# Searching Encrypted Data in the Cloud: the Quest for Practical Security

Bernardo Ferreira

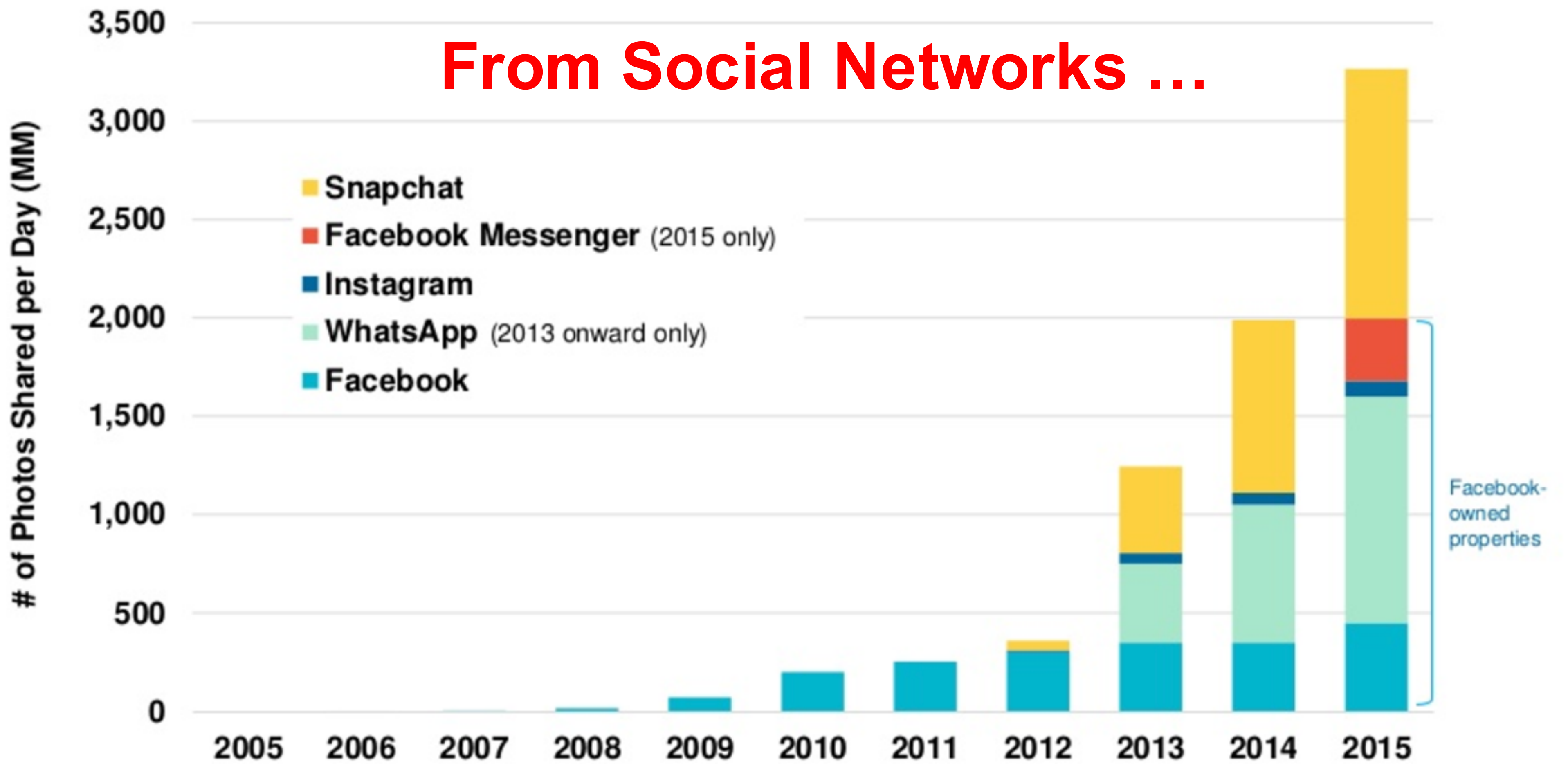
Universidade NOVA de Lisboa
NOVA LINCS

January
2018

73rd IFIP WG10.4 Meeting

# The Cloud is here to stay…

## Daily Number of Photos Shared on Select Platforms, Global, 2005 – 2015

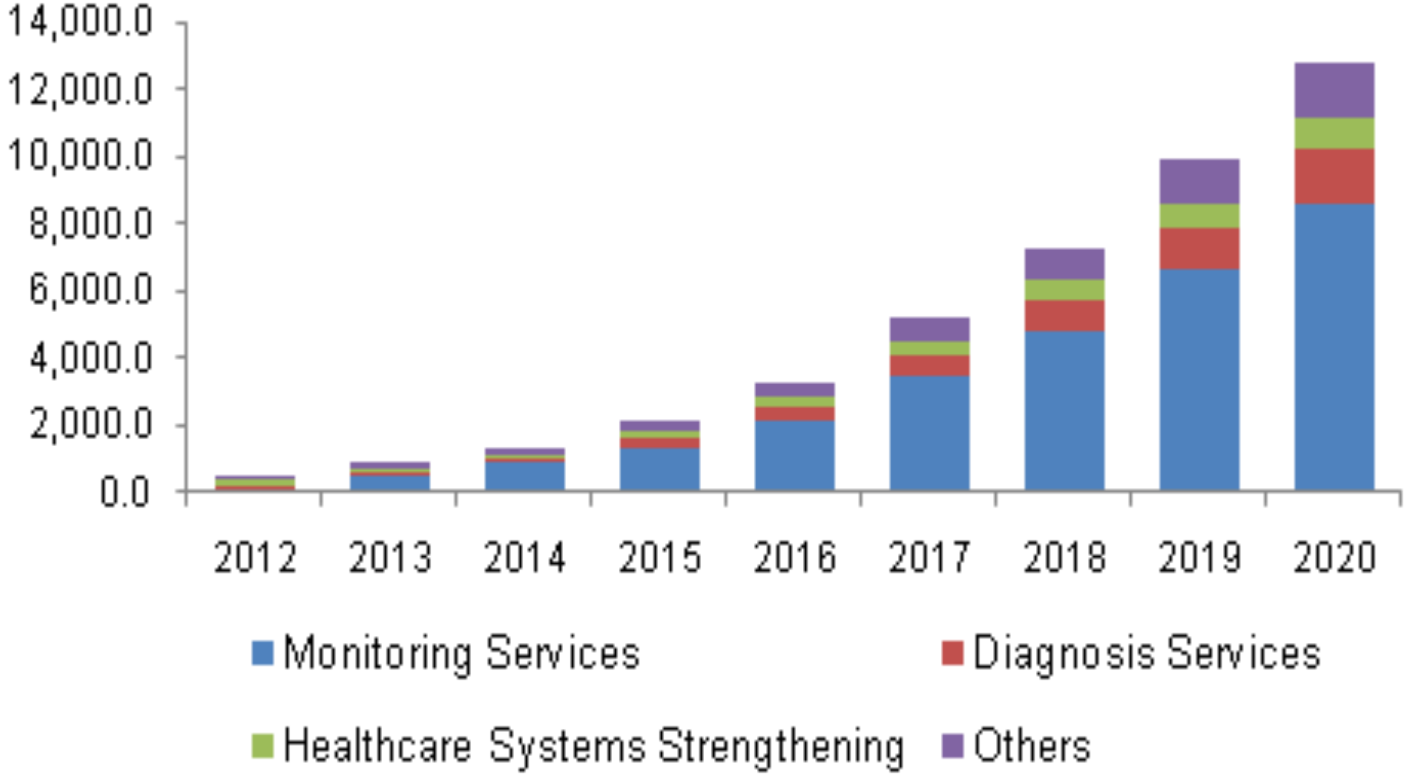


**From Social Networks …**

Source: M. Meeker. Internet Trends 2016. Code Conference, 2016

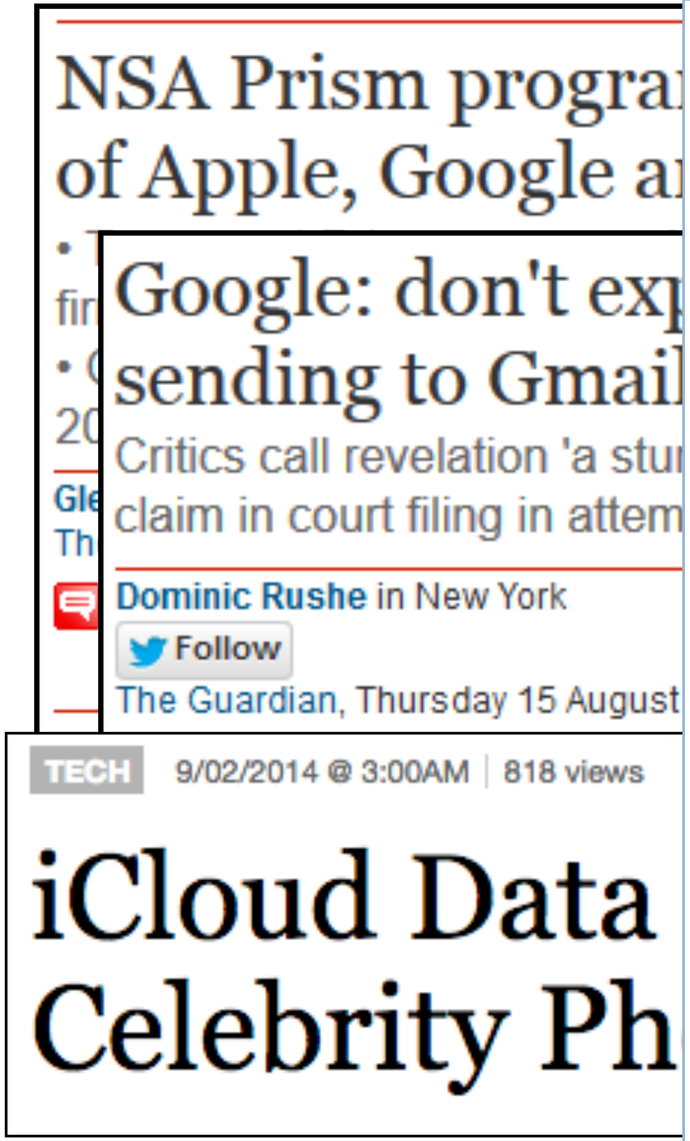
# Cloud adoption is on the rise…

## … to more sensitive applications



Cloud-based Healthcare Management Services

# Cloud Security Issues



NSA Prism progra
of Apple, Google a

• T
fir
• C
20

Gle
Th

Google: don't exp
sending to Gmail

Critics call revelation 'a stu
claim in court filing in attem

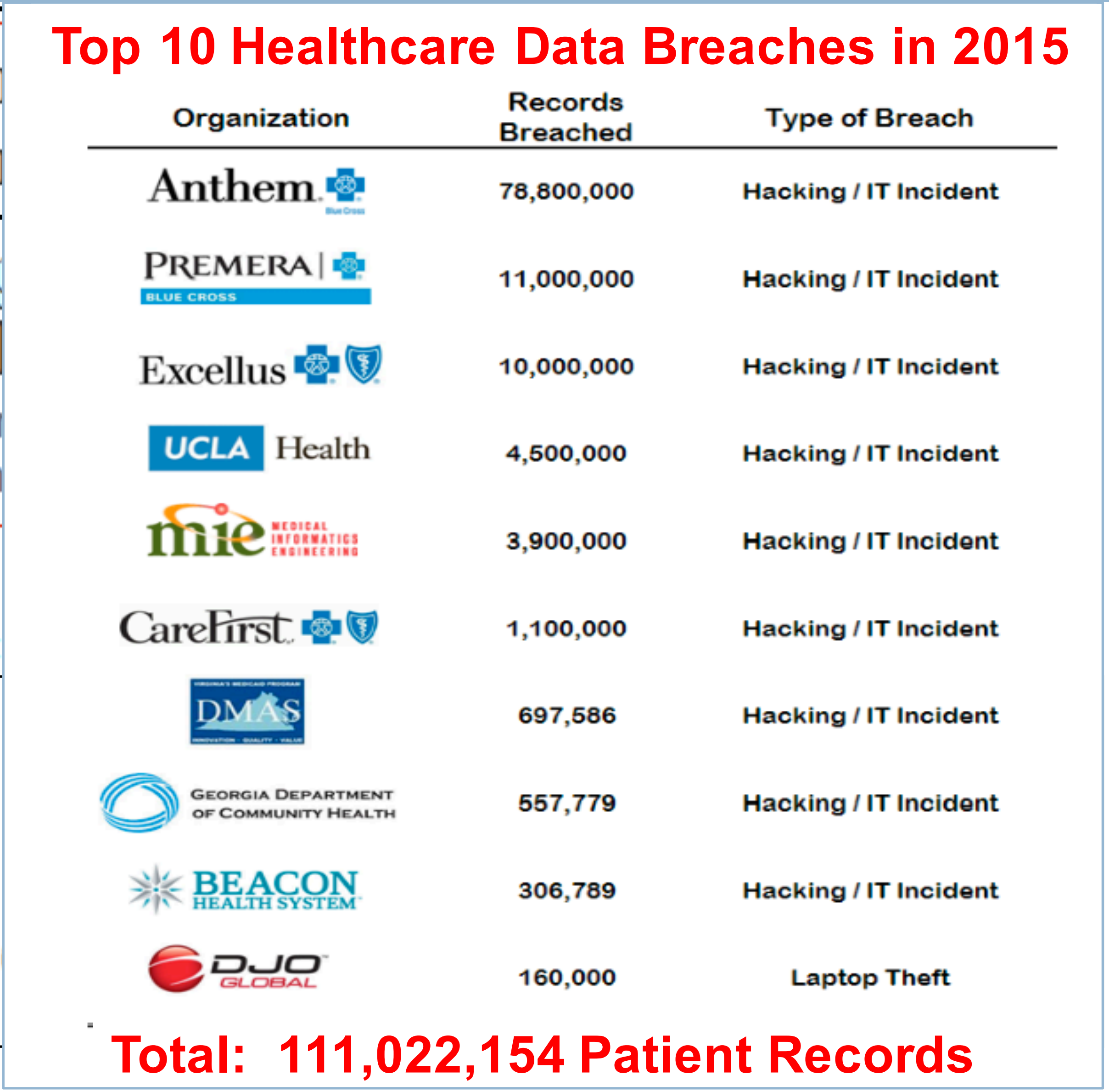
Dominic Rushe in New York

Follow

The Guardian, Thursday 15 August

TECH 9/02/2014 @ 3:00AM | 818 views

iCloud Data
Celebrity Ph



## Top 10 Healthcare Data Breaches in 2015

| Organization | Records Breached | Type of Breach |
|---|---|---|
| Anthem. Blue Cross | 78,800,000 | Hacking / IT Incident |
| PREMERA BLUE CROSS | 11,000,000 | Hacking / IT Incident |
| Excellus | 10,000,000 | Hacking / IT Incident |
| UCLA Health | 4,500,000 | Hacking / IT Incident |
| mie MEDICAL INFORMATICS ENGINEERING | 3,900,000 | Hacking / IT Incident |
| CareFirst. | 1,100,000 | Hacking / IT Incident |
| DMAS | 697,586 | Hacking / IT Incident |
| GEORGIA DEPARTMENT OF COMMUNITY HEALTH | 557,779 | Hacking / IT Incident |
| BEACON HEALTH SYSTEM | 306,789 | Hacking / IT Incident |
| DJO GLOBAL | 160,000 | Laptop Theft |

