advice!

• Regulatory Definition (HIPPA Example)
  – Protected health information is information, including demographic information, which relates to:
    – the individual’s past, present, or future physical or mental health or condition,
    – the provision of health care to the individual, or
    – the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual. Protected health information includes many common identifiers (e.g., name, address, birth date, Social Security Number) when they can be associated with the health information listed above.

• Common Definition
  – Data is any information about myself which is provided to an organization in order to render a service.
What about “data privacy” then?

Note – Lots of legal nuance here so don’t take this as legal advice!

• Industry Definition – Privacy Principles (ISO 29000:2011 Example)
  – A set of shared values governing the privacy protection of personally identifiable information (PII) when processed in information and communication technology systems

• Legal Definition – No common definition
  – In Canada, the right to privacy is rooted in the Canadian Charter of Rights and Freedoms. The Charter does not specifically mention privacy or the protection of personal information. However, it does afford protection under Section 7 (the right to life, liberty and the security of the person), and Section 8 (the right to be secure against unreasonable search or seizure).

• Common Definition
  – Data privacy is an expectation I have that the data I’ve provided to anyone will be properly protected.

Note – Lots of legal nuance here so don’t take this as legal advice!
Example: New user with email identifier

Is the email address trusted? What data is being bled if invalid?
Example: Assumption of valid email

*Increased risk exposure to multiple parties*

1. Identifies user as customer
2. Provides location of user
3. Provides name of user
4. Facilitates impersonation of user
5. Provides window of opportunity
Truism #2

“If your users don’t know what you’re doing with their data – you increase reputational risk if something goes wrong!”
“Consent” is a tricky concept

Note – Lots of legal nuance here so don’t take this as legal advice!

• Regulatory Definition (GDPR Recital 42 Example)
  – Where processing is based on the data subject’s consent, the controller should be able to demonstrate that the data subject has given consent to the processing operation.
    – In particular in the context of a written declaration on another matter, safeguards should ensure that the data subject is aware of the fact that and the extent to which consent is given.
    – In accordance with Council Directive 93/13/EEC a declaration of consent pre-formulated by the controller should be provided in an intelligible and easily accessible form, using clear and plain language and it should not contain unfair terms.
    – For consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended.
    – Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment.

• Common Definition
  – Before I consent to anything, I need to know what you’re going to do with my data, who you share it with, and how long the data is going to be around. Oh, and I can change my mind later you know!
“Trust” is the most complex of the concepts

• Reputational Risk (Public Relations Society of America)
  – Reputation risk is the risk of losses that might result from damage to an organization’s reputation, such as lost revenue; increased operating, capital or regulatory costs; or lost shareholder value (decreased stock price). Many types of events can impact an organization’s reputation and cause these losses, not all of which are the organization’s own fault.

• Brand Equity (American Marketing Association)
  – From a consumer perspective, brand equity is based on consumer attitudes about positive brand attributes and favorable consequences of brand use

• Common Definition
  – I won’t give my data to just anyone. If the provider is one whose brand I trust, or one which solves a real problem for me, I’m more likely to give personal data to them.
Top 12 questions when managing collected data

Goal is to minimize data collected and maximize customer value

1. Does the customer explicitly know the data is being collected?
2. Are they clear on how the data is being used in your service?
3. If the data is sent to any third party, is the customer clear on where it might be?
4. Will the customer need to opt-in or opt-out of data usage or collection?
5. Who internally could have access to the data, and why?
6. Does the data require any specialized processing (e.g. encryption or hashes)
7. How would you know if someone accessed the data?
8. Have you informed the customer how long the data is retained for?
9. If regulations allow users access to their data, what is the process?
10. If regulations allow users to delete or correct their data, what is the process?
11. What data is being transferred as part of “phone home” systems?
12. If web services are used, how is transmitted data handled?
GDPR actions not just for data breaches
Truism #3

“When a data incident occurs – the only data exfiltrated is data you retained"
Reputational damage factors

- M&A Goodwill impairments
- SEC Filing requirements
- Churn of major accounts
- Bankruptcy reorganization
- Regulatory fines
- Supply chain impact
- Increased costs of customer acquisition
Truism #4

“As applications evolve, original decisions around data collection become opaque resulting in increased potential leakage.”
Anonymizing shared data doesn’t guarantee privacy

Current methods for anonymising data leave individuals at risk of being re-identified, according to new UCLouvain and Imperial research.
Truism #5

“Given access to any data, people will find a way to use, and potentially misuse it.”
Web service APIs change risk dynamic

**API Lifecycle**
- Twitter API shutdown August 2018
- Google+ shutdown April 2019
- Salesforce API versioning

**Data usage and control**
- GDPR data processor vs data controller
- Data sovereignty and jurisdiction
- Data mashups and inference scenarios

