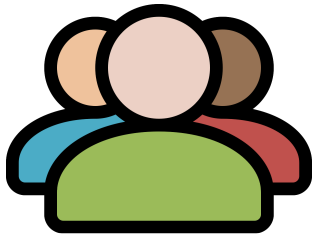
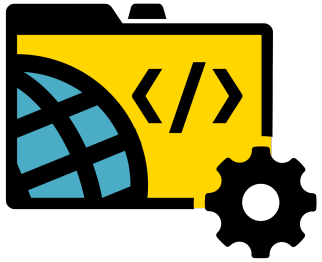
How Do Data Breaches Happen?

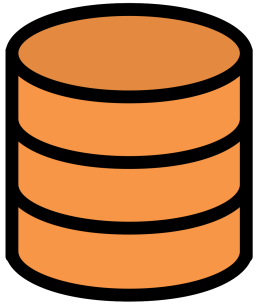


Users

Service Provider

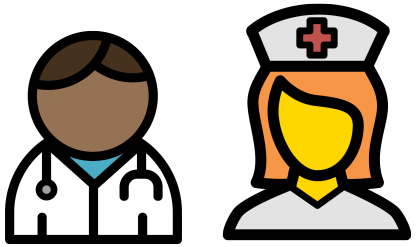


Application



Data

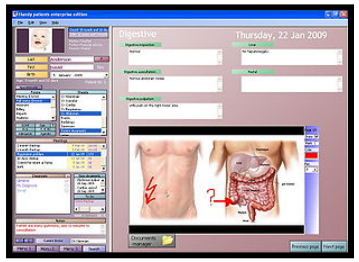
How Do Data Breaches Happen?



Users

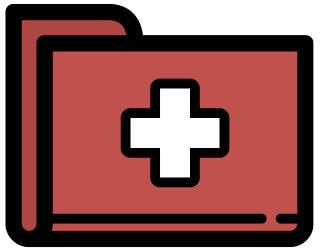
Doctors, Nurses

UCLA Health



Application

Electronic Medical Record System



Data

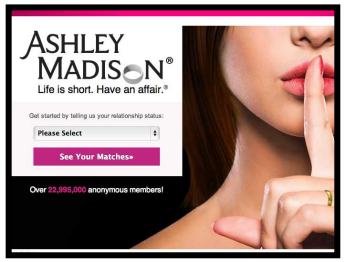
Medical Records

How Do Data Breaches Happen?

Users

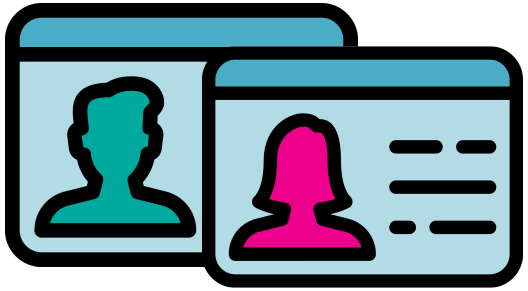
Customers

ASHLEY MADISON®.COM
Life is Short. Have an Affair.®



Application

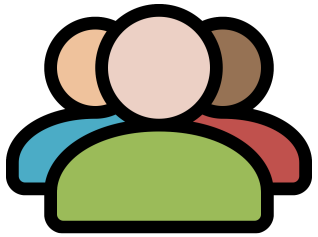
Online Dating



Data

Dating Profiles

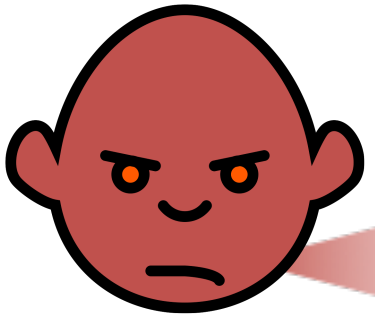
How Do Data Breaches Happen?



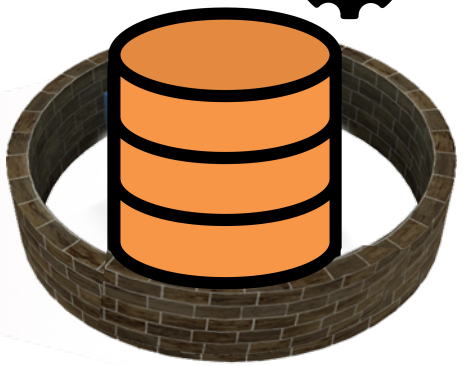
Users

Service Provider

Application



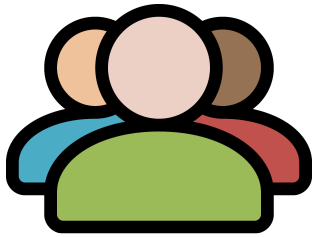
Adversary



Data

How Do Data Breaches Happen?

Encryption Key

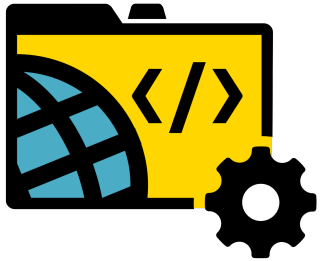


Users

Service Provider

How can we protect the data?

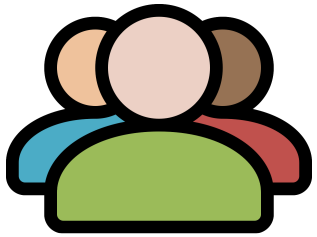
Encryption



Application

Data

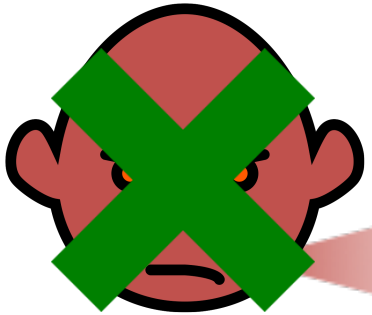
How Do Data Breaches Happen?



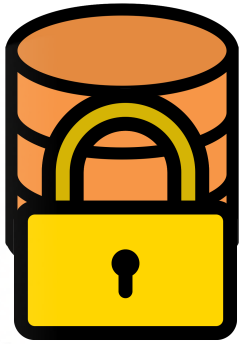
Users

Service Provider

Application



Adversary

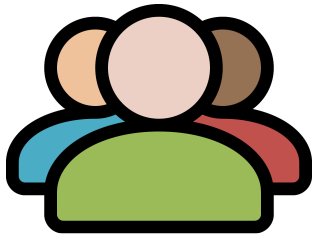


Data

How Do Data Breaches Happen?