**Total: 111,022,154 Patient Records**

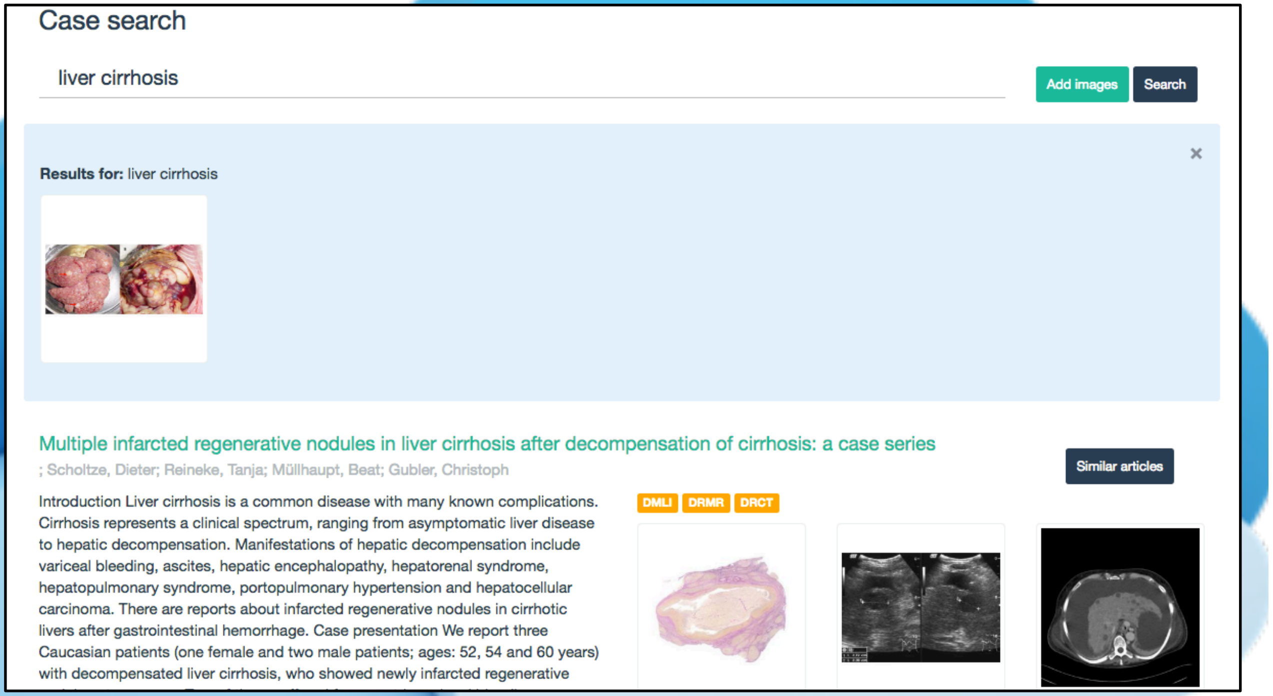
# Challenges

- Protect data privacy in the Cloud
  - In rest, transit and during computations
  - From external and internal attacks

- Support search on encrypted data
  - Search is relevant as cloud database size increases
  - Must be efficient, secure, and provide query expressiveness

# Example Use Case

- □ Cloud-backed Medical Database with sensitive patient records and similarity searching



Case search

liver cirrhosis

**Results for:** liver cirrhosis

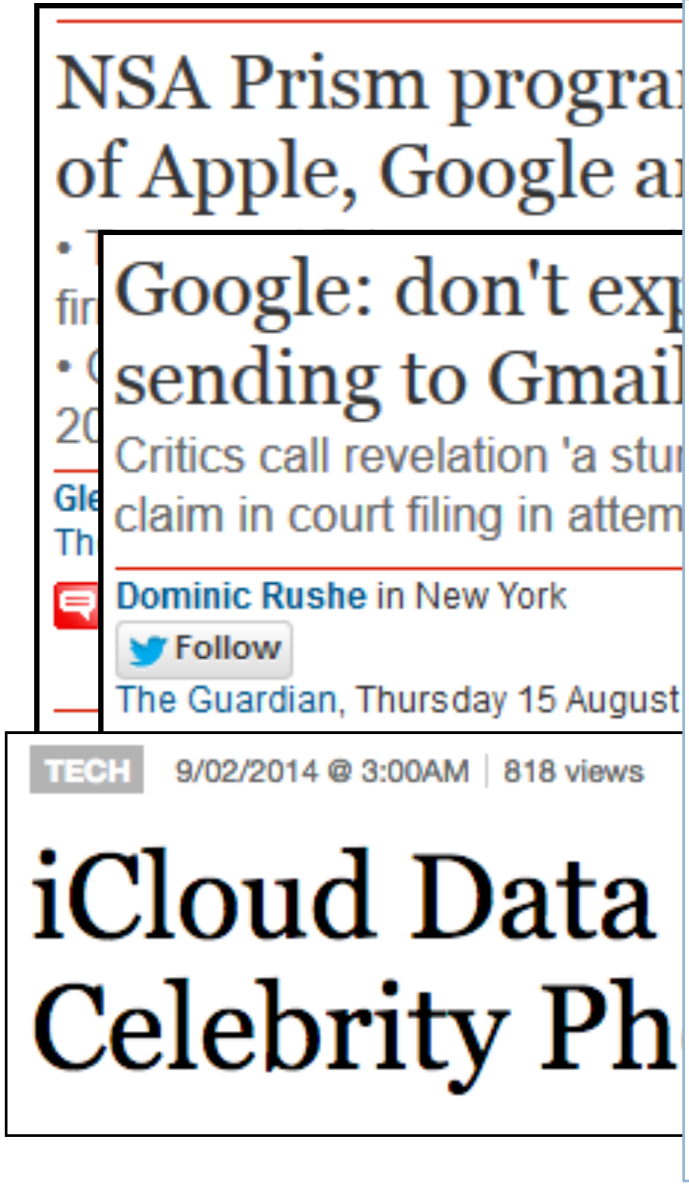
Multiple infarcted regenerative nodules in liver cirrhosis after decompensation of cirrhosis: a case series

; Scholtze, Dieter; Reineke, Tanja; Müllhaupt, Beat; Gubler, Christoph

Introduction Liver cirrhosis is a common disease with many known complications. Cirrhosis represents a clinical spectrum, ranging from asymptomatic liver disease to hepatic decompensation. Manifestations of hepatic decompensation include variceal bleeding, ascites, hepatic encephalopathy, hepatorenal syndrome, hepatopulmonary syndrome, portopulmonary hypertension and hepatocellular carcinoma. There are reports about infarcted regenerative nodules in cirrhotic livers after gastrointestinal hemorrhage. Case presentation We report three Caucasian patients (one female and two male patients; ages: 52, 54 and 60 years) with decompensated liver cirrhosis, who showed newly infarcted regenerative

