On Deploying Property-Preserving Encryption

Paul Grubbs
Cornell University/Skyhigh Networks
Outline

Look at applications of property-preserving encryption (PPE)

Discuss gaps in understanding of how PPE is used

Open problems + Motivate further work

Disclaimer: I am still a consultant for SHN

My opinions are my own
Business Software

It used to be...

BusinessCo

“On-premise”

Now it’s becoming...

“Customer relationship management” and our case study for the talk

Most of you know at least one of these:

salesforce  service now
Ms. Business uses Salesforce

Load customer data

Get all customers with first name Alice

Get customers with >$1,000,000 value

Keyword search

<table>
<thead>
<tr>
<th>Accounts</th>
<th>Customer</th>
<th>Zip</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alice Cooper</td>
<td>60652</td>
<td>500,000</td>
</tr>
<tr>
<td></td>
<td>Bob Jones</td>
<td>46032</td>
<td>1,600,000</td>
</tr>
<tr>
<td></td>
<td>Alice Zandra</td>
<td>95014</td>
<td>1,200,000</td>
</tr>
</tbody>
</table>

Numerical comparisons
We need to use encryption for Salesforce now.

Data residency laws

Consumer privacy laws

Voluntary (security-minded CIO/CISO)

Industry regulations
Ms. Business uses Salesforce, with encryption

Load customer data

Get encryption key from BusinessCo

Empire Encryption

Get all customers w/ name ‘Alice’

Load encrypted data

Design goals:
- Maximize Security
- Functionality
- Minimize cost

Cost of solution + retraining thousands of users

<table>
<thead>
<tr>
<th>AcctName</th>
<th>Zip</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a7f45edbc</td>
<td>94521</td>
<td>95734857</td>
</tr>
<tr>
<td>94dabc467</td>
<td>12379</td>
<td>97563543</td>
</tr>
<tr>
<td>1273548fd</td>
<td>40378</td>
<td>96784657</td>
</tr>
</tbody>
</table>
Design spectrum of encryption proxies

Use as much Salesforce functionality as possible

Property-preserving encryption (PPE)

Increasing cost of proxy

Standard encryption

Re-implement needed functionality locally (in the proxy)
Deep dive into keyword search + encryption
Keyword search on text fields

Get all customers with first name Alice

Alice Cooper, Alice Zandra

Accounts

<table>
<thead>
<tr>
<th>CustName</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice Cooper</td>
<td>60652</td>
</tr>
<tr>
<td>Bob Jones</td>
<td>46032</td>
</tr>
<tr>
<td>Alice Zandra</td>
<td>95014</td>
</tr>
</tbody>
</table>
Standard industry solution

Get all customers with first name Alice

Alice Cooper, Alice Zandra

Encryption Proxy

Get all customers with first name Fhbruyf

Fhbruyf b47394, Fhbruyf djdlvldl8...

Accounts

<table>
<thead>
<tr>
<th>CustName</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fhbruyf b47394</td>
<td>95421</td>
</tr>
<tr>
<td>Hdiel d849g9</td>
<td>16478</td>
</tr>
<tr>
<td>Fhbruyf djdlvldl8</td>
<td>94738</td>
</tr>
</tbody>
</table>

Deterministic encryption: word-by-word, length-preserving

We wanted to do better, so we turned to academic research on searchable symmetric encryption

Enables keyword and phrase queries with no overhead but security is problematic.
Searchable symmetric encryption (academic abstraction)

- Store encrypted search index for some documents
- Encrypted search token for “Alice”
- Doc IDs 1, 3
- Get docs 1, 3
- Encryptions of 1, 3

Trusted client

Untrusted server

Encrypted index

Encrypted documents

[CJJJKRS’14]: simple, parallelizable, scalable, handles updates
Searchable symmetric encryption (our deployment)

Client (encryption proxy)

Store encrypted search index for some documents

Encrypted search token for “Alice”

Doc IDs 1, 3

Encrypted index

Get docs 1, 3

Encryptions of 1, 3

Both the client and index are hosted by SHN, only documents are on Salesforce
Complexities in SSE deployment