Can we protect the data and let the application work?

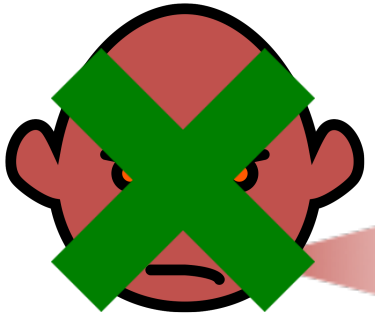
How Do Data Breaches Happen?



Users

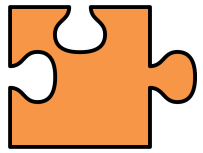
Service Provider

Application



Adversary

Data



Encryption Quiz

Match the characteristics of each encryption:

B	Property Preserving
E	Searchable
D	Secure Computation
A	Homomorphic
C	Functional

A. Computations performed on encrypted data matches the result of the computation on the plaintext.

B. Encrypted data is in the same order as the plaintext.

C. A secret key that allows someone to learn the function that is being encrypted.

D. Several parties can compute a function using inputs that are kept private.

E. Encrypted data that can be searched using encrypted keywords.

Property Preserving Encryption

PPE is widely deployed:

CryptDB
(*SOSP 2011*)

Google
Encrypted BigQuery

Microsoft Research
Cipherbase = Encrypt(Database)

skyhigh

SAP SEEED

Perspecsys
Making the Public Cloud Private™

CipherCloud®
Trust in the Cloud™

Property Preserving Encryption

Microsoft SQL Server 2016 Preview

Always Encrypted

BENEFITS

- Enhanced in-memory performance provides up to 30x faster transactions, more than 100x faster queries than disk-based relational databases and real-time operational analytics

New Always Encrypted technology helps protect your data at rest and in motion, on-premises and in the cloud, with master keys sitting with the application, without application changes

- Built-in advanced analytics provide the scalability and performance benefits of building and running your advanced analytics algorithms directly in the core SQL Server transactional database
- Business insights through rich visualizations on mobile devices with native apps for Windows, iOS and Android
- Simplify management of relational and non-relational data by querying both with T-SQL using PolyBase
- Faster hybrid backups, high availability and disaster recovery scenarios to back up and restore your on-premises databases to Microsoft Azure and place your SQL Server AlwaysOn secondaries in Azure

Property Preserving Encryption

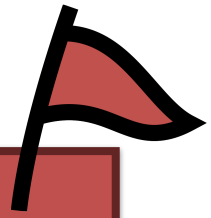
Why is PPE so popular?

Deployability
No change to application and database servers

Expressiveness
Supports most common SQL queries

Efficiency
~25% overhead

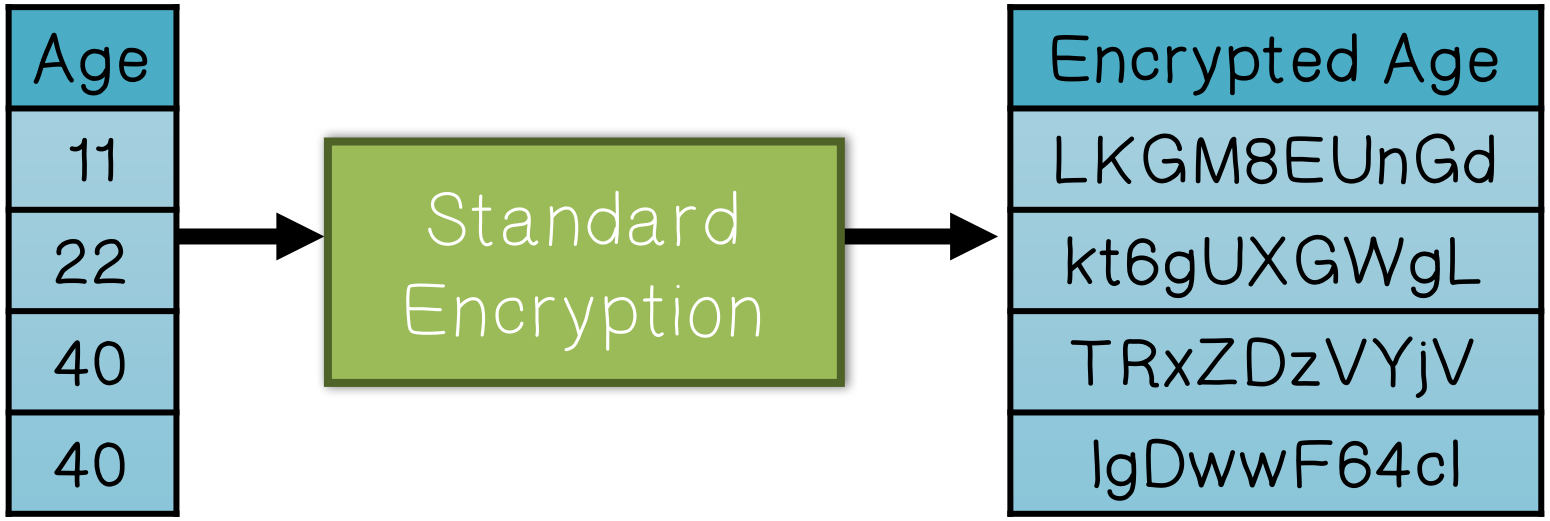
Security
???

