Secure Data Preservers for Web Services

Byung-Gon Chun
Yahoo! Research

Joint work with
Jayanthkumar Kannan (Google) and Petros Maniatis (Intel Labs)

Users Entrust Web Services with Their Data

TigerDirect.com
SHOP BY PHONE 1-800-800-8300

Credit card number

WebMD
Better information. Better health.

Health records

E*TRADE

Trading strategy

bing

Web click logs

Users Entrust Web Services with Their Data

TigerDirect.com

SHOP BY PHONE 1-800-800-8300

Credit card number

WebMD®

Better information. Better health.

Health records

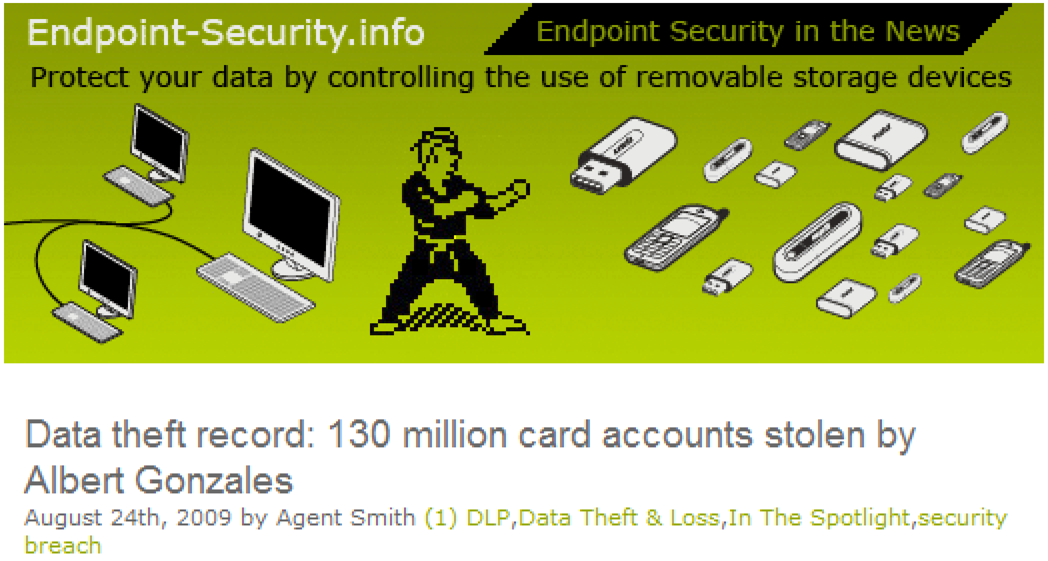
- How their data will be used
- What parts will be shared
- With whom they will be shared

Exposure of Sensitive Data

- dataloss.db lists 400 data loss incidents in 2009; on average exposed half-a-million customer records

Exposure of Sensitive Data

- dataloss.db lists 400 data loss incidents in 2009; on average exposed half-a-million customer records

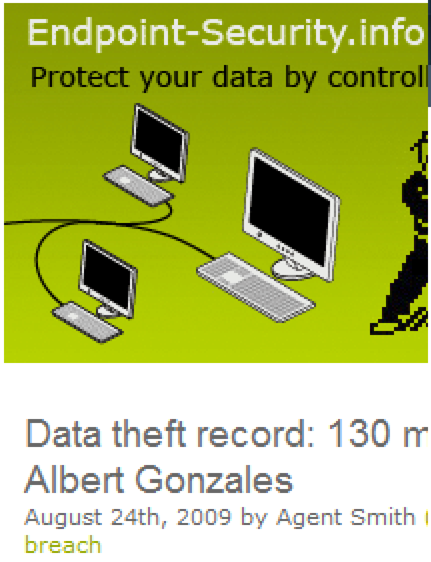

Endpoint-Security.info
Endpoint Security in the News
Protect your data by controlling the use of removable storage devices

