On Deploying Property-Preserving Encryption

Paul Grubbs

Cornell University/Skyhigh Networks

Outline

Encryption that trades leakage for functionality

Look at applications of property-preserving encryption (PPE)

Discuss gaps in understanding of how PPE is used

Open problems + Motivate further work

Disclaimer:

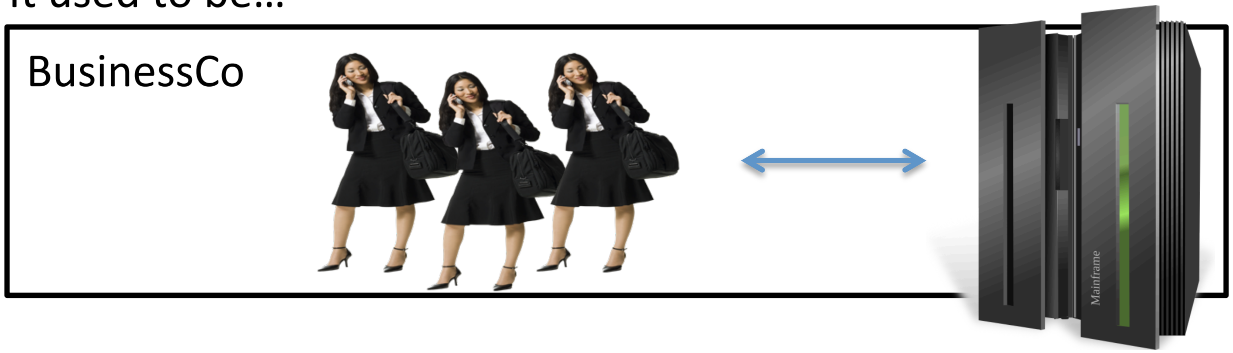
Former employee of Skyhigh Networks (SHN)

I am still a consultant for SHN

<u>My opinions are my own</u>

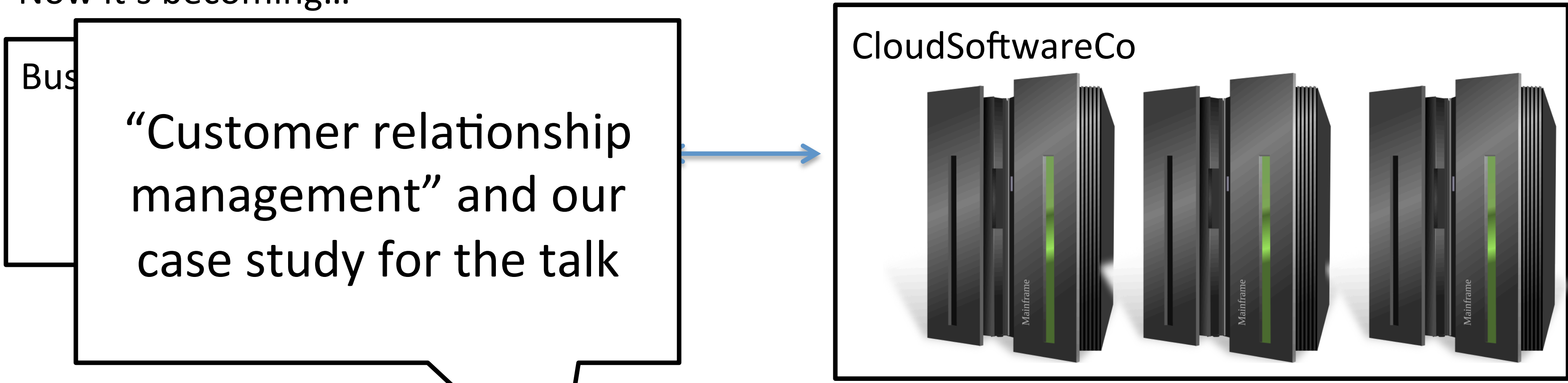
Business Software

It used to be…



"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 servicenow jive box

Ms. Business uses Salesforce

Keyword search

salesforce

Load customer data

Get all customers w/ first name Alice

Get customers with >$1,000,000 value

Numerical comparisons

Accounts

Customer	Zip	Value
Alice Cooper	60652	500,000
Bob Jones	46032	1,600,000
Alice Zandra	95014	1,200,000

A change

We need to use encryption for Salesforce now.