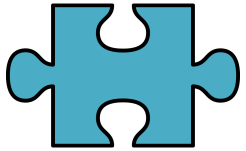


PPE Leak Quiz One

Standard Encryption

Leaks nothing except size





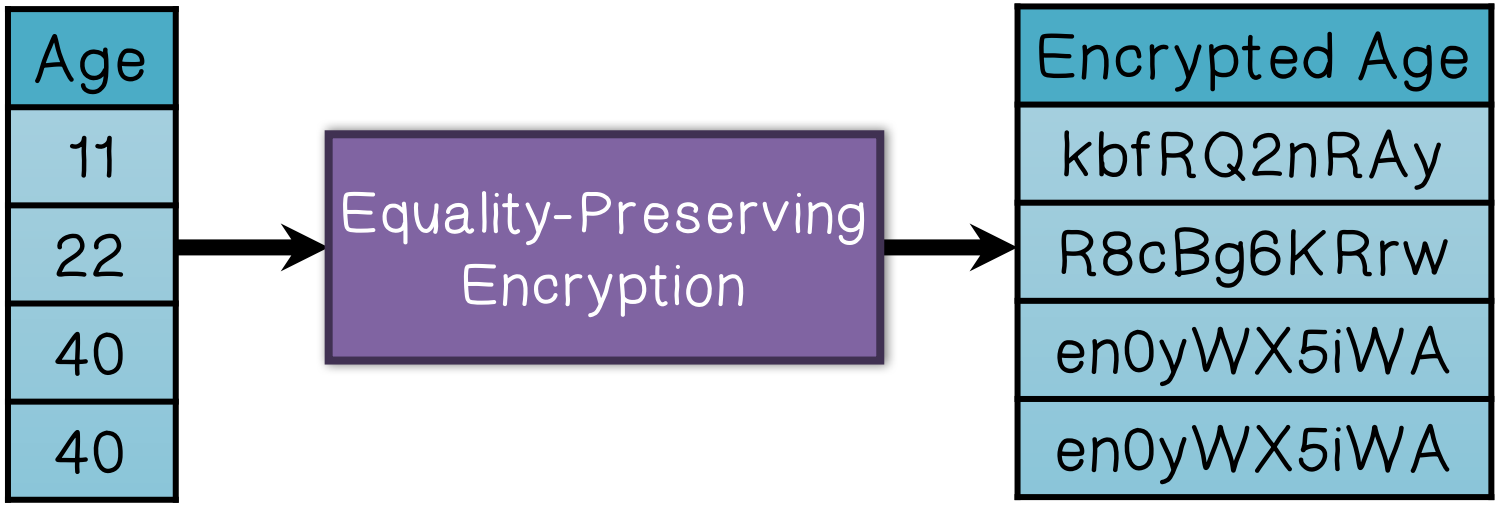
PPE Leak Quiz One

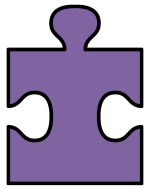
In the given property preserving encryption:

What is preserved? equality

What is leaked? frequency

Property-Preserving Encryption



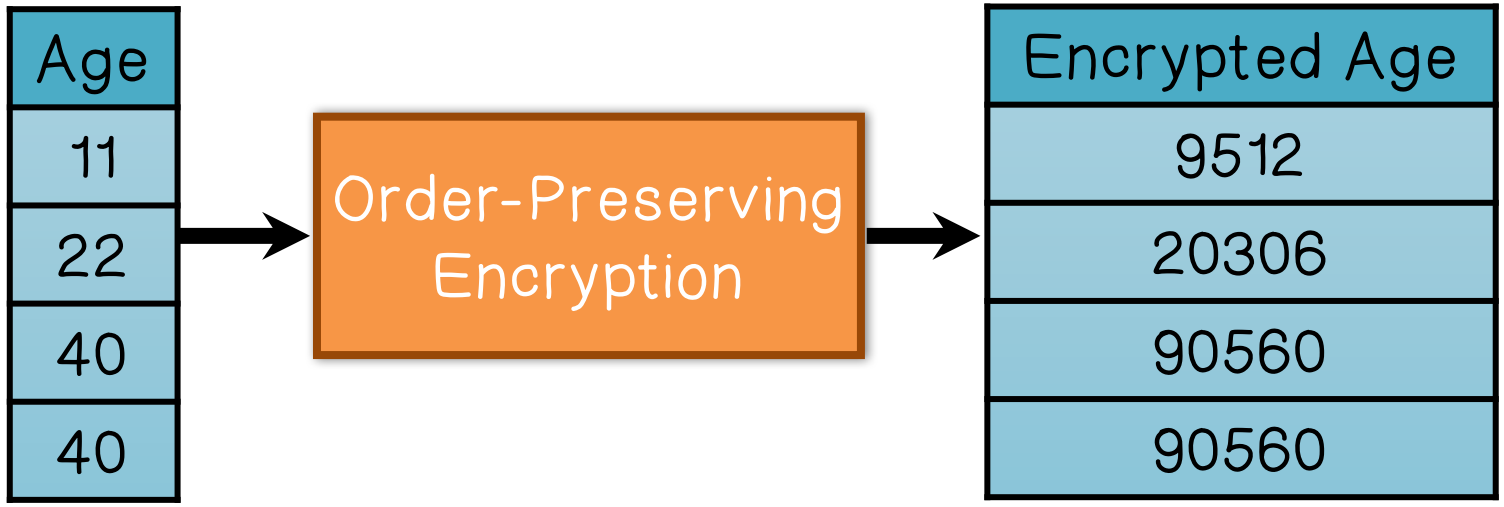


PPE Leak Quiz One

What is leaked in order-preserving encryption?

Order	Frequency

Property-Preserving Encryption



PPE Leakage

What does this leakage mean for real applications?

Open EMR
open-emr.org

Electronic Medical Records

PPE Leakage

Choosing Attributes

Attributes	Query Type
Sex	Equality
Race	Equality
Age	Order
Admission Month	Order
Patient Died	Equality
Primary Payer	Equality
Length of Stay	Order
Mortality Risk	Order
Disease Severity	Order
Major Diagnostic Category	Equality
Admission Type	Order
Admission Source	Equality

PPE Leakage

Sensitivity of the Attributes

Attributes Sensitive to Hospitals
Patient Died
Length of Stay
Mortality Risk
Disease Severity
Major Diagnostic Category

Attributes Sensitive to Patients
Sex
Race
Age
Admission Month
Patient Died
Primary Payer
Length of Stay
Mortality Risk
Disease Severity
Major Diagnostic Category
Admission Type
Admission Source

Data for Attributes

AHRQ Advancing Excellence in Health Care

H·CUP

2009

200 hospitals with
12,975 to 121,664 patients

TEXAS Department of State Health Services

2008

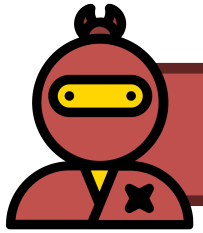
AHRQ Advancing Excellence in Health Care

H·CUP

2004

Encrypted Data

Auxiliary Data



INFERENCE ATTACK!!.