Data theft record: 130 million card accounts stolen by Albert Gonzales

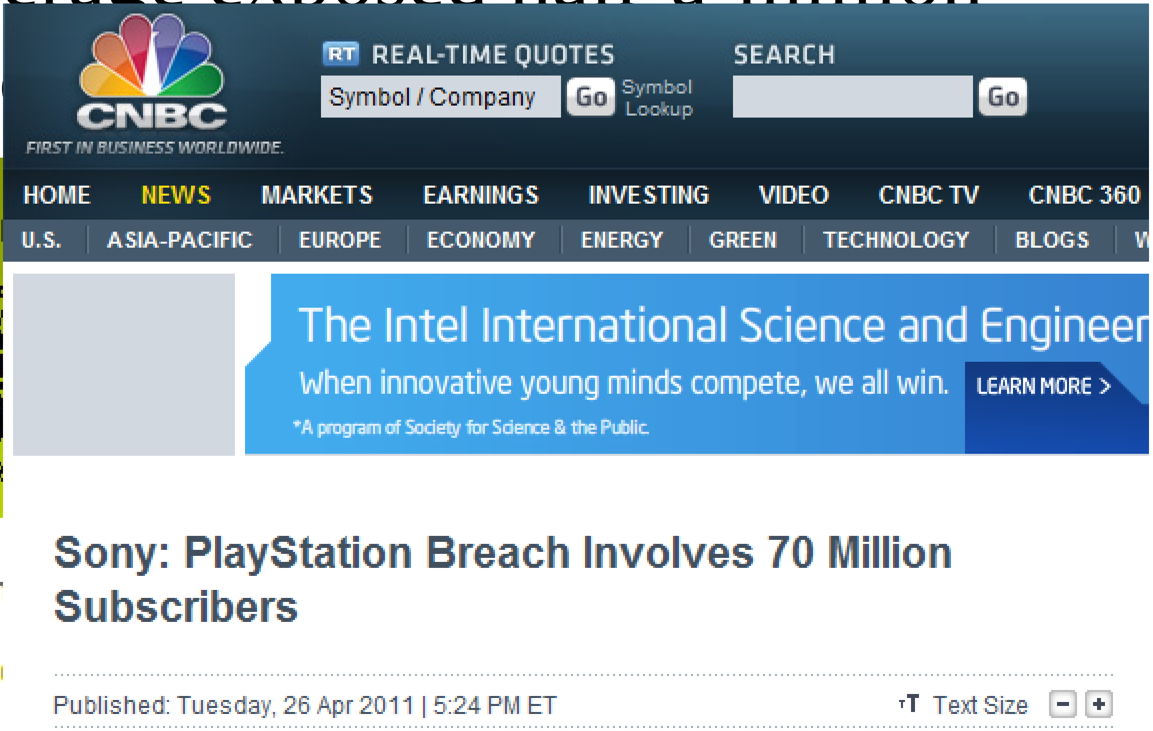
August 24th, 2009 by Agent Smith (1) DLP,Data Theft & Loss,In The Spotlight,security breach

Exposure of Sensitive Data

- dataloss.db lists 400 data loss incidents in 2009; on average exposed half-a-million customer re



Data theft record: 130 m
Albert Gonzales
August 24th, 2009 by Agent Smith
breach



Sony: PlayStation Breach Involves 70 Million Subscribers

Published: Tuesday, 26 Apr 2011 | 5:24 PM ET

Exacerbated by Giving Up Data Usage Control

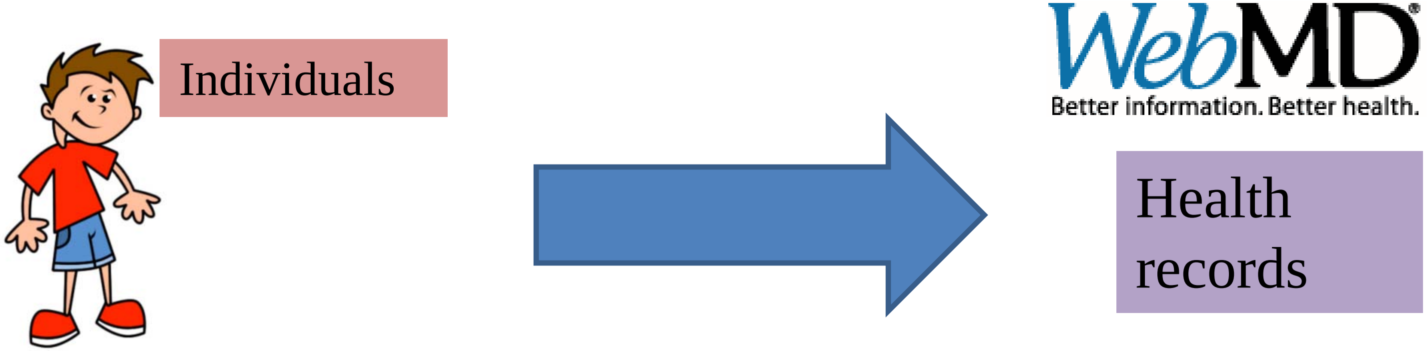
Individuals

Health records

WebMD

Better information. Better health.