**Data and privacy breaches**
- Facebook API tokens
- [24]7.io and Delta, Kmart, Sears
- Third-party data bleeds
- Phone home tracking
Managing consent can be complicated

- Multiple people
- Software updates
- Mobile devices
- Changes in jurisdiction
- Right to withdraw
- Right to access
Dissecting the privacy and security design decisions of modern applications

Example: An IoT application is more than just firmware
Modern application = Proprietary Code + Open Source Components + API Usage + Application Behavior and Configuration
IoT development requires multiple disciplines

- **Limited CPU resources**
- **Limited RAM for features**
- **C/C++ typical**
- **MQTT common protocol**

- **Responsive application**
- **View device data**
- **View historical information**

**MQTT Broker**
- Lightweight protocol
- High volume
- Pub/Sub interface

**OTA**
- Avoid MITM
- Certification of image

**IoT Device**
- Limited CPU resources
- Limited RAM for features
- C/C++ typical
- MQTT common protocol

**Core Data**

**Analysis Engine**
- **Authentication and Authorization**
- **Analysis Engine MQTT WebSocket**

**Web UI**

**Mobile Interface**
- iOS/Android application
- Configure device
- View device data
- Receive notifications

**Data stored for analysis**

**Encrypt data published via MQTT**

**Configure via Bluetooth**

represents constraints in the system
Identify security targets from platform requirements

Role: Security Architect with CISO and Product Owner guidance

Design Goal:
Select an IoT toolchain meeting product and cost requirements

Privacy Concern:
Device instability leading to data disclosure and reputational damage

Tasks and requirements:
1. Select platform supporting desired protocols
   • Protocol implementations must be resilient
2. Select candidate vendor or open source stack
3. Validate protocols against cost and stability
   • Define protocol fuzzing framework
4. Report on security targets during development

Privacy Concern:
Device instability leading to data disclosure and reputational damage
Select development frameworks and environment

Role: Development Lead with Product Owner guidance

Design Goal:
Select frameworks capable of meeting time to market and security targets

Privacy Concern:
Architectures which rely upon external data processing

Tasks and requirements

1. Select languages based on security reqs
2. Define build environment
3. Identify commercial and open source frameworks and libraries
   • Define governance for security updates
4. Enable IDE security plugins
5. Enable build time CI analysis
Continuous security assessments during development

Role: Developer with Development Lead guidance

Development Goal:
Identify security weaknesses prior to code commits

Privacy Concern:
Poor security training and limited review of data processing and collection activities

Tasks:

1. Transparent security review during coding
   - No disruption to existing workflows

2. Remediation and contextual guidance
   - Lower defect costs by shifting left
Continuous security assessments during build

Role: Release Engineer with guidance from QA and Product Owner

Release Goal:
Ensure release meets quality, security and functional targets

Privacy Concern:
Identify weak test coverage, limited security testing and poor data handling

Tasks and requirements:

1. Build triggered from merge/pull request
2. Detailed scans run parallel to build process
3. Optionally fail builds based on security targets/exceptions
4. Centralized security progress tracking
Confirm governance and security target progress

Role: Security Architect

Governance Goal:
Ensure release meets security and functional targets

Privacy Concern:
Identify and redress latent data management practices as applications evolve

Tasks:
1. Centralized review of security results
2. Review by common taxonomy
   - (OWASP Top 10, SANS Top 25)
3. Triage issue status via defect workflows
4. Measure progress against governance targets
5. Define security targets for future releases
**Key open source centric takeaways**

**Data security starts with informed project contributors**
- Question why any data is collected, how its processed, and how long will it be retained
- Train all reviewers and developers to identify sensitive data usage or leakage
- Disclose in docs/pubs data collection and processing actions – don’t assume consumers will read code

**Data governance requires cooperation between Devs and project leadership**
- Don’t merge with LGTM reviews where data is collected or manipulated
- Don’t assume component used in a single regulatory jurisdiction
- Document all decisions impacting risk assessments surrounding data usage

**Manage risk impact of legacy components with forward facing actions**
- Define data governance policy for project, and update README stating minimum version
- Review default configurations and disable any allowing arbitrary data collection or processing
- Document data management policies for versions in “maintenance mode”
References

- https://gdpr-info.eu/
- https://laws-lois.justice.gc.ca/eng/acts/P-8.6/
- https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375
- https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html
- https://gdpr-info.eu/recitals/no-42/
- https://www.datascience.com/blog/what-is-a-churn-analysis-and-why-is-it-valuable-for-business
- https://aytm.com/blog/brand-equity-basics-1/
Thank You