Plaintext Data

Encryption Attacks

Equality-Preserving Encryption

Frequency Analysis

[Al-Kindi 9th Century]

Order-Preserving

ℓ_p-Optimization Attack

New Attack [Naveed, Kamara, Wright, CCS 2015]

Attack Analysis

Encrypted Column

Number of days patient stayed in the hospital

Patient ID	#Days
Patient 1	467xt2J23t
Patient 2	467xt2J23t
Patient 3	42wBrfQ7yn
Patient 4	467xt2J23t
Patient 5	fNewgCgM20
Patient 6	467xt2J23t
Patient 7	fNewgCgM20
Patient 8	467xt2J23t
Patient 9	zt1NEtbkn9
Patient 10	467xt2J23t
Patient 11	42wBrfQ7yn
Patient 12	467xt2J23t
Patient 13	j0th0LelsP
Patient 14	zt1NEtbkn9
Patient 15	zt1NEtbkn9
Patient 16	467xt2J23t
Patient 17	42wBrfQ7yn
Patient 18	467xt2J23t
Patient 19	467xt2J23t
Patient 20	42wBrfQ7yn

Sorted Histogram

#Days	Frequency
467xt2J23t	10
42wBrfQ7yn	4
zt1NEtbkn9	3
fNewgCgM20	2
j0th0LelsP	1

Sorted Auxiliary Histogram

#Days	Frequency
1	11
2	5
3	3
4	2
5	1

#Days Ciphertext	#Days Plaintext
467xt2J23t	1
42wBrfQ7yn	2
zt1NEtbkn9	3
fNewgCgM20	4
j0th0LelsP	5

Recovered Column

Patient ID	#Days
Patient 1	1
Patient 2	1
Patient 3	2
Patient 4	1
Patient 5	4
Patient 6	1
Patient 7	4
Patient 8	1
Patient 9	3
Patient 10	1
Patient 11	2
Patient 12	1
Patient 13	5
Patient 14	3
Patient 15	3
Patient 16	1
Patient 17	2
Patient 18	1
Patient 19	1
Patient 20	2

ℓ_p-Optimization Attack

Encrypted Column

Number of days patient stayed in the hospital

Patient ID	#Days
Patient 1	467xt2J23t
Patient 2	467xt2J23t
Patient 3	42wBrfQ7yn
Patient 4	467xt2J23t
Patient 5	fNewgCgM20
Patient 6	467xt2J23t
Patient 7	fNewgCgM20
Patient 8	467xt2J23t
Patient 9	zt1NEtbkn9
Patient 10	467xt2J23t
Patient 11	42wBrfQ7yn
Patient 12	467xt2J23t
Patient 13	j0th0LelsP
Patient 14	zt1NEtbkn9
Patient 15	zt1NEtbkn9
Patient 16	467xt2J23t
Patient 17	42wBrfQ7yn
Patient 18	467xt2J23t
Patient 19	467xt2J23t
Patient 20	42wBrfQ7yn

Sorted Histogram

#Days	Frequency
467xt2J23t	10
42wBrfQ7yn	4
zt1NEtbkn9	3
fNewgCgM20	2
j0th0LelsP	1

Sorted Auxiliary Histogram

#Days	Frequency
1	11
2	5
3	3
4	2
5	1

#Days Ciphertext	#Days Plaintext
467xt2J23t	1
42wBrfQ7yn	2
zt1NEtbkn9	3
fNewgCgM20	4
j0th0LelsP	5