© NovaMedSearch, NOVA LINCS

# Cloud Security Issues



NSA Prism progra
of Apple, Google a

• T
fir
• (
20

Gle
Th

Google: don't exp
sending to Gmail

Critics call revelation 'a stu
claim in court filing in attem

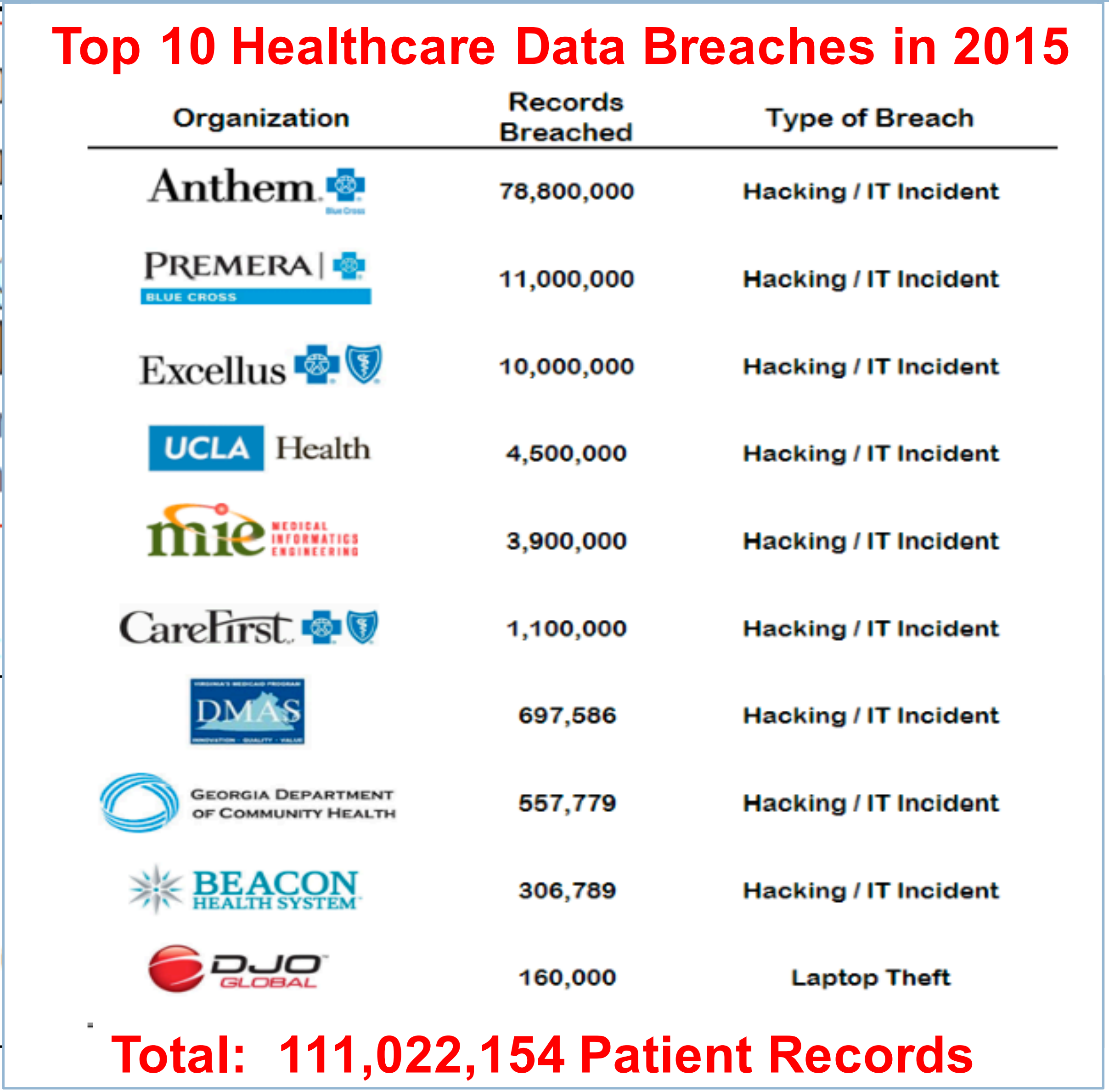
Dominic Rushe in New York

Follow

The Guardian, Thursday 15 August

TECH 9/02/2014 @ 3:00AM | 818 views

iCloud Data
Celebrity Ph



## Top 10 Healthcare Data Breaches in 2015

| Organization | Records Breached | Type of Breach |
|---|---|---|
| Anthem. Blue Cross | 78,800,000 | Hacking / IT Incident |
| PREMERA BLUE CROSS | 11,000,000 | Hacking / IT Incident |
| Excellus | 10,000,000 | Hacking / IT Incident |
| UCLA Health | 4,500,000 | Hacking / IT Incident |
| mie MEDICAL INFORMATICS ENGINEERING | 3,900,000 | Hacking / IT Incident |
| CareFirst | 1,100,000 | Hacking / IT Incident |
| DMAS | 697,586 | Hacking / IT Incident |
| GEORGIA DEPARTMENT OF COMMUNITY HEALTH | 557,779 | Hacking / IT Incident |
| BEACON HEALTH SYSTEM | 306,789 | Hacking / IT Incident |
| DJO GLOBAL | 160,000 | Laptop Theft |

**Total: 111,022,154 Patient Records**

# Privacy Attacks on the Cloud

- Two main adversaries to consider in cloud apps:

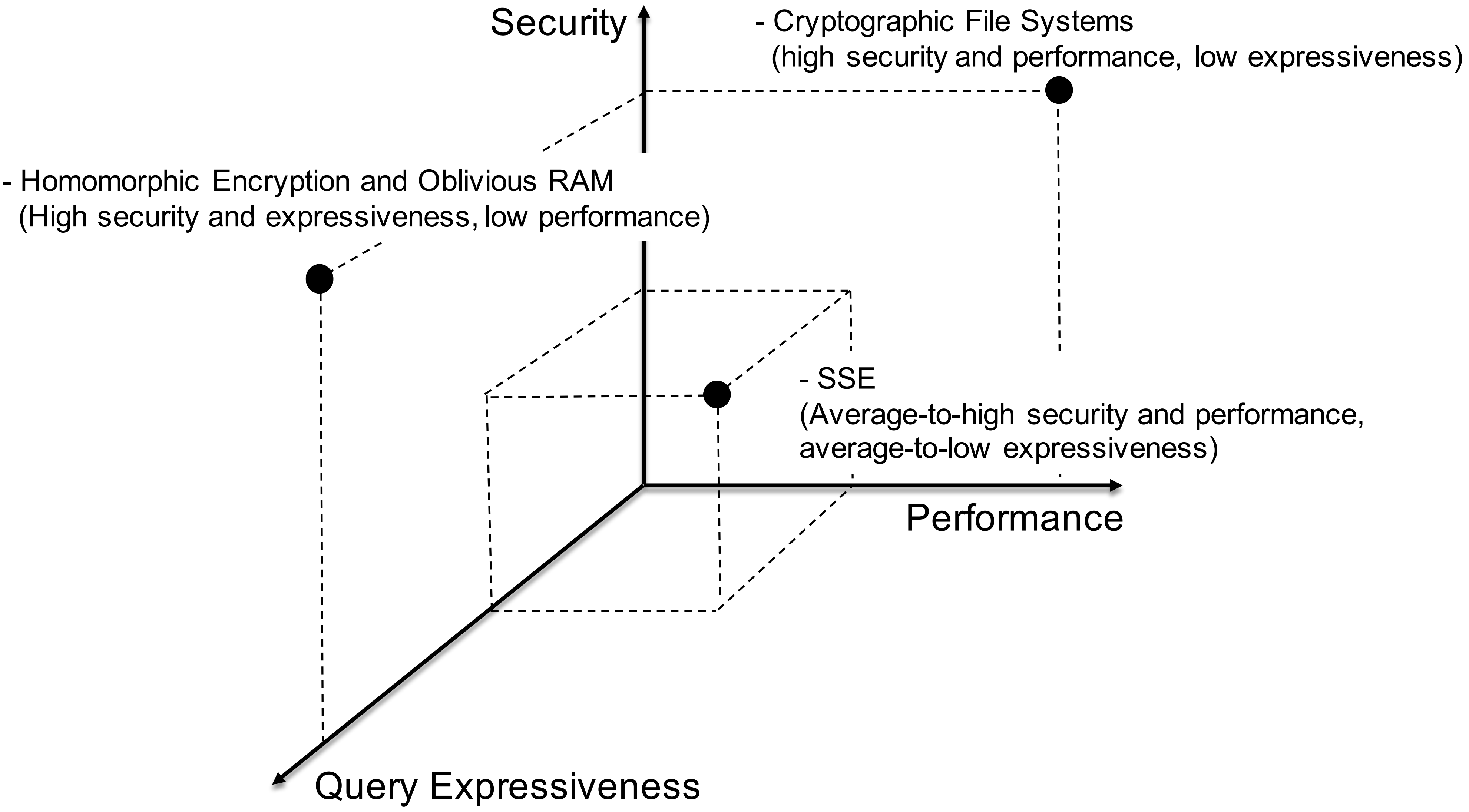
  - Internet Hacker (e.g. iCloud Data Breach)
    - Snapshot attacker – may gain temporary access to cloud servers and perform a snapshot copy of all data
    - Adversarial ability is a subset of Cloud Provider, but should still be considered as a separate adversary

  - Cloud Provider (e.g. PRISM Program)
    - Has access to all data and can observe all traffic and data accesses
    - Assumed to be honest but curious – passive attacks
    - Active attacks may also be interesting to consider