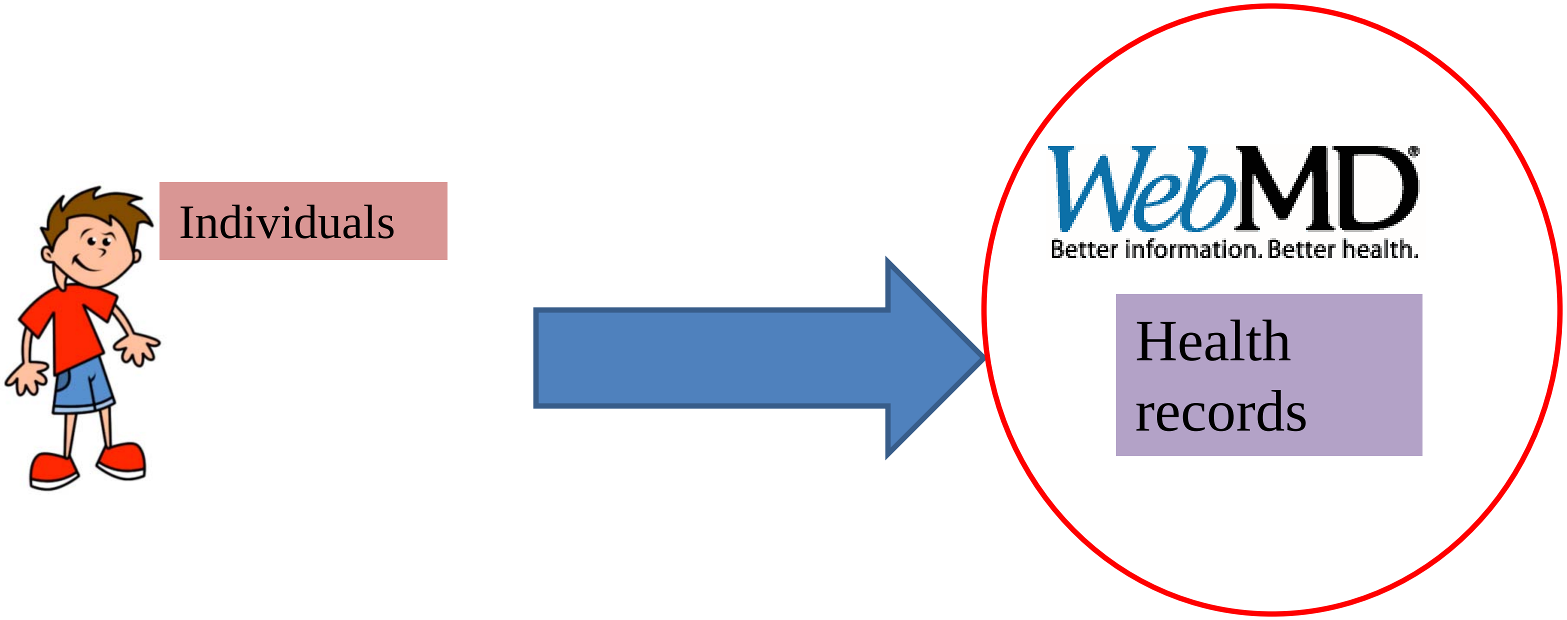
Exacerbated by Giving Up Data Usage Control



Individuals

Health records

Exacerbated by Giving Up Data Usage Control



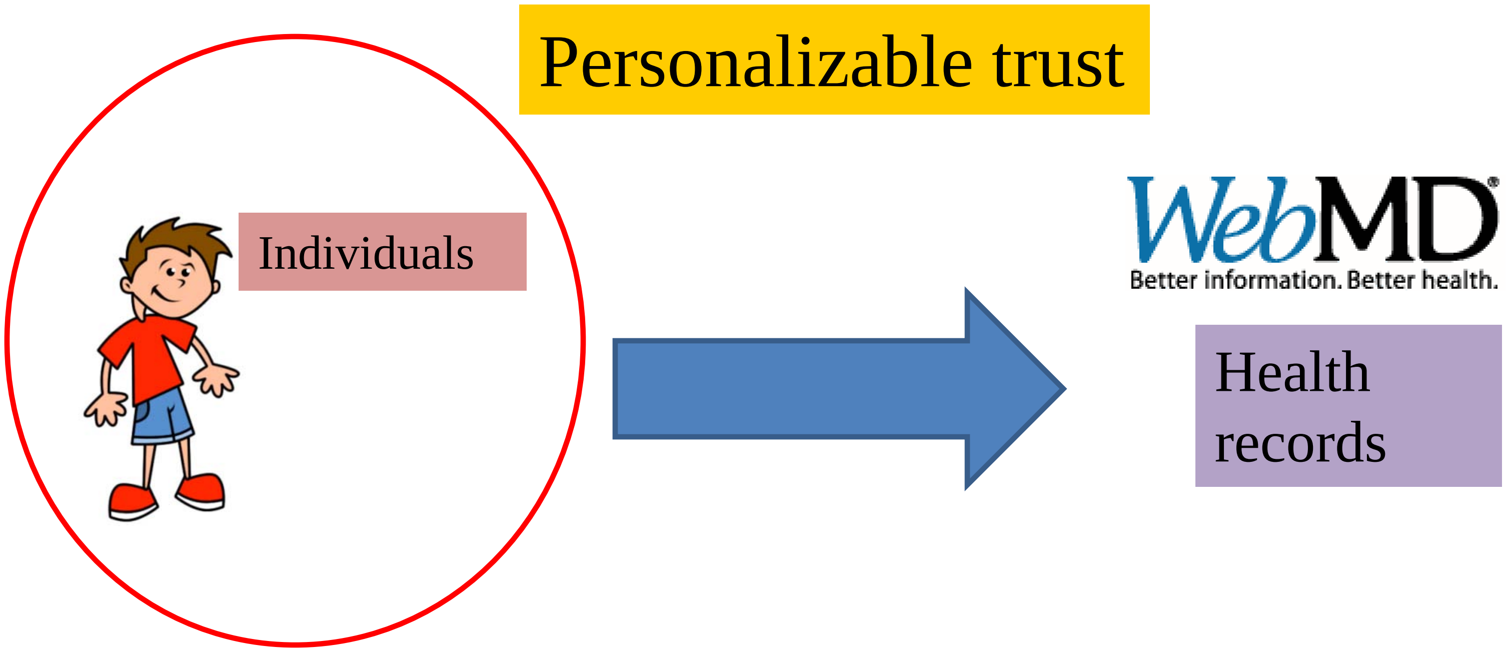
Individuals

WebMD
Better information. Better health.

