Data Centric Security for Personal Information

Chad Childers
IT Security & Strategy
Ford Motor Company
Overview

- Background
  - 1950’s Thinking vs. 2010’s Reality
  - Data Centric Security
- The Personal Information Landscape
- Implementations for different risk levels
- Conclusions
- References / PKI / Cloud Security
1950’s Thinking vs. 2010’s Reality

- **Traditional Thinking**
  - Monolithic government/corporate databases
  - Internal users and administrators
  - Stable business models with slow changes

- **Reality**
  - Mix of contractor, offshore, outsourced, cloud, and alliance data sharing
  - Sale of personal information for marketing purposes
  - Rapid Business Change or websites shut down
20th Century networks

- Firewalls
- Employees
- DMZ
- Standalone PCs
21st Century networks

- Deperimeterization
- Endpoint Security
- Consumerization
- Alliances
- Outsourcing
- Cloudsourcing
- Ubiquitous Computing
Data Centric Security vs...

- Network Security – trust everything on the network, trusted users, firewall controls
- Endpoint Security – untrusted network, trust and secure the host, application controls
- Data Centric Security – untrusted host, secure the data itself, metadata/honor system applied to endpoint applications or cryptographic controls
Data Centric Security theories

- Different levels of privacy based on multiple cryptographic keys
- Cryptographic theory work on performing bulk operations on data without access to decrypt that data
- Business drivers for data classification, labels (metadata) transmitted and external parties trusted to implement and enforce associated security requirements
Peanut Butter vs Risk Driven Security

- Applying all controls uniformly may have worked in simpler times
- Apply controls based on business drivers, risk assessment, CIA, or individual preference
- Allow affected individuals to choose their own security level
- Different security for different sets of users
Risk Driven Security
CIA application classification

- Confidentiality
  - Data classification
  - Information disclosure

- Integrity
  - Dollar value
  - Repudiation

- Availability
  - Service Level Agreement
  - Denial of Service
Business Change

**First Virtual**
- 1994 Founded by Borenstein et al for online payments, predecessor to PayPal
  - Ann Arbor office - MIME protocols, insecurity of credit card information via the Internet
- 1998 Direct Marketing Email (CRM or spam) becomes primary business, renamed MessageMedia
- 2001 Sold to DoubleClick

**Firefly**
- 1995 Founded by MIT Media Lab as a social network specializing in music recommendation, predecessor to Facebook, Pandora, movie recommendation paradigm from Netflix
- Creators of Open Profiling, which became the W3C Platform for Privacy Preferences (P3P)
- 1998 Sold to Microsoft
- 1999 Shut down, only sign-on technology kept (Passport)
Business Change 2

- **LiveJournal**
  - 1999 Founded in US
  - 2005 Sold to SixApart in Japan
  - 2007 Sold to SUP in Russia

- **Netflix**
  - 2010 Removal of Friends/Community
  - 2011 Divestiture of DVD business, removal of “saved queue” data
Sale of Personal Information

- 1993 Drivers Privacy Protection Act in response to 1983-9 actress stalker attack/murder via access to DMV records, bans state sale of PII
- 2007 Florida DMV sold PII of 10M drivers to mass marketers, sued under DPPA, settled for $10M
- 2010 Florida makes $63M per year selling PII to Shadowsoft, West, etc
- 2010 Missouri DOR sold PII to Shadowsoft, which sold the information through publicdata.com, settled for return of information
- 2011 West Publishing DPPA suit for collecting and
1988 Video Privacy Protection Act in response to 1987 Supreme Court nominee Bork’s rental records being given to a newspaper

2007 Facebook Beacon advertising system launch

2008 Blockbuster sued under VPPA for giving rental information to Facebook Beacon

2009 Netflix sued under VPPA for improper use in Netflix Prize competition

2011 Facebook offers list of friends Netflix viewing activity in 44 countries, pushes for repeal of VPPA
Data Centric Security levels

- Metadata only – trust all infrastructure and systems
- Encryption only – trust key holders, applications and hosts
- Multi-level encryption – granular levels of trust
- Smart encryption/package – trust paradigm and initial entry application
- Distributed data – trust data owner’s local infrastructure, brokering application
Metadata only

- Interoperability
  - Programs can read metadata properties
  - Vocabularies can evolve and be extended
- Add CIA vocabulary
- Gaps/attacks?
  - Must trust all parties to comply
  - Insufficient for CI security assurance
  - Network and Endpoint security are still critical, not truly data centric security, but a necessary first step
- ccREL: The Creative Commons Rights Expression Language
- IBM Data Centric Security Model
Encryption only

- Payload encrypted with symmetric key
- Symmetric key encrypted with public key(s) to allow multiple recipients
- Could be extended
  - Add CIA vocabulary
  - Require encryption strength based on Confidentiality
- Gaps/attacks?
  - Must trust applications/users to modify and re-encrypt data
  - Sufficient in transit
  - Insufficient for Integrity assurance
  - Brute force or crypto algorithm
- S/MIME encrypted container
Multi-Level Encryption