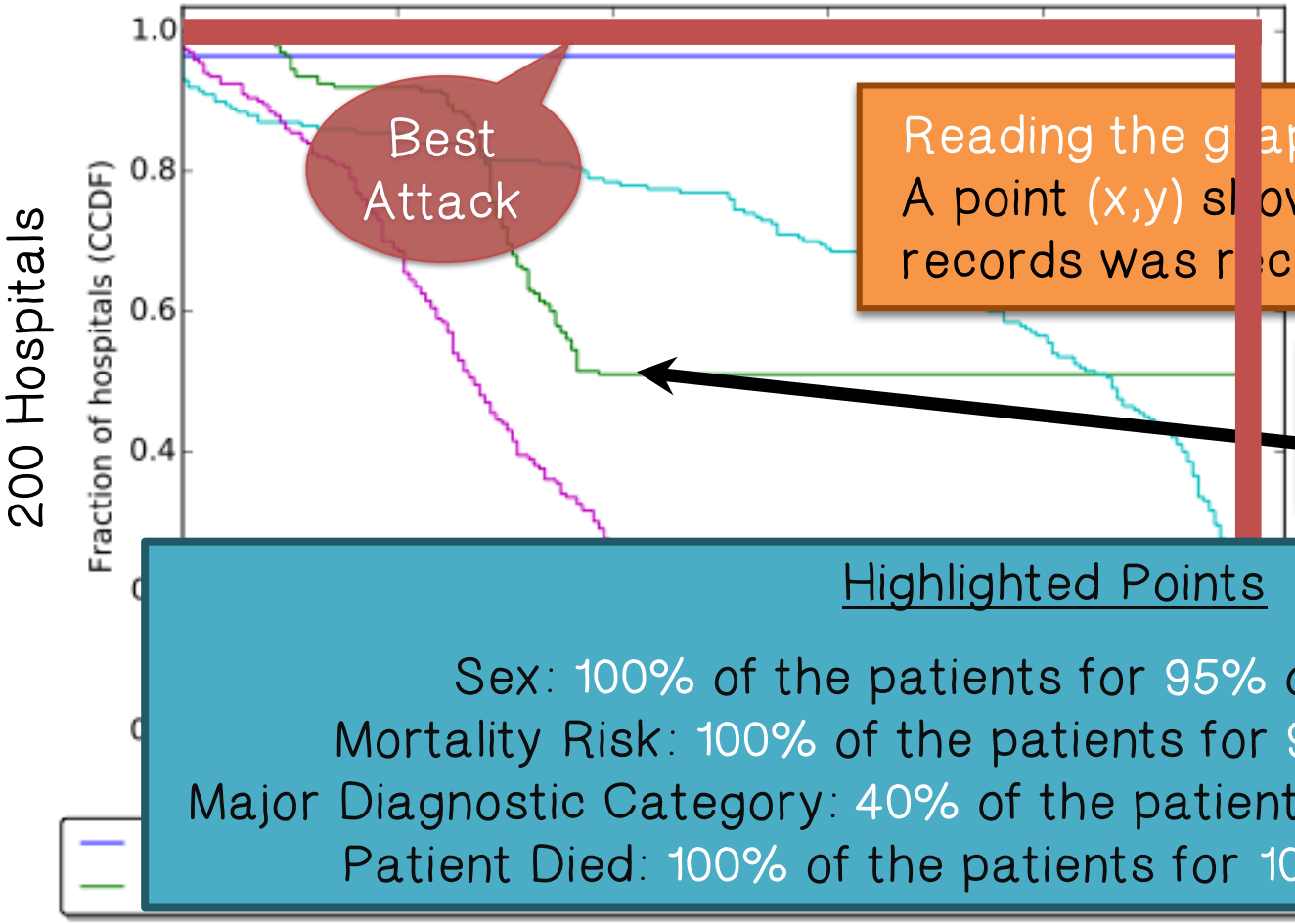
Recovered Column

Patient ID	#Days
Patient 1	1
Patient 2	1
Patient 3	2
Patient 4	1
Patient 5	4
Patient 6	1
Patient 7	4
Patient 8	1
Patient 9	3
Patient 10	1
Patient 11	2
Patient 12	1
Patient 13	5
Patient 14	3
Patient 15	3
Patient 16	1
Patient 17	2
Patient 18	1
Patient 19	1
Patient 20	2

Optimization Attack Analysis

Encrypted Data H·CUP 2009

Auxiliary Data H·CUP 2004



Reading the graph
A point (x,y) shows that at least x fraction of records was recovered for y fraction of hospitals.

At least 40% of records were recovered for 50% of hospitals

Highlighted Points

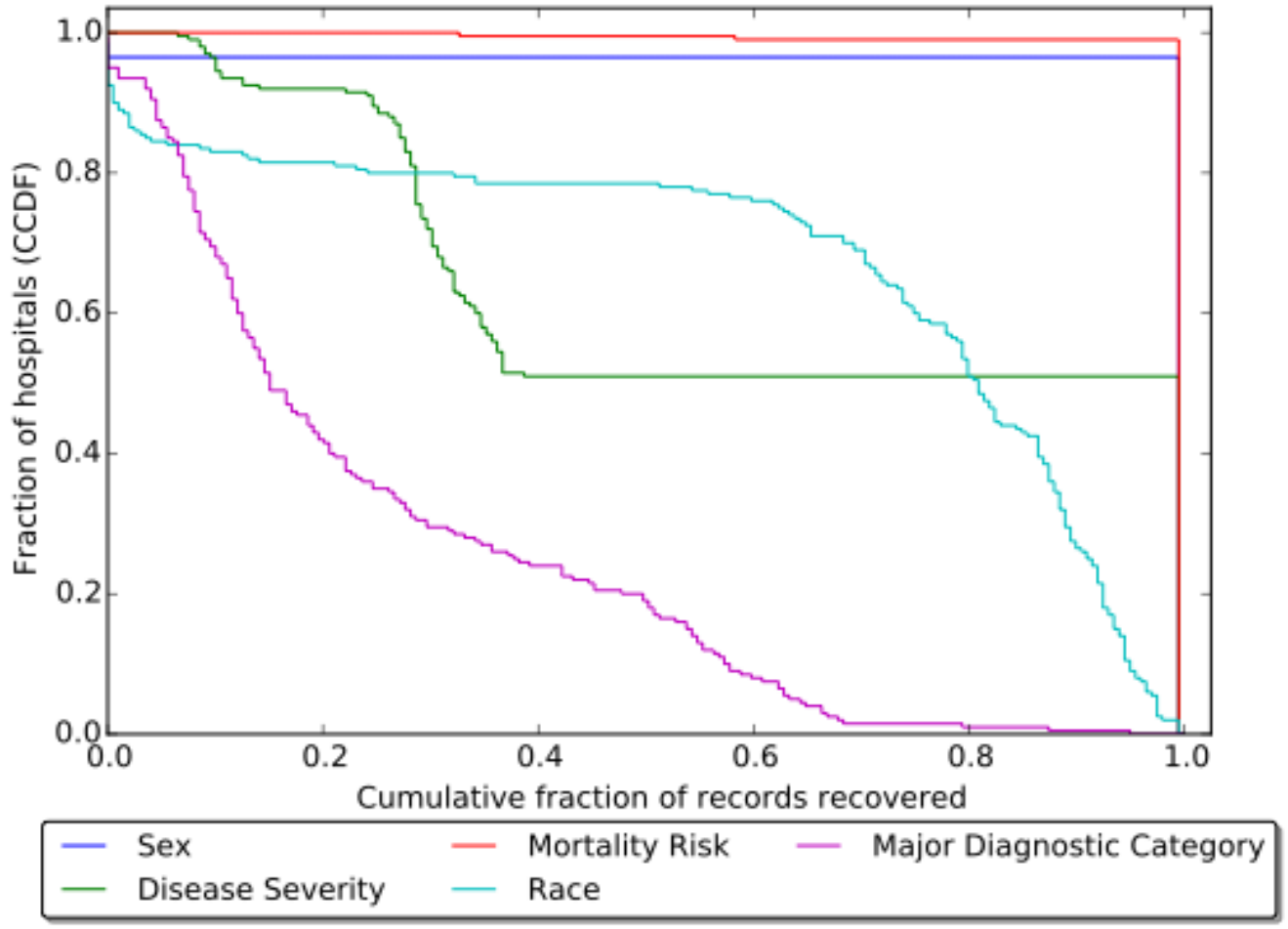
Sex: 100% of the patients for 95% of the hospitals
Mortality Risk: 100% of the patients for 99% of the hospitals
Major Diagnostic Category: 40% of the patients for 28% of the hospitals
Patient Died: 100% of the patients for 100% of the hospitals