Health records

- How their data will be used
- What parts will be shared
- With whom they will be shared

Give Control Back to Users



Personalizable trust

Individuals

Health records

- How their data will be used
- What parts will be shared
- With whom they will be shared

Roadmap

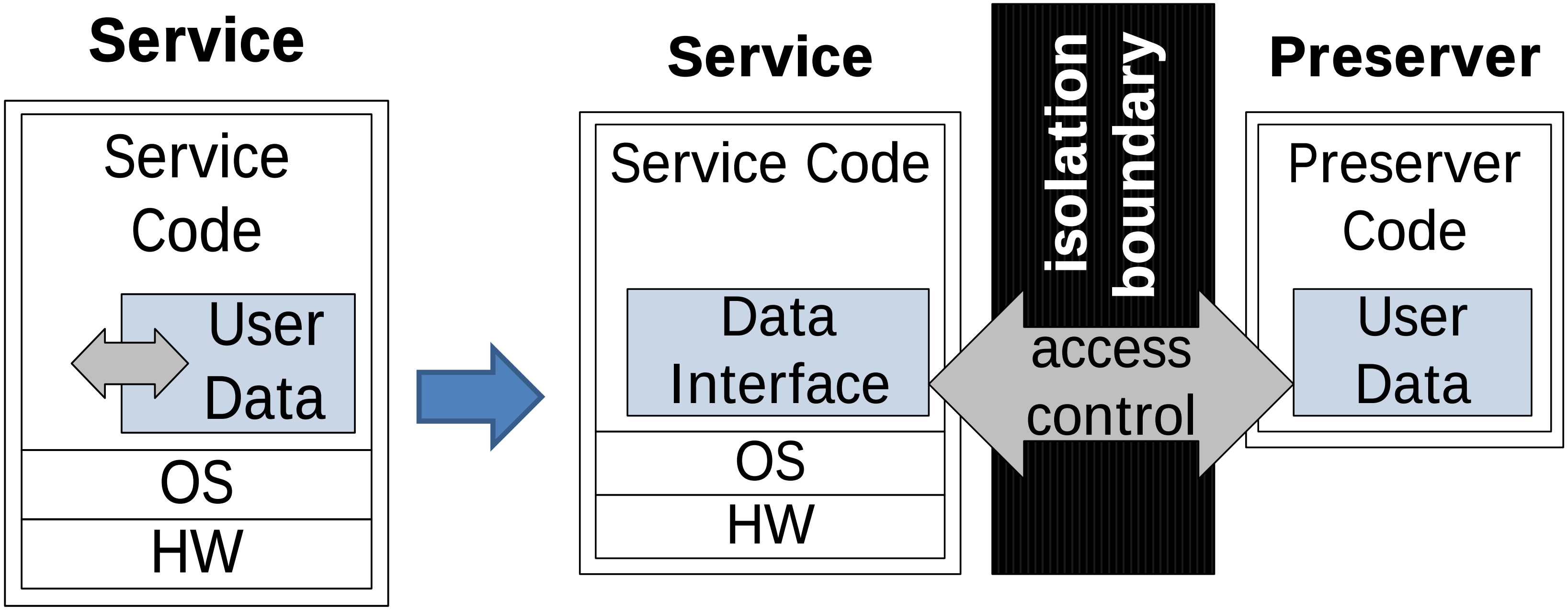
- Motivation
- Secure Data Preserver
- Design
- Evaluation

Our Approach

- Entrusting raw data violates least privilege

- Encapsulate sensitive data and enforce well-defined interface for service to access data

Secure Data Preserver (SDaP)