Data residency laws

Consumer privacy laws

HIPAA COMPLIANCE

Voluntary (security-minded CIO/CISO)

Industry regulations

PCI DSS COMPLIANT

Ms. Business uses Salesforce, with encryption

Get encryption key from BusinessCo

salesforce

Load customer data

Load encrypted data

Encryption Proxy

???

Get all customers w/ name 'Alice'

Accounts

AcctName	Zip	Value
a7f45edbc	94521	95734857
94dabc467	12379	97563543
1273548fd	40378	96784657

How does the proxy satisfy queries (search, report generation, etc.) on data? At scale?

Design goals:
Maximize
- Security
- Functionality

Minimize cost

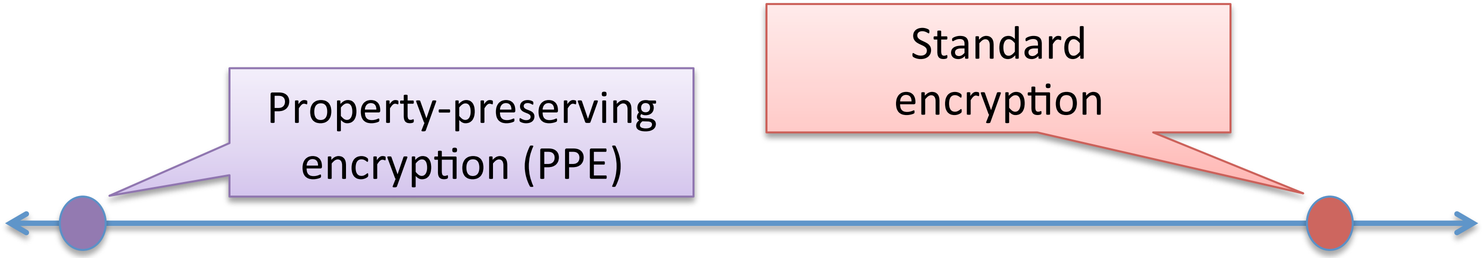
Cost of solution + retraining thousands of users

Perspecsys
Making the Public Cloud Private

skyhigh

CipherCloud™
Building Trust in the Cloud

Design spectrum of encryption proxies



Property-preserving encryption (PPE)

Standard encryption

Use as much Salesforce functionality as possible

Increasing cost of proxy

Re-implement needed functionality locally (in the proxy)