Optimization Attack Analysis

Encrypted Data — H·CUP 2009

Auxiliary Data — TEXAS Department of State Health Services 2008

200 Hospitals

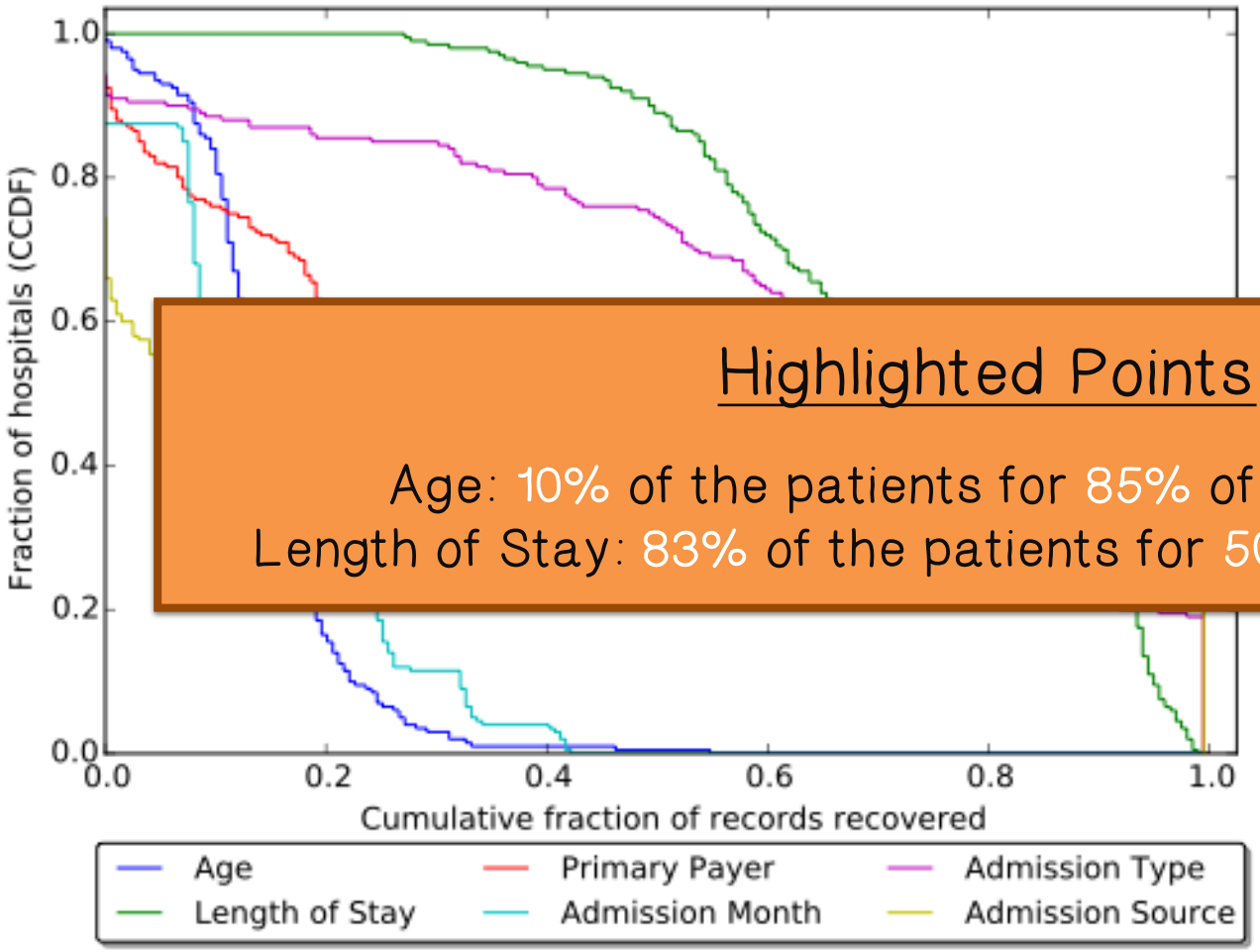


Optimization Attack Analysis

Encrypted Data — H·CUP 2009

Auxiliary Data — H·CUP 2004

200 Hospitals



Highlighted Points

Age: 10% of the patients for 85% of the hospitals

Length of Stay: 83% of the patients for 50% of the hospitals

Cumulative Attack

Sorting Attack

Cumulative Attack

New Attack [Naveed, Kamara, Wright, CCS 2015]

Order-Preserving Encryption

Cumulative Attack

Linear Sum Assignment Problem (LSAP)