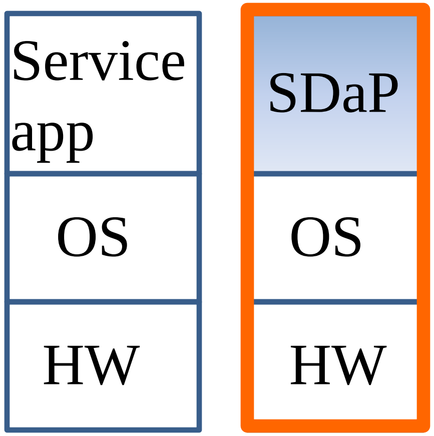


(a) Service + User Data

(b) Service + Preserver

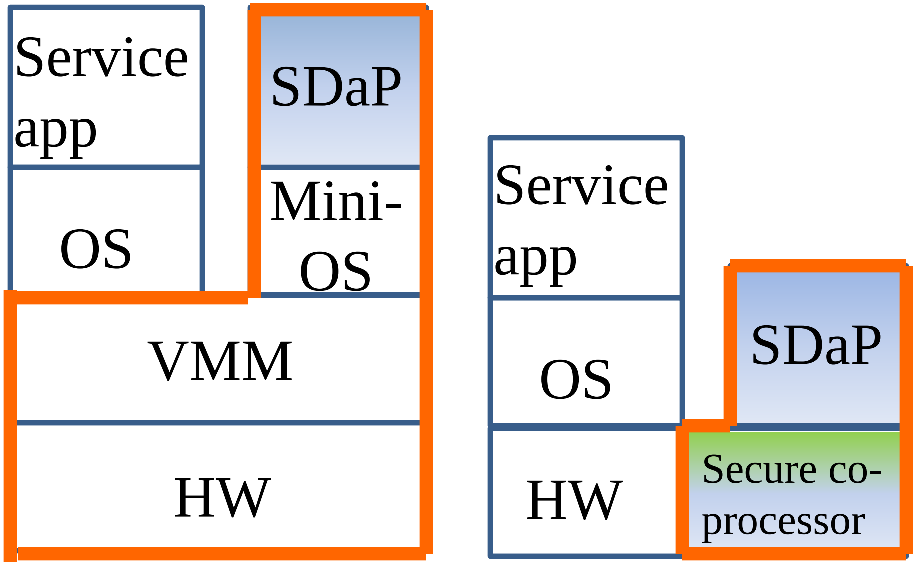
Preserver Deployment Scenarios

Trusted third party or client



Faulty service app
Faulty service operator

Co-location



Faulty service app

Faulty service app
Faulty service operator

What Apps Are Suitable?

- Sensitive query
  - User provides sensitive query, service provides data stream
  - E.g., Trading, Health
- Analytics on sensitive data
  - Service performs data mining on user’s sensitive data
  - E.g., Targeted advertising, Recommendation
- Proxy
  - User provides credentials to another service

What Apps Are Suitable?

- Sensitive query
  - User provides sensitive query, service provides data

* Limitation

Data-centric service reading and updating users' data at fine granularity

- E.g., Docs, Social networking apps

- Proxy
  - User provides credentials to another service

Roadmap

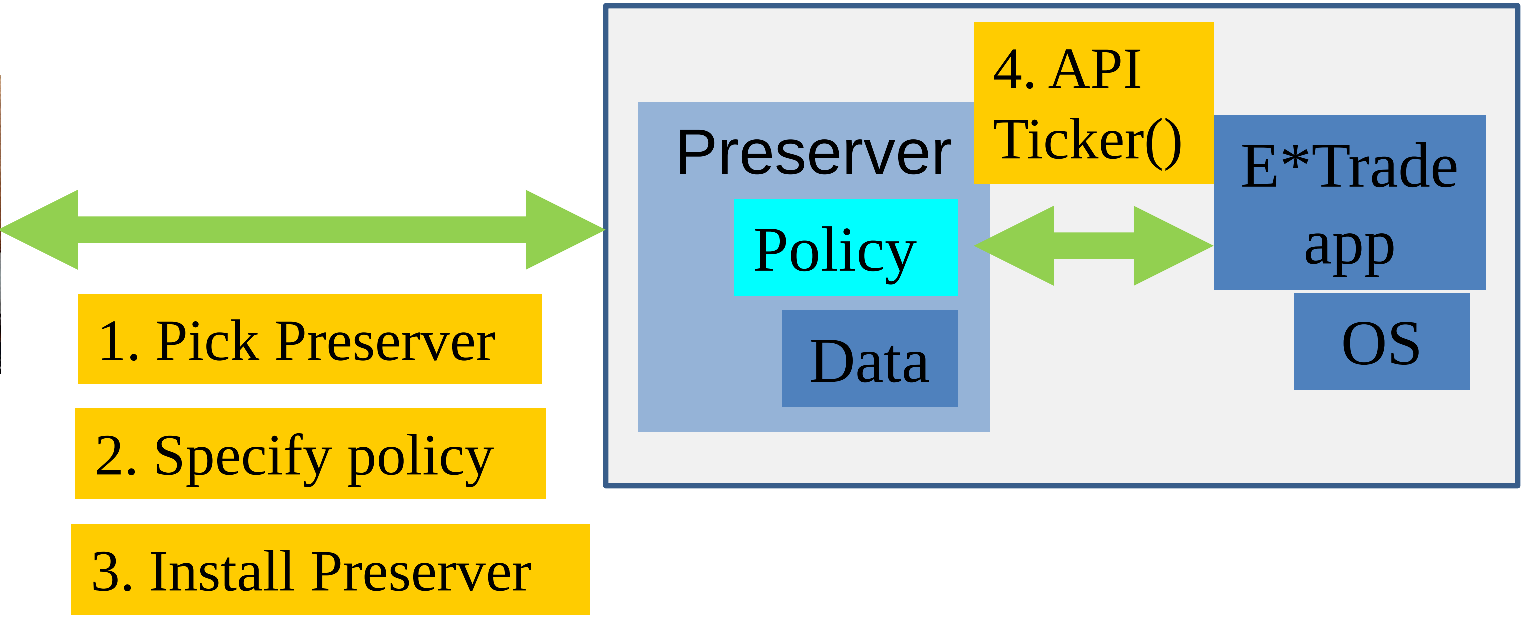
- Motivation
- Secure Data Preserver
- Design
- Evaluation