• Threat model is different
  – SHN stores index, not Salesforce
  – Still valuable to protect against compromise
    • Theft of hard disk vs. penetration of software
    • Regulation is concerned with ‘data residency’

• A *lot* of engineering effort
  – Geo-replicated multi-tenant Cassandra clusters
  – ~1 person-year of work
  – 60-ish % of engineering : updates
  – Potentially dozens of large (160 million objects) customers
  – *Roughly 31 updates per millisecond per customer*

• Open questions:
  – Stateless dynamic SSE *or* state that doesn’t need synchronization
    • Hard to get needed throughput for updates with synchronization
  – No preprocessing/indexing stage (no static index)
  – Security?
Deep dive into range queries + encryption
Range queries

Get all customers with >$1,000,000 value

Bob Jones, Alice Zandra

Accounts

<table>
<thead>
<tr>
<th>AcctName</th>
<th>Zip</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice Cooper</td>
<td>60652</td>
<td>500,000</td>
</tr>
<tr>
<td>Bob Jones</td>
<td>46032</td>
<td>1,600,000</td>
</tr>
<tr>
<td>Alice Zandra</td>
<td>95014</td>
<td>1,200,000</td>
</tr>
</tbody>
</table>
Encrypted range queries

Get all customers with >$1,000,000 value

Bob Jones, Alice Zandra

Encryption Proxy

Make report on >$96000478 value

Use order-preserving encryption (OPE)

Accounts

<table>
<thead>
<tr>
<th>AcctName</th>
<th>Zip</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a7f45edbc</td>
<td>94521</td>
<td>95734857</td>
</tr>
<tr>
<td>94dabc467</td>
<td>12379</td>
<td>97563543</td>
</tr>
<tr>
<td>1273548fd</td>
<td>40378</td>
<td>96784657</td>
</tr>
</tbody>
</table>
Two kinds of OPE

• Stateless OPE [BCLO ‘09]
  – Deterministic, fast(ish)
  – Ciphertexts 3 bits longer than plaintexts
  – Unclear security

• Interactive OPE [PLZ ‘13] [KS ‘14] [K ‘15]
  – Proxy must store state (‘stateful’)
  – Other ciphertexts change with insertions (‘mutable’)


Complexities in OPE deployment

• Interactive is non-starter
  – Global, synchronized state
  – Implementing correctly: person-years of effort for unsure performance
  – Mutability requires additional complexity & custom code, so increased attack surface
• Stateless OPE easier, but still
  – Fixed domain size
  – Efficiency (needed some creativity to make fast)
    • CryptDB: 25-50ms
    • SHN: 2-3ms
• Active attacks possible (“marketing automation CPA”)
• Open questions:
  – Domain extension for OPE
  – Trade security for strict order
  – Security? (Next talk!)
**Recent leakage-abuse attacks on PPE**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Techniques</th>
<th>Recovery Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKK12</td>
<td>Searchable encryption</td>
<td>Query recovery</td>
</tr>
<tr>
<td>CGPR15</td>
<td>Searchable encryption</td>
<td>Partial message recovery</td>
</tr>
<tr>
<td>NKW15</td>
<td>FPE, OPE</td>
<td>Plaintext recovery</td>
</tr>
</tbody>
</table>

**Punchline:** PPE can be badly broken in some settings
Hold on a sec...

Crypto researcher on PPE

PPE is OVER!

Use as much Salesforce functionality as possible
Use as much Salesforce functionality as possible

Property-preserving encryption (PPE)
Property-preserving encryption (PPE)

Standard encryption
Standard encryption

Cost of using encryption increases greatly
Cost of using encryption increases greatly

Re-implement needed functionality locally (in the proxy)
Re-implement needed functionality locally (in the proxy)
Role of researchers?

- Settings where people want to use PPE
- Settings where PPE use is ‘ok’

Researchers can help find this intersection, guide decision-making about tradeoffs
Conclusion

• PPE is deployed and used
• PPE use will continue to grow
• Interesting opportunity for researchers to have *real-world* impact
  – Tons of cool open problems!!!
Thanks for listening!
Questions?

Special thanks to Hani Dawoud, Steve Myers, Tom Ristenpart, and the team at SHN