Deep dive into keyword search + encryption

Keyword search on text fields

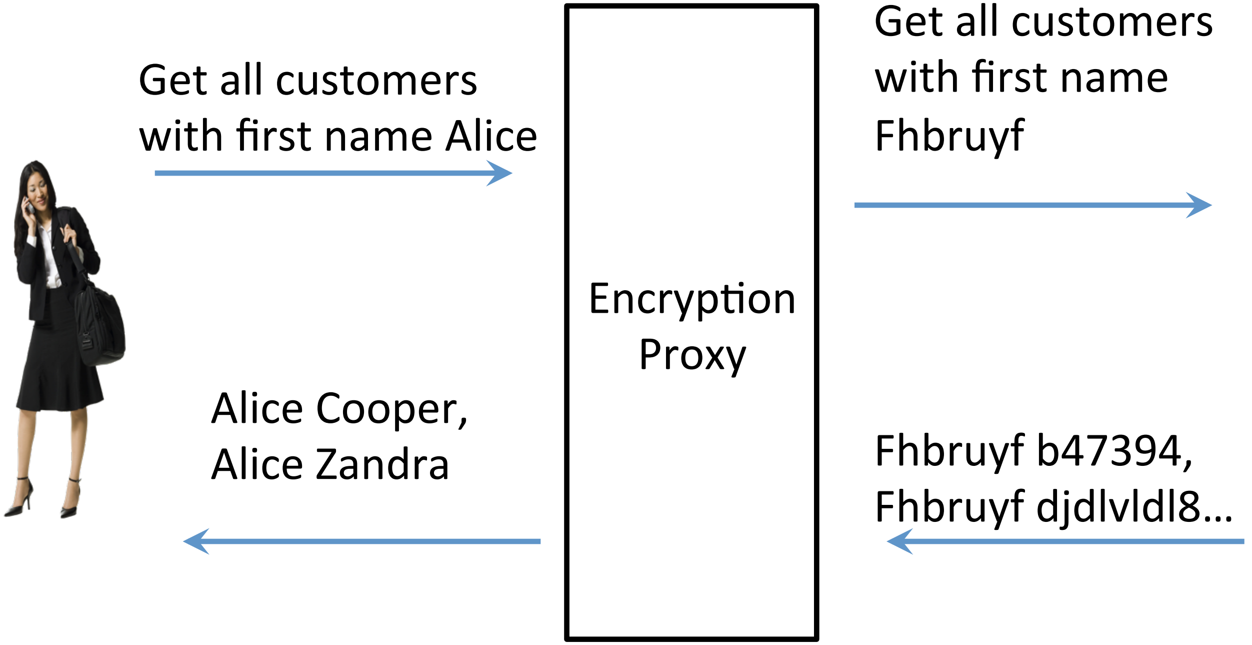
Get all customers with first name Alice

Alice Cooper, Alice Zandra

Accounts

CustName	Zip
Alice Cooper	60652
Bob Jones	46032
Alice Zandra	95014

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

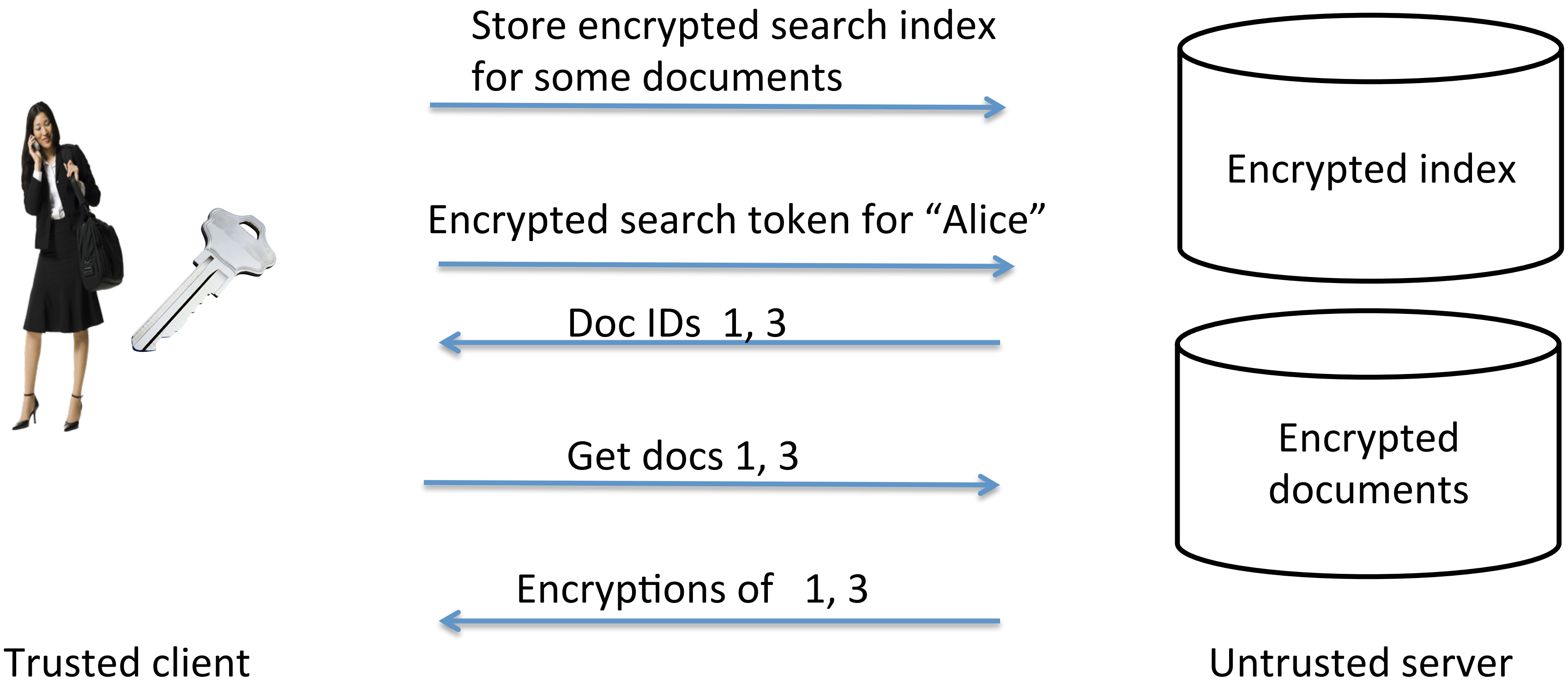
CustName	Zip
Fhbruyf b47394	95421
Hdiel d849g9	16478
Fhbruyf djdlvldl8	94738

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

Enables keyword and phrase queries with no overhead but security is problematic.