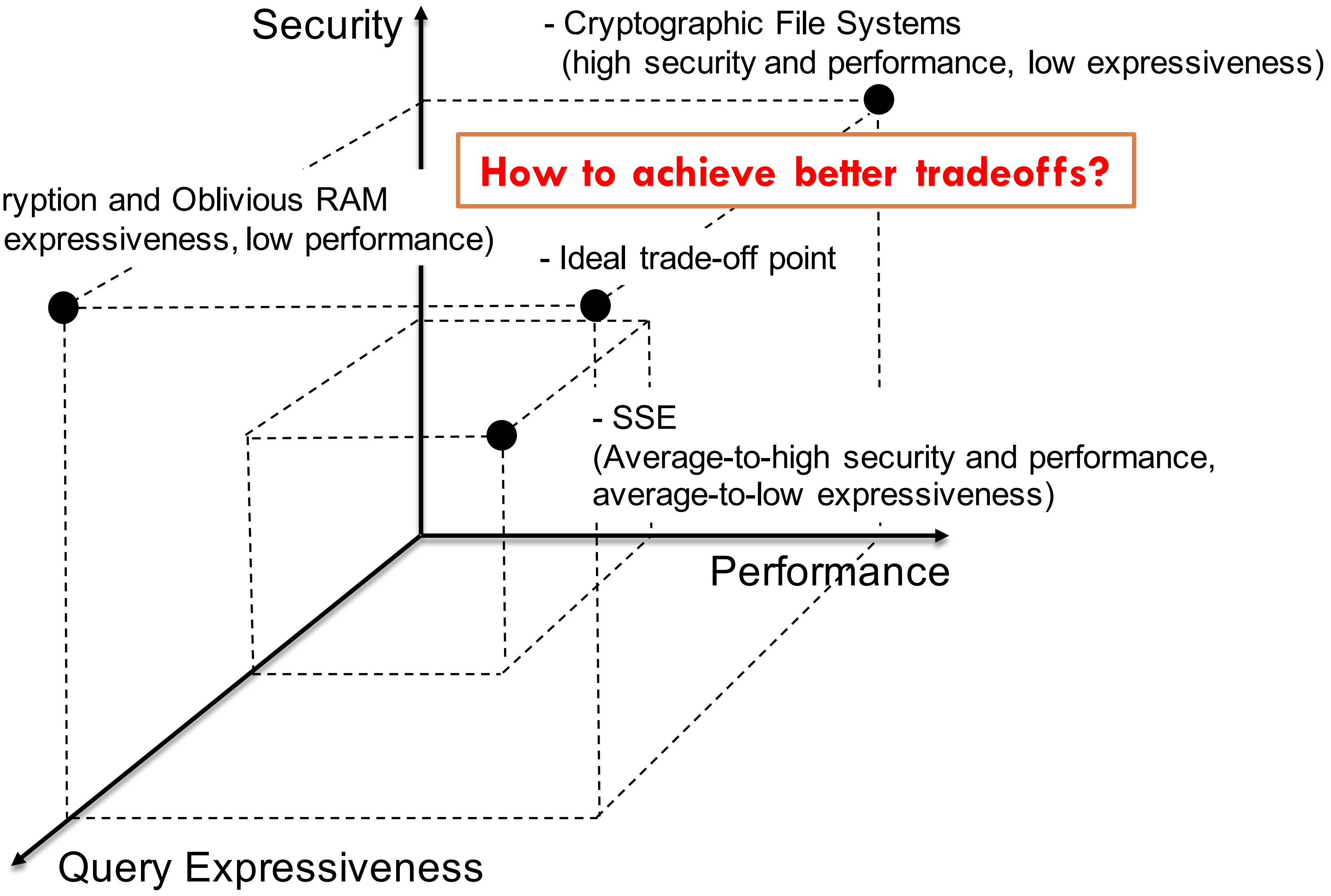
# Existing Solutions

- Cryptographic File Systems
  - ✓ Standard encryption of data at rest and in transit
  - ✗ Computations must be performed on client side
- Oblivious-RAM, Fully Homomorphic Encryption
  - ✓ Arbitrary complex computations on encrypted data
  - ✗ Orders of magnitude away from practical performance
- Searchable Symmetric Encryption (SSE)
  - ✓ Allows efficiently searching encrypted data
  - ✗ High client-side overhead
  - ✗ Limited usability and query expressiveness
  - ✗ Leaks some information patterns w/ operations

# The Security-Performance-Expressiveness Trade-off


Security
- Cryptographic File Systems
(high security and performance, low expressiveness)
- Homomorphic Encryption and Oblivious RAM
(High security and expressiveness, low performance)
- SSE
(Average-to-high security and performance,
average-to-low expressiveness)
Performance
Query Expressiveness

# The Security-Performance-Expressiveness Trade-off



- Cryptographic File Systems (high security and performance, low expressiveness)

**How to achieve better tradeoffs?**

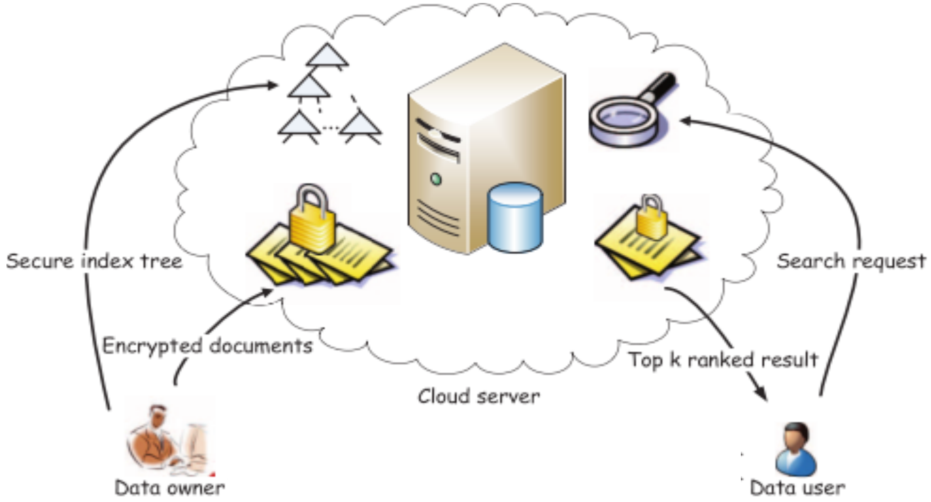
- Homomorphic Encryption and Oblivious RAM (High security and expressiveness, low performance)

- Ideal trade-off point

- SSE (Average-to-high security and performance, average-to-low expressiveness)

# Searchable Symmetric Encryption (SSE)

- Based on an Encrypted Data Structure
  - Reveals no information at rest – semantic security
  - Used in conjunction with a cryptographic token allows performing an encrypted operation
    - E.g. encrypted exact-match search, range queries
  - However, reveals some patterns with queries
    - Repetition of (enc.) queries, repetition of (enc.) query results



# Searchable Symmetric Encryption (SSE)

- Security
  - Snapshot Attacker countered by Enc. Data Structure
    - A snapshot of the database reveals nothing
  - Cloud Provider only partly addressed
    - Patterns leaked + possible background information may reveal contents of queries and database
- Performance
  - Practical and efficient, but most update and search overhead on client
- Query expressiveness - severely limited
  - Designed for exact-match searching of text documents
  - Extending severely limits security and/or performance

# NOVA LINCS Research on SSE

- First Research Vector
  - Improve usability and performance, preserve security guarantees
- Second Research Vector
  - Achieve high security, usability and performance
- Future Research Vectors…

# NOVA LINCS Research on SSE

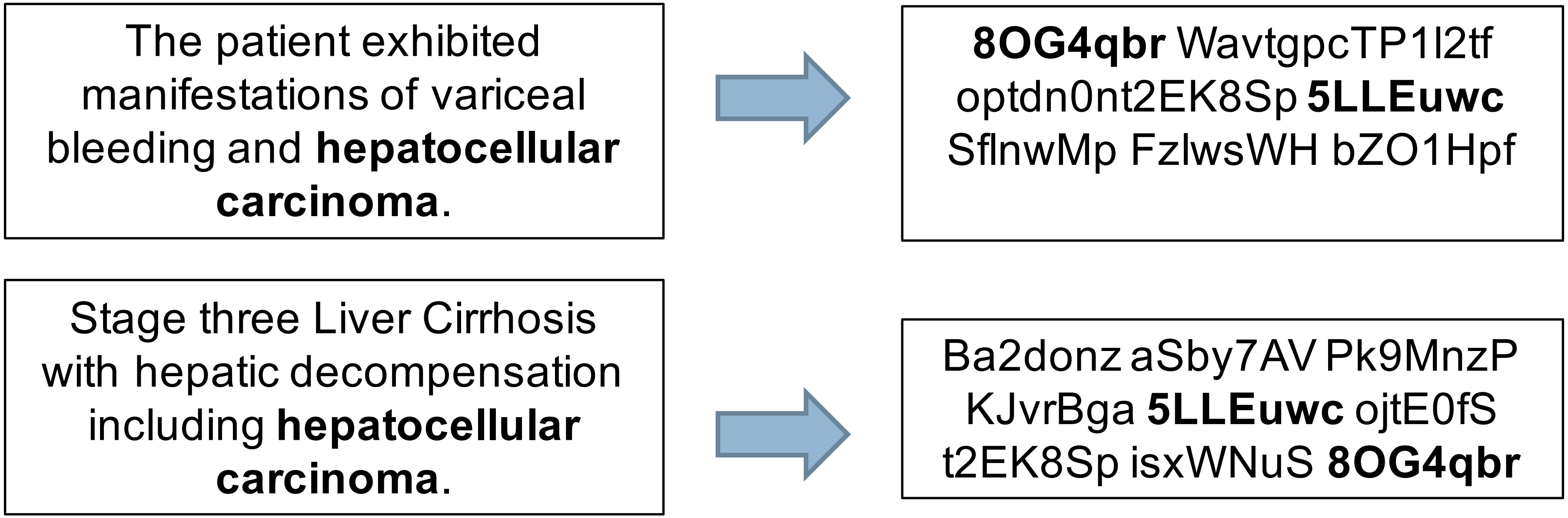
- First Research Vector
  - Improve usability and performance, preserve security guarantees
- Second Research Vector
  - Achieve high security, usability and performance
- Future Research Vectors…

# First Research Vector

- In frequently queried systems, patterns eventually leaked for all search space
  - What if we reveal them from the start? (i.e. w/ updates)
    - Encrypt data with w/ controlled-leakage property-preserving schemes
    - Cloud receives and indexes encrypted data based on patterns leaked
  - Result: efficient and privacy-preserving outsourcing of indexing computations to the Cloud

# First Research Vector

- Text Data (B. Ferreira, H. Domingos - OAIR'13)
  - User det. encrypts keywords, destroys docs. structure
  - Cloud builds index from encrypted keywords
    - Efficient support of multi-keyword ranked queries



The patient exhibited manifestations of variceal bleeding and **hepatocellular carcinoma**.