Preserver Design Goals

- Simple Interface
- Flexible deployment
- Fine-grained use policy
- Trust but mitigate risk

Preserver Operational View

E*TRADE



1. Pick Preserver

2. Specify policy

3. Install Preserver

4. API Ticker()

Preserver

Policy

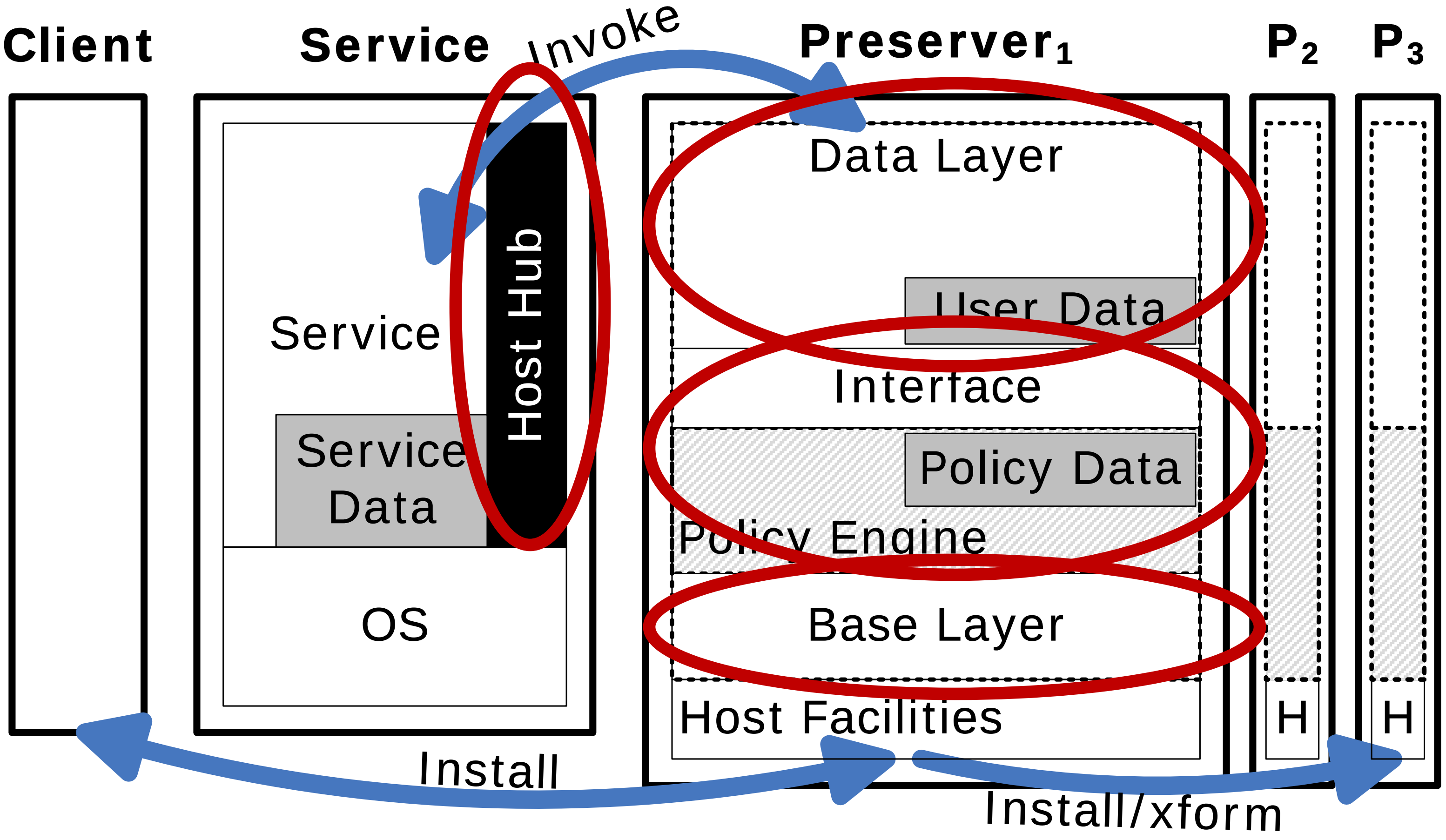
Data

E*Trade app

OS

Preserver Architecture

Hosting Invocation Transformation



Preserver Hosting

- Which services can host users’ preservers
- Hosting policy
  - – Declarative language based on SecPAL

1. *alice* SAYS CanHost(M) IF OwnsMachine(*amazon*, M)

- Hosting mechanism
  - – Hosting protocol based on Diffie-Hellman protocol

Preserver Hosting

- Which services can host users’ preservers
- Hosting policy
  - Declarative language based on SecPAL

2. *alice* SAYS CanHost(M) IF TrustedService(S), OwnsMachine(S,M), HasCoprocessor(M)