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

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

Encrypted index

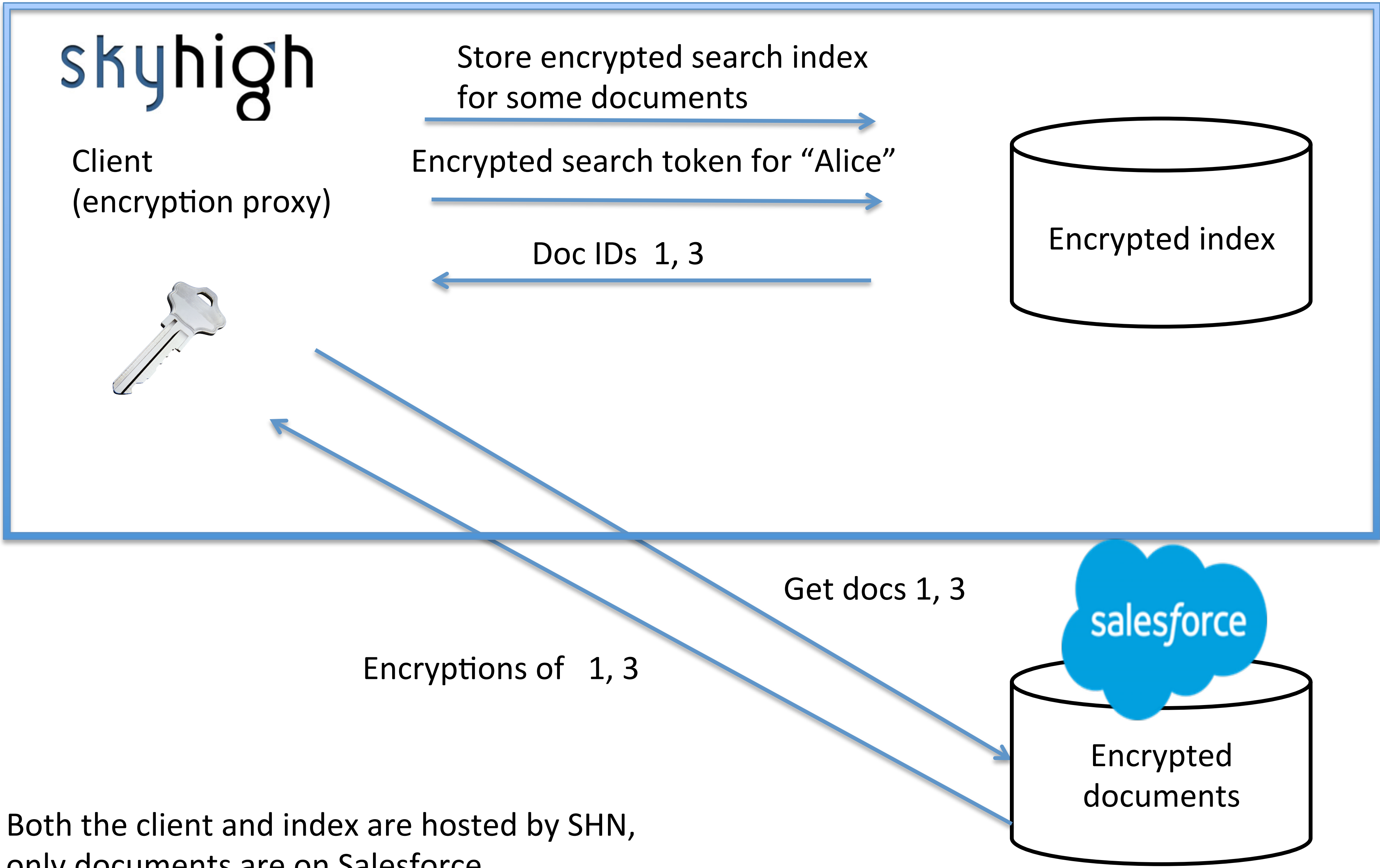
Encrypted documents

Trusted client

Untrusted server

[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

Encrypted documents

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

AcctName	Zip	Value
Alice Cooper	60652	500,000
Bob Jones	46032	1,600,000
Alice Zandra	95014	1,200,000