Takes $O(n^3)$ time using Hungarian Algorithm

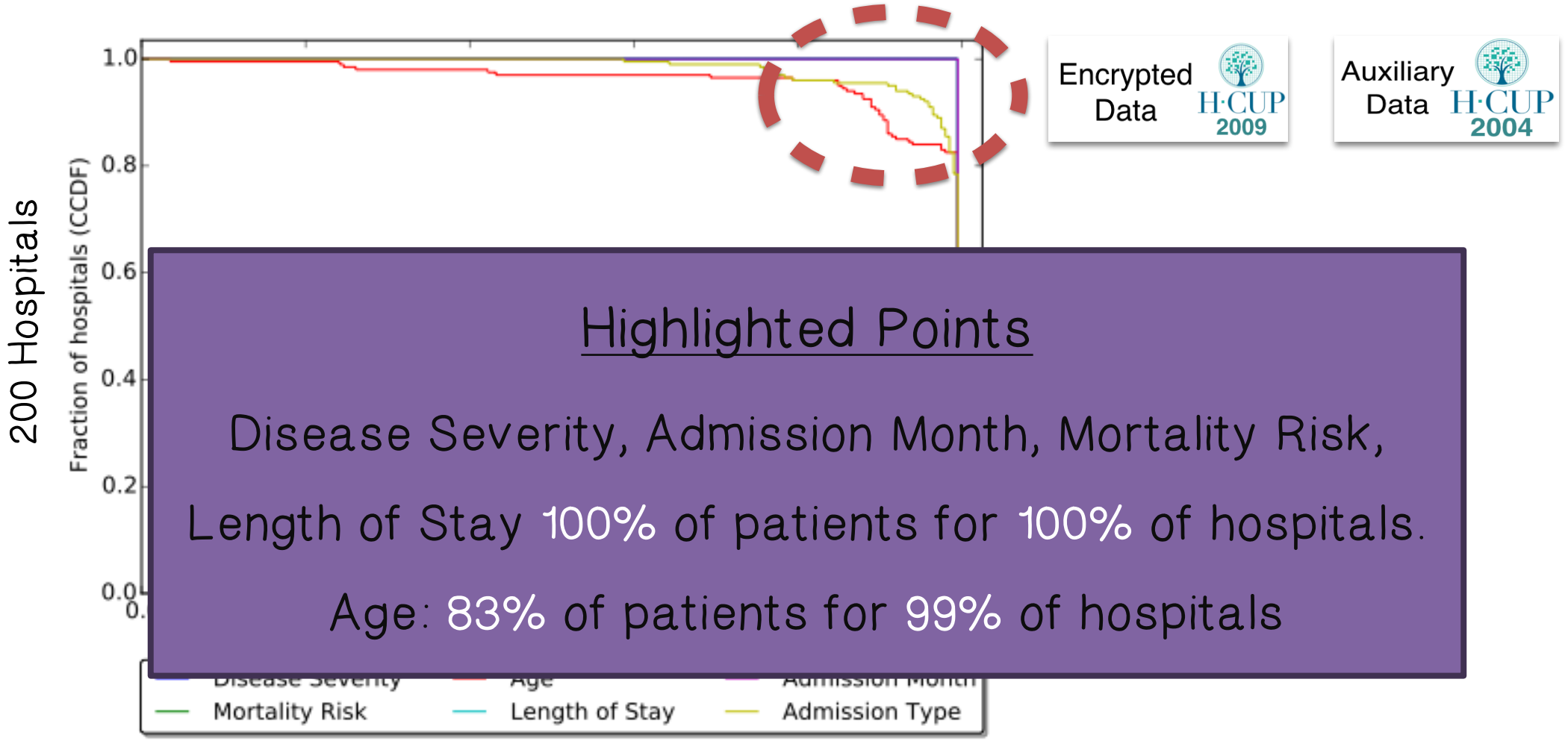
Cumulative attack exploits both order and frequency

Cumulative-Atk$(\mathbf{c}, \mathbf{z})$:

1. $\psi \leftarrow \text{Hist}(\mathbf{c})$ and $\varphi \leftarrow \text{CDF}(\mathbf{c})$;
2. $\pi \leftarrow \text{Hist}(\mathbf{z})$ and $\mu \leftarrow \text{CDF}(\mathbf{z})$;
3. output

$$\arg\min_{X \in \mathbb{P}} \sum_{i=1}^{|\mathbb{M}_k|} \left(|\psi_i - X_i \cdot \pi| + |\varphi_i - X_i \cdot \mu| \right)$$

Cumulative Attack Analysis



Encrypted Data H·CUP 2009

Auxiliary Data H·CUP 2004

Highlighted Points

Disease Severity, Admission Month, Mortality Risk,

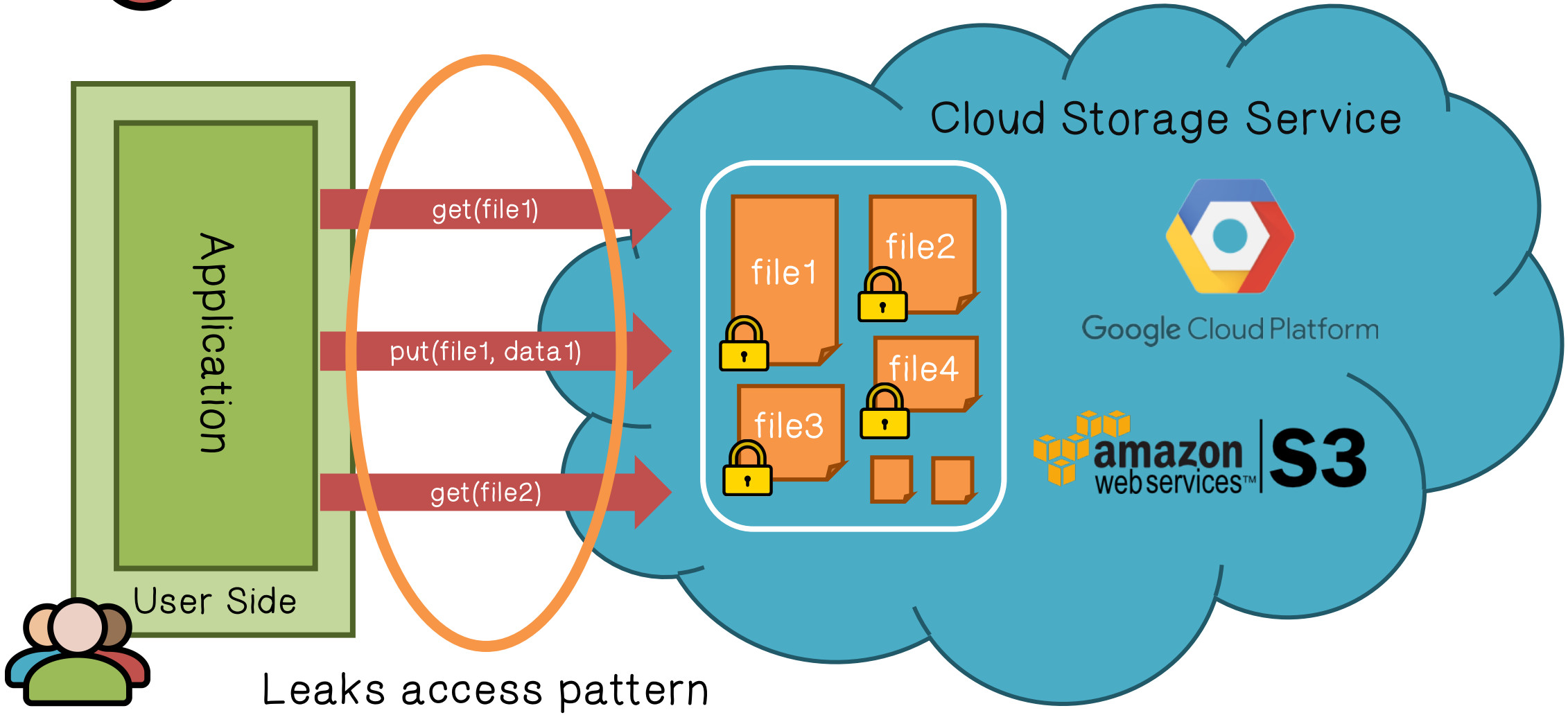
Length of Stay 100% of patients for 100% of hospitals.