- Payload contains separate sub-packages encrypted at different levels
- Different levels encrypted with different keys
- Allows very granular control of Confidentiality
- May also be perfectly suited for Web of Trust implementations
- Gaps/attacks the same as encryption
- pDCS
Web of Trust

- PKI public key signing is by a paid central CA
- Web of Trust public key signing is only done in person, by someone you know and trust
- Applying this concept to Data Centric Security for Personal Data crypto
  - Allow encrypted, multi-level, or smart data packages to only be decrypted by individuals or entities who have been given authority in person (fully trusted)
  - Allow multi-level or smart data packages to give certain data or de-identified data to partially trusted (level 2 or below) individuals or entities
Smart Encrypted Package

- Allow operations on data without allowing actual access to data – potential approaches:
  - Advanced encryption algorithm
  - One way hash of source data to allow specified operations
  - Code to perform data operations within encrypted container

- Potential extensions
  - Provide de-identified data
  - Validate access with owner
  - Some use cases may be allowed without application validation
  - Logging may be internal to code

- Gaps/attacks?
  - Brute force
  - Crypto algorithms unproven
  - API implementations unproven
  - Some operations may be intrinsically unsafe

- Virtual Smart Card
Smart Encryption Module

- Split into use case broker module (agent API) and encrypted data package
- Existing DRM solutions work for viewing use case
- Require network connection to validate access with data owner

Potential extensions
- Validate applications by code signing
- Add use cases in application code
- Require logging of changes within container or by owner

Gaps/attacks?
- Man in the middle
- Control of data owner permission server
- Permission mechanism implementation
- Broker module must be trusted and executable could be attacked
- Brute force or crypto algorithm
Distributed Data

- Code to perform data operations does not contain data – potential approaches:
  - Data owner retains data, trusted local broker sends only extract or one-way hash needed to perform operations to trusted server broker
  - Data owner deals with only trusted clearinghouse

- Gaps/attacks?
  - Bad for Availability, speed/latency – who cares if Social Media has to wait!
  - Attacks on clearinghouse
  - Local attacks easier – but would never expose millions of records at a time!
Distributed Data 2

- If private individual is defined as owner of data
  - Allows owner to make data updates across all data users quickly and reliably
  - Allows owner to change personal or customer relationship access rights
  - Allows owner to be alerted and make decisions for access over a given level of Confidentiality
  - Allows owner to require payment for each marketing sale access of data
- Tax return preparation software
- Data Owner Data Brokering
Data Ownership

- In some countries, an individual may be defined as owner of their PII – is there a lawyer in the house?

- If an individual spends hundreds of hours creating data, do they then have some right to that data?
  - Blogging or photography
  - Social media relationship data
  - Lists, favorite movies, songs
  - Address books
Conclusions

- The profit motive drives creation of silos, which have intrinsic limits and weaknesses.
- A data centric paradigm could be a complete game-changer, like the birth of the web – removes intrinsic limits.
- Distributed/smart levels are not the first step, and for some applications, they aren’t the right level.
- Next: how can specific level implementations improve the Personal Information Landscape.
Apply the right controls
At the right time
For the right cost
Top 7 security concerns in the Cloud

*Cloud Security Alliance*

- Abuse & Nefarious Use
- Insecure Interfaces and APIs
- Malicious Insiders
- Shared Technology Issues
- Data Loss or Leakage
- Account or Service Hijacking
- Unknown Risk Profile
“Cloud Service Risk”- Dave Cullinane – eBay
VP/CISO, Cloud Security Alliance Director

- Our function: anticipate, understand, act
  - Find out where the business is going
  - Inform the business of the type of risk in the situation
- Only 30% evaluate risk first
- Is there a plan to protect customer information
- 50% of eBay security problems from micro projects, slip by without risk evaluation
Stormy Weather

- You pay more for trust ability, risk transparency.
- Characteristics indicative of high risk:
  - Accessibility
  - Complexity
  - Extensibility
- Confidentiality vs Availability
  - Cannot be simultaneously optimized, what matters to business?
- All clouds are not created equal… too ambiguous.
  - Only pretty babies are a Cloud
PKI = Public Key Infrastructure

- public key: n and e
- private key: d

Euclid’s algorithm

\[ ed \equiv 1 \mod (p - 1)(q - 1) \]
\[ d = e^{-1} \mod ((p - 1)(q - 1)) \]

- Large primes p and q are destroyed
- Public key is then signed by CA
- Factoring primes is very hard!
References

- pDCS: Security and Privacy Support for Data-Centric Sensor Networks, Min Shao et al, Pennsylvania State University
- Elevating the Discussion on Security Management - The Data Centric Paradigm, Tyrone Grandison et al, IBM Research
- ccREL: The Creative Commons Rights Expression Language
  [http://wiki.creativecommons.org/CcREL](http://wiki.creativecommons.org/CcREL)
- Towards the Australian Data Commons
  [https://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf](https://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf)
- Cloud Security Alliance
  [http://www.cloudsecurityalliance.org](http://www.cloudsecurityalliance.org)
- Photograph © Andrew Dunn, licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license.
- Privacy Rights and Data Security Compliance: Legal and Business Best Practices, Brian Balow, SecureWorld