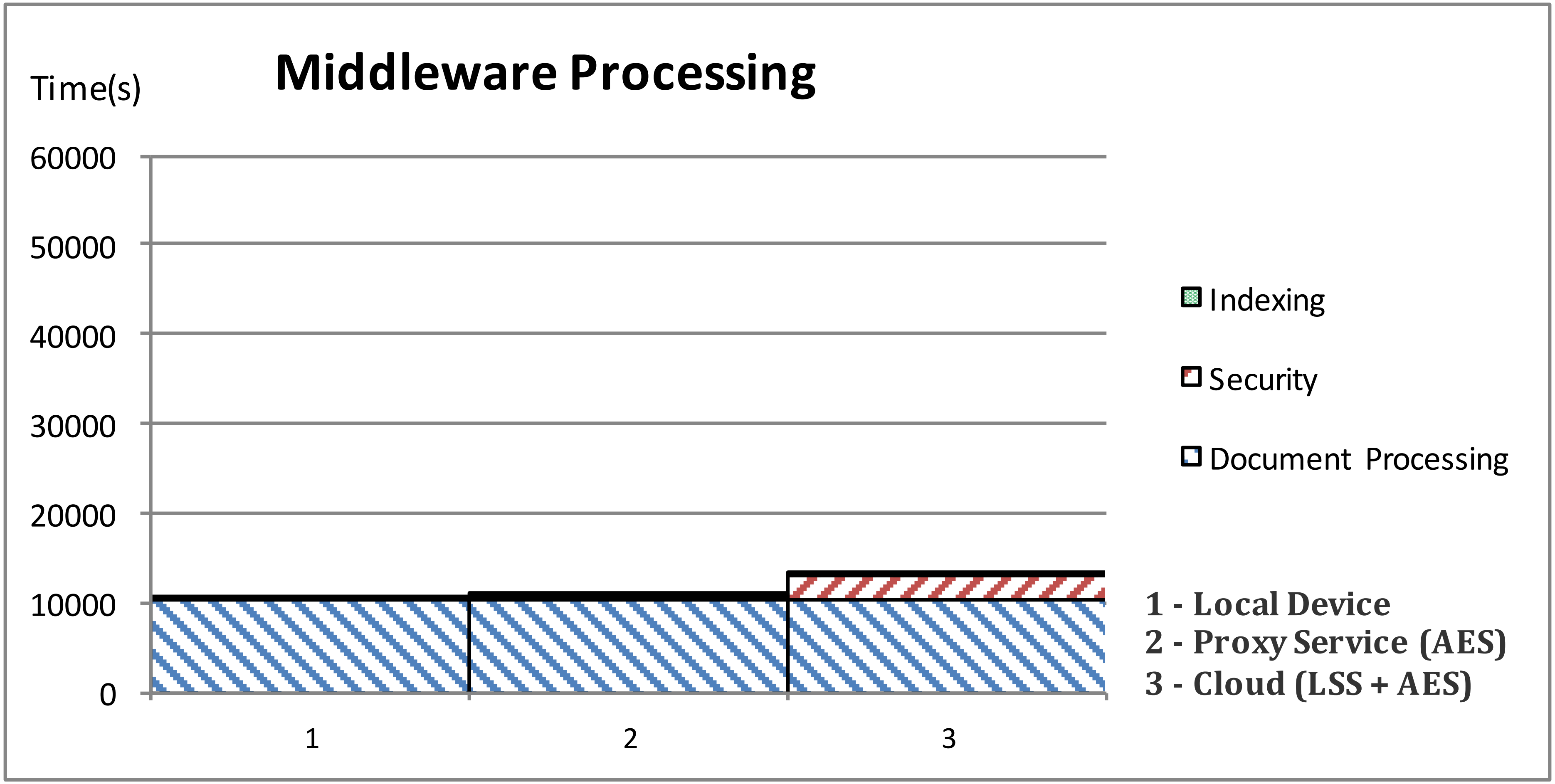
**8OG4qbr** WavtgpcTP1l2tf optdn0nt2EK8Sp **5LLEuwc** SflnwMp FzlwsWH bZO1Hpf

Stage three Liver Cirrhosis with hepatic decompensation including **hepatocellular carcinoma**.

Ba2donz aSby7AV Pk9MnzP KJvrBga **5LLEuwc** ojtE0fS t2EK8Sp isxWNuS **8OG4qbr**

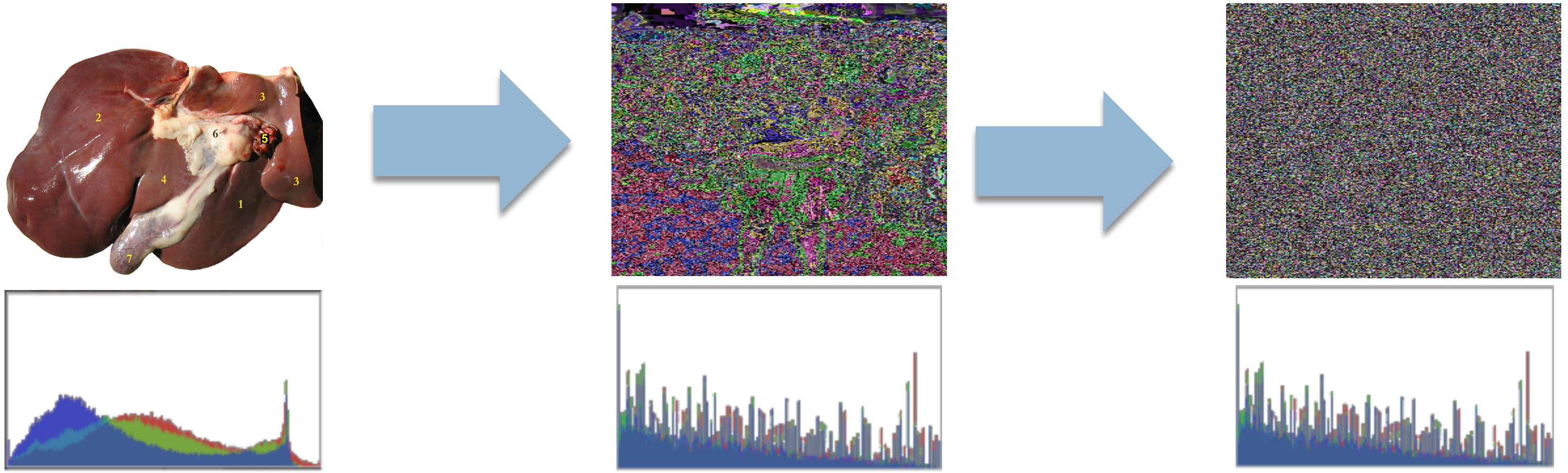
# First Research Vector

- □ Text Data (B. Ferreira, H. Domingos - OAIR’13)



# First Research Vector

- Image Data (B. Ferreira et al. - SRDS'15)
  - Separate color from texture and encrypt in separate
    - Texture encrypted with probabilistic encryption
    - Color encrypted with deterministic encryption
  - Cloud trains and builds index on color data
    - Efficient support of color-based similarity search