Encrypted range queries

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

Encryption
Proxy

Bob Jones,
Alice Zandra

Make report on
>$96000478 value

salesforce

Accounts

AcctName	Zip	Value
a7f45edbc	94521	95734857
94dabc467	12379	97563543
1273548fd	40378	96784657

Use order-preserving
encryption (OPE)

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

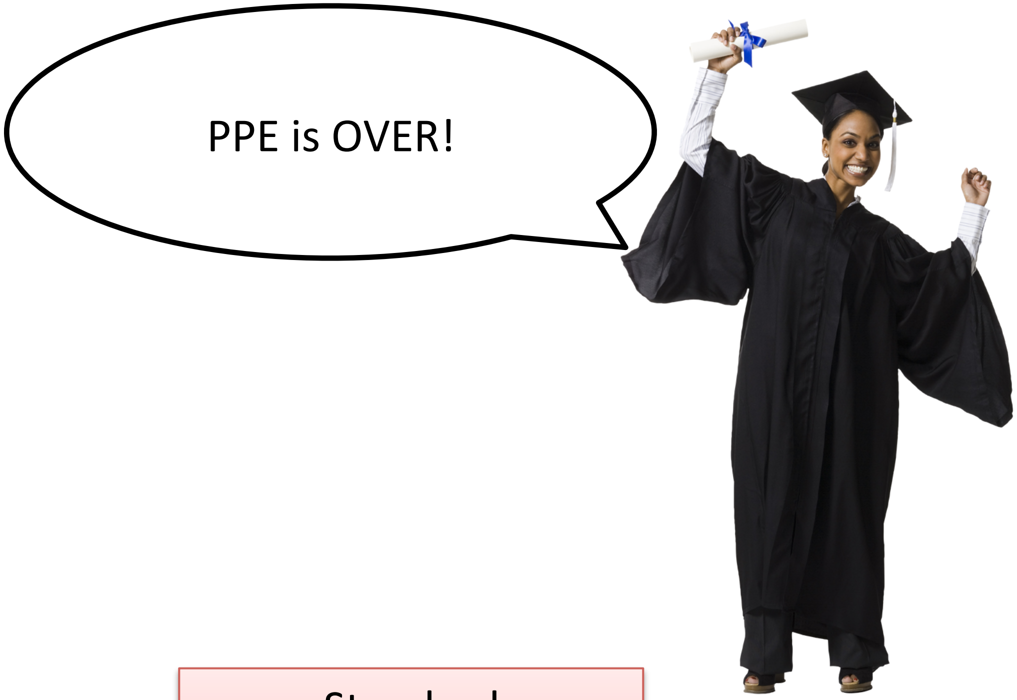
- Interactive is non-starter
  - Global, synchronized state
  - Implementing correctly: person-years of effort for unsure performance
  - Mutability requires additional complexity & custom code, so increased attack surface
- Stateless OPE easier, but still
  - Fixed domain size
  - Efficiency (needed 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

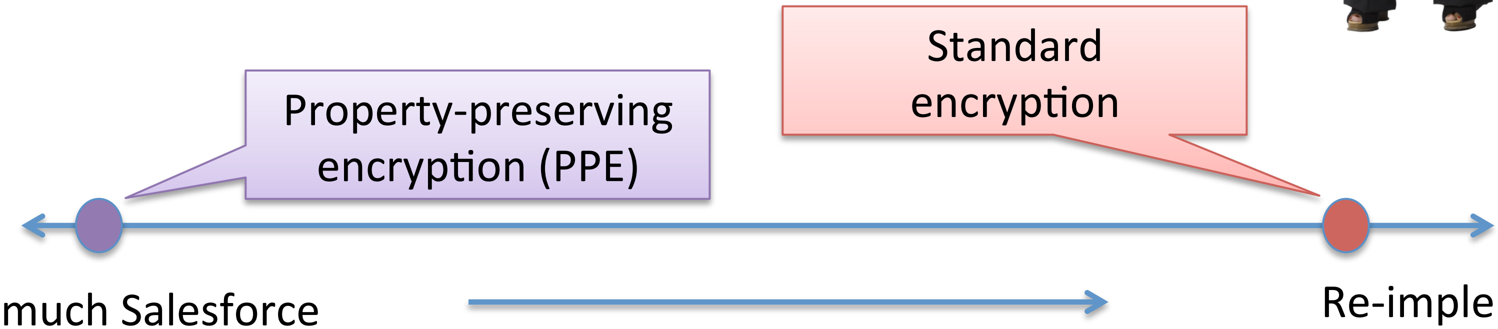
IKK12	Searchable encryption	Query recovery
CGPR15	Searchable encryption	Partial message recovery
NKW15	FPE, OPE	Plaintext recovery