Age: 83% of patients for 99% of hospitals

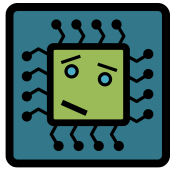
Attack Recap

Attribute	Equality-Preserving Encryption Attack	Order-Preserving Encryption Attack
Sex	100% patients, 95% hospitals	—
Race	60% patients, 70% hospitals	—
Age	10% patients, 85% hospitals	83% patients, 99% hospitals
Admission Month	20% patients, 55% hospitals	100% patients, 100% hospitals
Patient Died	100% patients, 100% hospitals	—
Primary Payer	90% patients, 38% hospitals	—
Length of Stay	83% patients, 50% hospitals	100% patients, 100% hospitals
Mortality Risk	100% patients, 99% hospitals	100% patients, 100% hospitals
Disease Severity	100% patients, 51% hospitals	100% patients, 100% hospitals
Major Diagnostic Category	40% patients, 28% hospitals	—
Admission Type	60% patients, 65% hospitals	79% patients, 100% hospitals
Admission Source	90% patients, 38% hospitals	—

Suppose We Don't Trust the Cloud



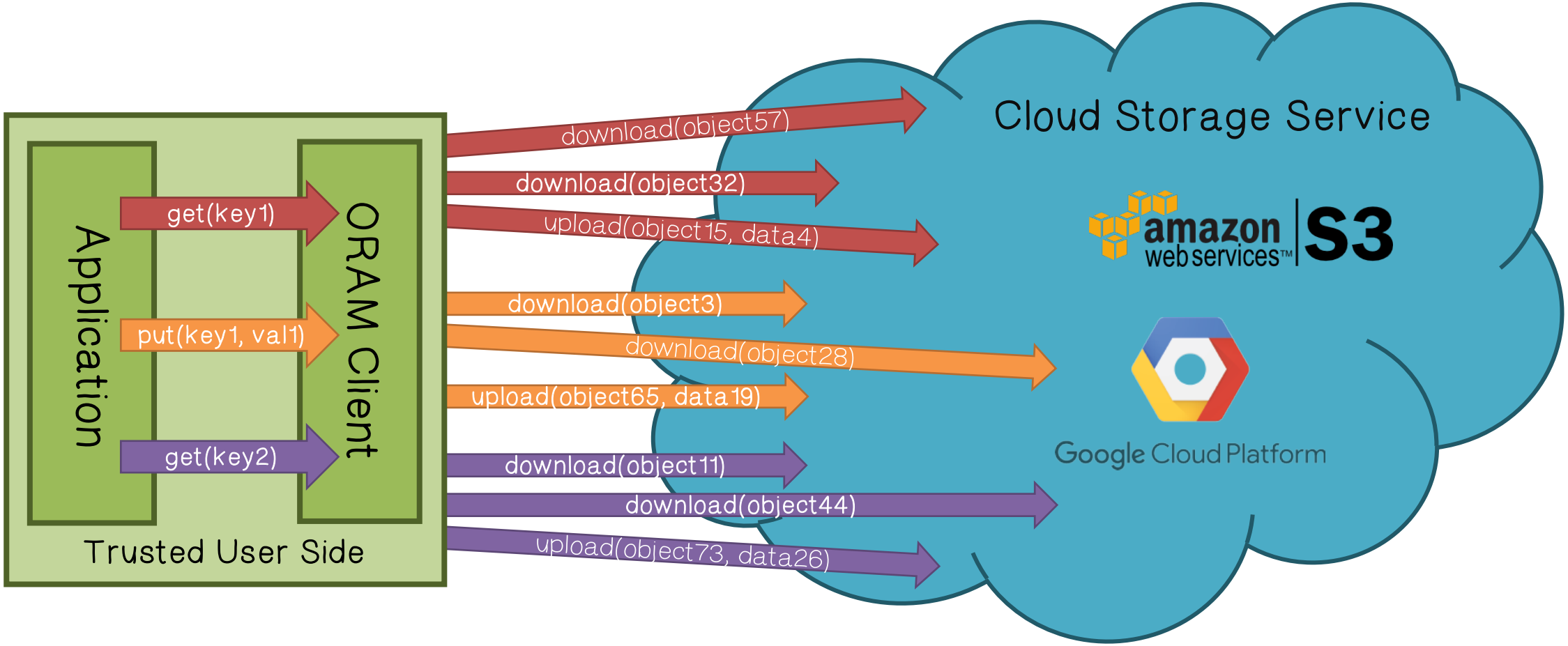
Leaks access pattern

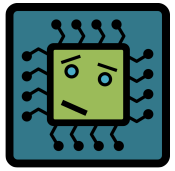


Oblivious RAM

✔ Obliviousness:	! Techniques
• For any fixed size request sequence, the associated storages accesses observed (by the cloud) are statistically independent of the requests	• Operates on fixed size data blocks • Encrypt blocks with ciphertext indistinguishability • Dummy accesses, re-encryption, shuffling, etc.

Oblivious RAM



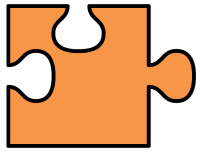


Oblivious RAM

Some ORAM Systems

- Tree-based: PathORAM
- Layered-based: LayeredORAM
- Large messages-based: PracticalOS
- Partition-based: ObliviStore

1. **[**PathORAM**]** Stefanov, Emil, et al. *"Path ORAM: An Extremely Simple Oblivious RAM Protocol."* CCS 2013.
2. **[**LayeredORAM**]** Goodrich, Michael, et al. *"Oblivious RAM simulation with efficient worst-case access overhead."* CCSW 2011.
3. **[**PracticalOS**]** Goodrich, Michael, et al. *"Practical oblivious storage."* CODASPY 2012.
4. **[**ObliviStore**]** Stefanov, Emil, and Elaine Shi. *"Oblivistore: High performance oblivious cloud storage."* S&P 2013.



ORAM Quiz

Select the statements that are true with regards to ORAM

- [x] Client must have a private source of randomness
- [ ] Data does not have to be encrypted, since there is no access pattern
- [x] Each access to the remote storage must have a read and a write