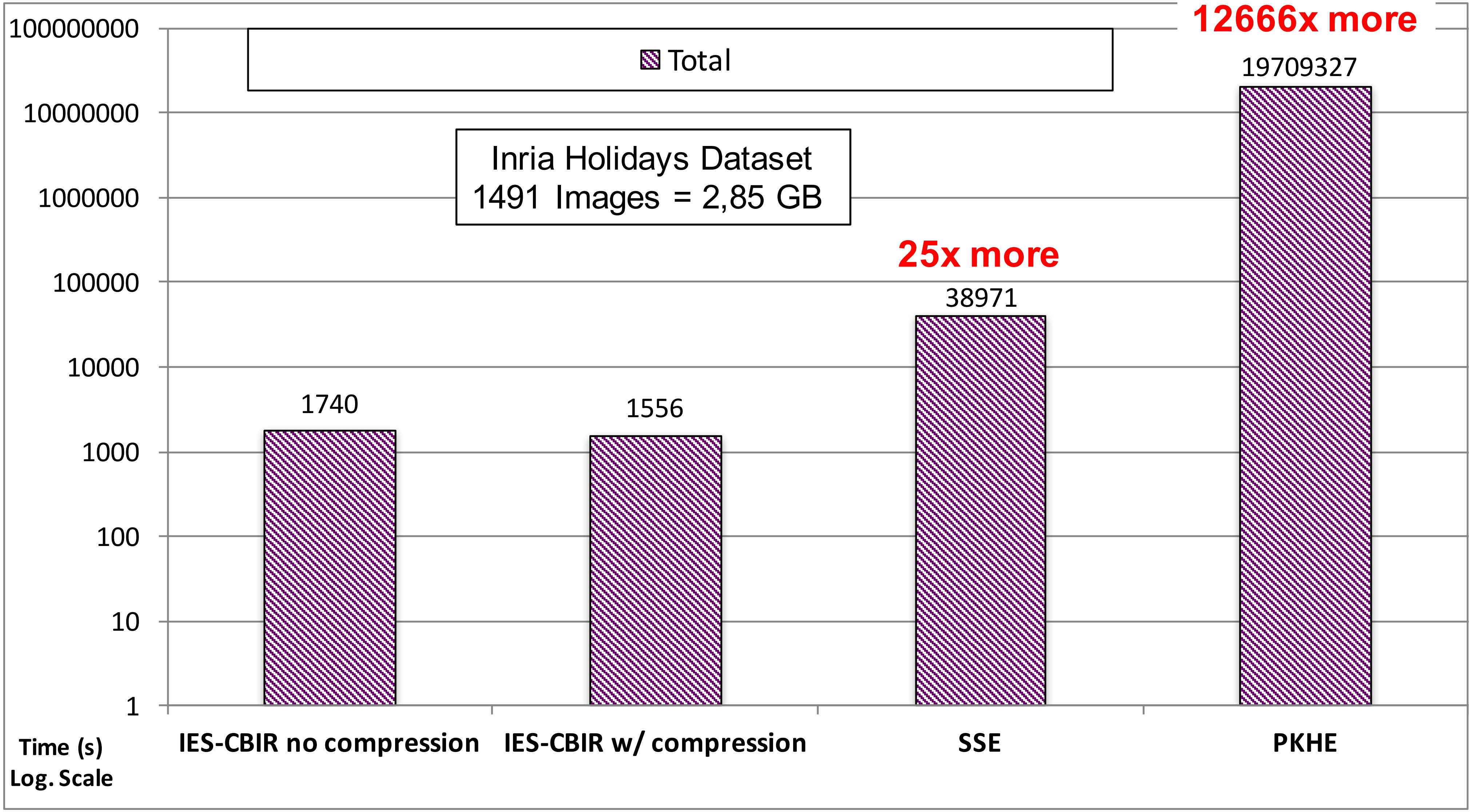


Plaintext Image

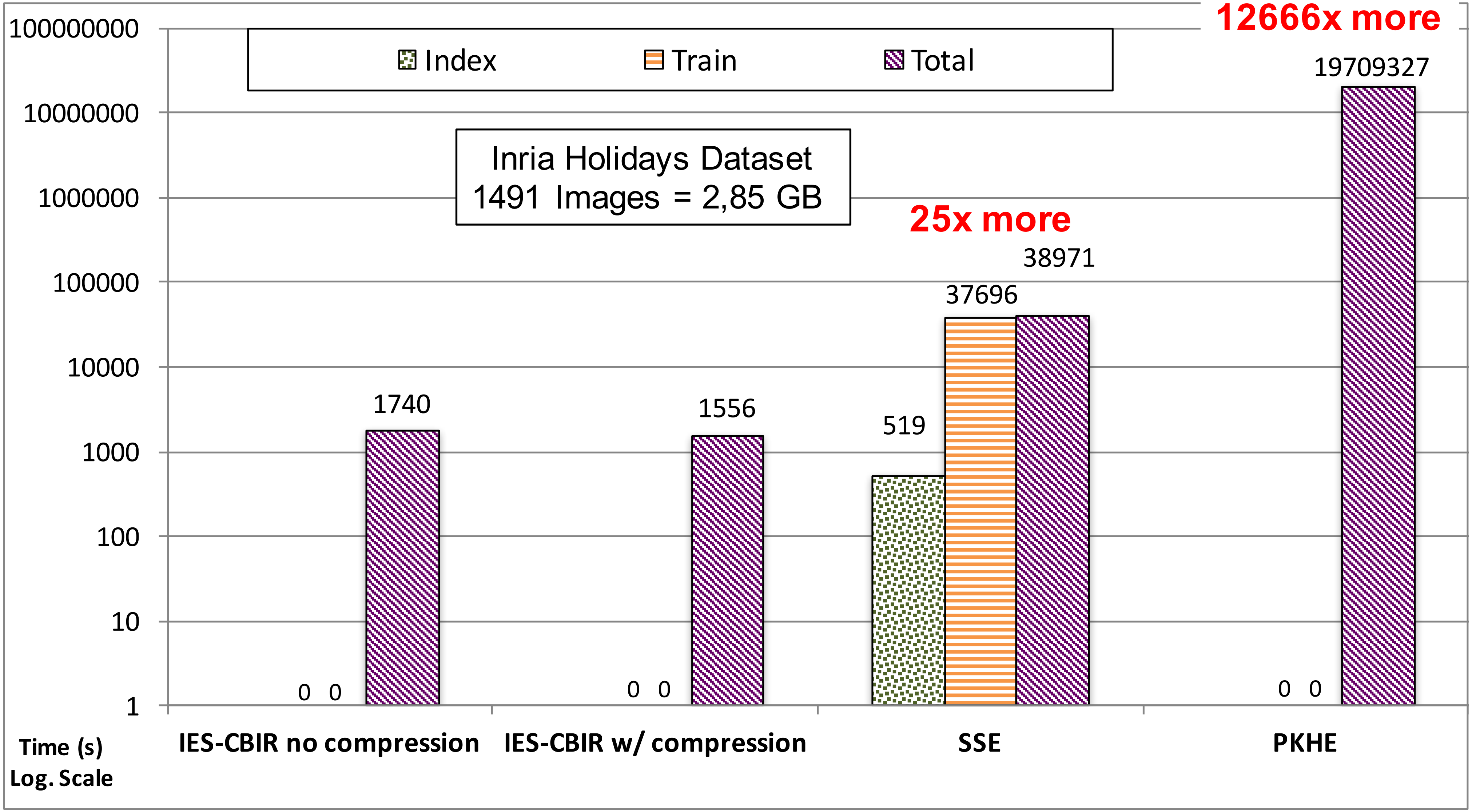
Color Encryption Only

Color & Texture Encryption

# Privacy-Preserving Content-Based Image Retrieval
# Update Performance Results

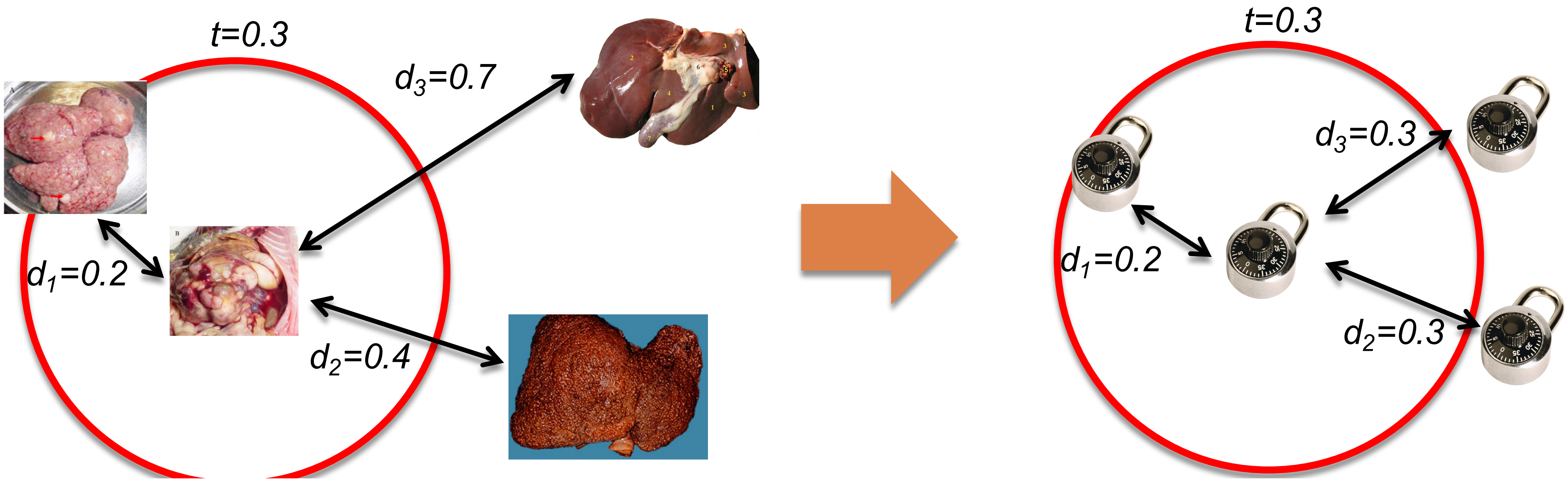


# Privacy-Preserving Content-Based Image Retrieval Update Performance Results

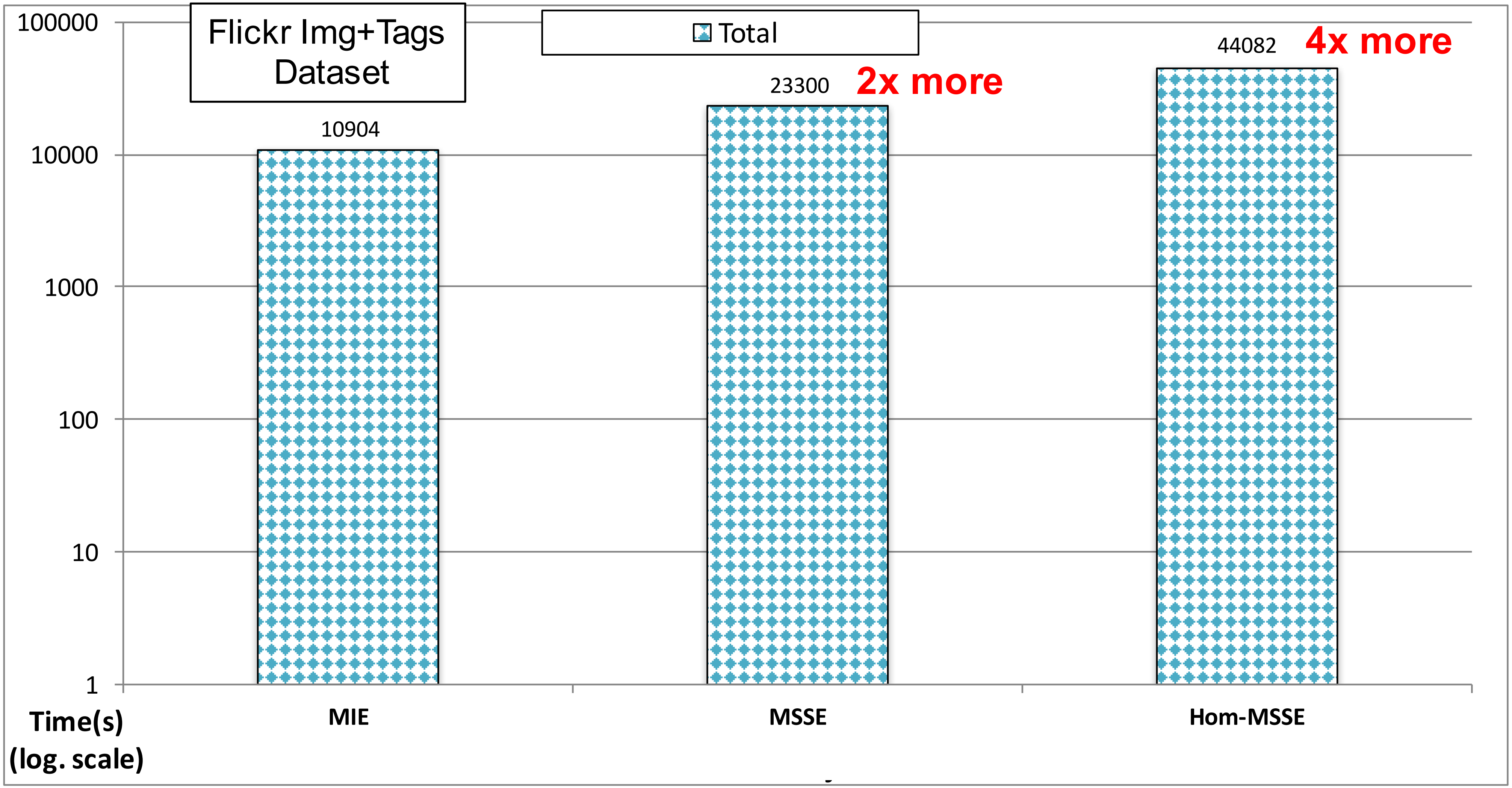