Punchline: PPE can be badly broken in some settings

Crypto researcher on PPE



Hold on a sec…



Property-preserving encryption (PPE)

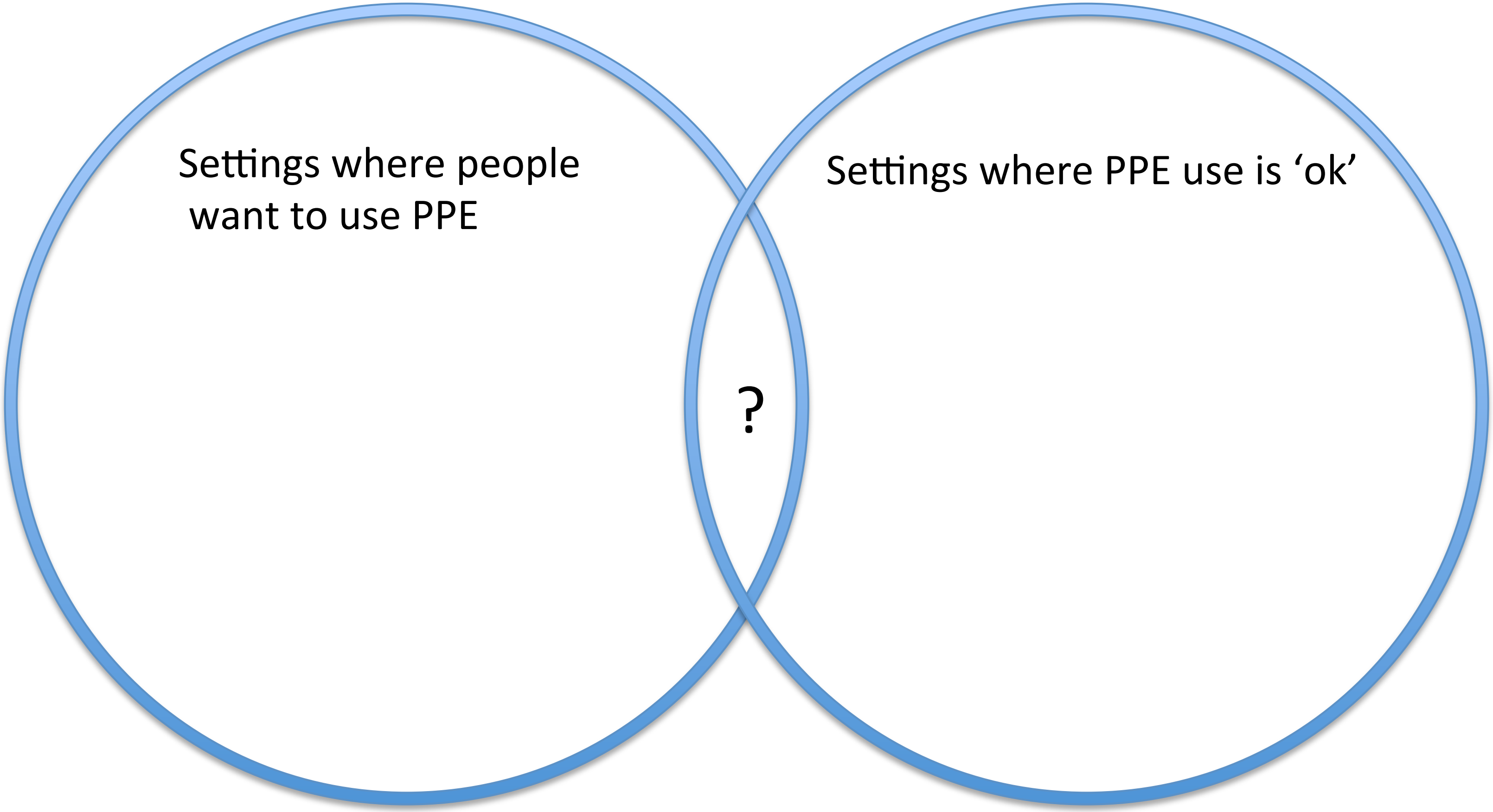
Standard encryption

Use as much Salesforce functionality as possible

Cost of using encryption increases greatly

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