- Hosting mechanism
  - Hosting protocol based on Diffie-Hellman protocol

Preserver Hosting

- Which services can host users' preservers
- Hosting policy
  - Declarative language based on SecPAL

3. *alice* SAYS *amazon* CANSAY TrustedService(S)

- Hosting mechanism
  - Hosting protocol based on Diffie-Hellman protocol

Preserver Invocation

- Constrain interface invocation parameters with SecPAL
- Two kinds: stateless, stateful

1. *alice* SAYS CanInvoke(*amazon*, A) IF LessThan(A, 50)

- Transfer of invocation policies: exo-leasing

Preserver Invocation

- Constrain interface invocation parameters with SecPAL
- Two kinds: stateless, stateful

2. *alice* SAYS CanInvoke(*doubleclick*,A) IF
LessThan(A,Limit), Between(Time,”01/01/10”,”01/31/10”)
STATE (Limit=50,Update(Limit,A))

- Transfer of invocation policies: exo-leasing

Preserver Invocation

- Constrain interface invocation parameters with SecPAL
- Two kinds: stateless, stateful

3. *alice* SAYS *amazon* CANSAY CanInvoke(S,A) IF LessThan(A,Limit) STATE (Limit=50,Update(Limit,A))

- Transfer of invocation policies: exo-leasing

Preserver Transformation

- Filtering: retain a subset of data
  - E.g., only the web history in the last six months

- Aggregation: merging of raw data from mutually trusting users of a service
  - E.g., ad-click history of users

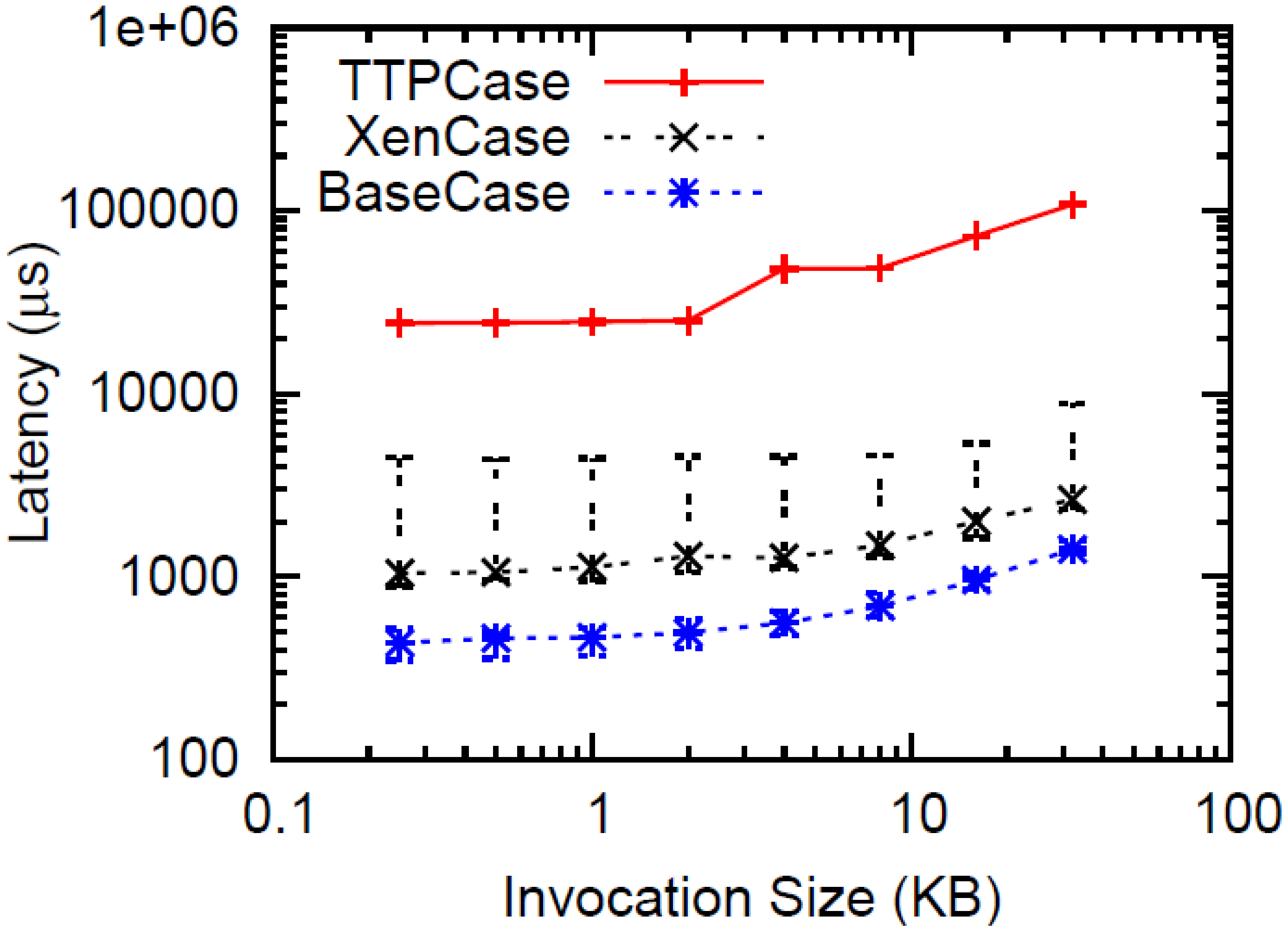
Roadmap

- Motivation
- Secure Data Preserver
- Design
- Evaluation

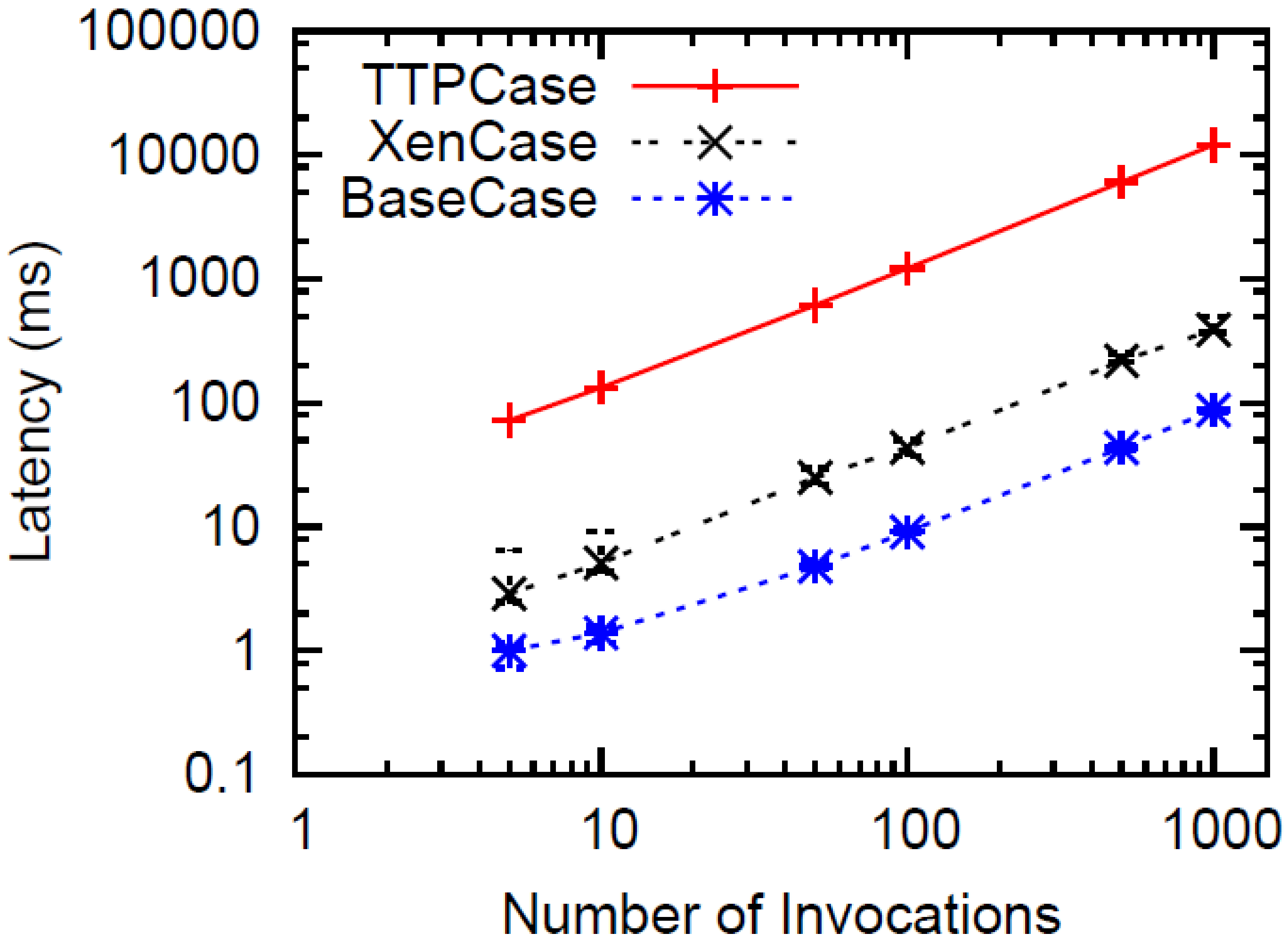
Evaluation

- Deployment options:
  - TTP, client, Xen-based co-location
- Three sample preservers:
  - Stock trading, targeted advertising, credit card xact
- Main results:
  - Cost of preserver
  - Comparison of deployment options
  - Security analysis: LS2-based theoretical analysis, Trusted Computing Base (TCB) comparison

Cost of Basic Invocation (Latency)



Cost of Stock Trading (Latency)



Discussion

- Find appropriate interfaces, verify them
- Easy refactoring
  - Even automated
- Apps with rich interfaces
  - Information flow control

Related Work

- Wilhelm’s mobile agent
- CLAMP
- BSTORE
- Decentralized privacy frameworks
- Information flow control

Conclusion

- Rearchitect web services around the principle of giving data usage control back to users

- Secure Data Preserver achieves this goal via data encapsulation and interface-based access control

Thank you!

Q & A