# First Research Vector

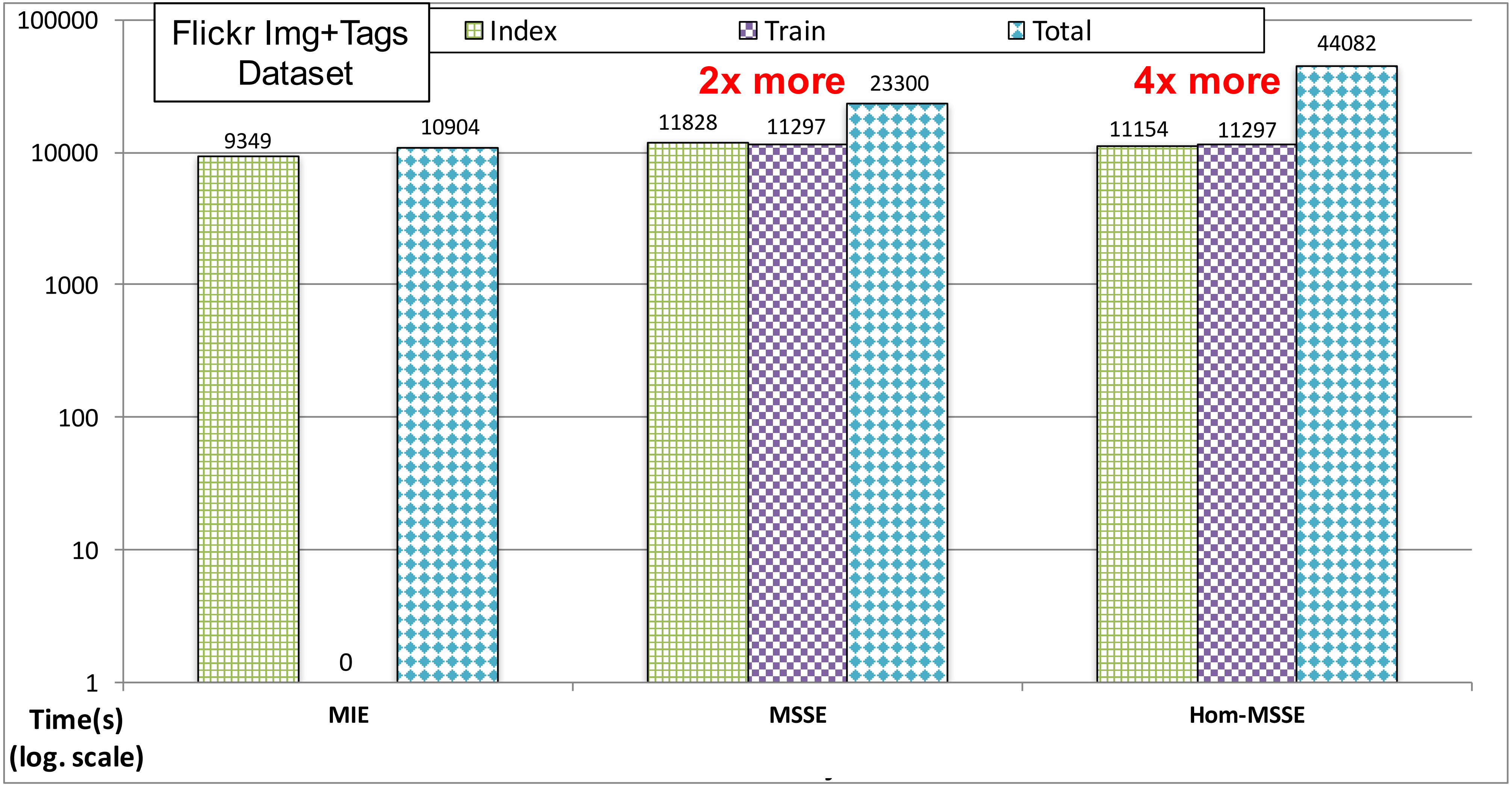
- Multimodal Data (B. Ferreira et al. - DSN’17)
  - DPE – Cryptographic encoding algorithms that preserve controllable distance function between plaintexts
    - Specialized implementations for different medias
  - Cloud leverages DPEs to train & index multimodal data
    - Particularly optimized for mobile devices



# Multimodal Indexable Encryption
## Update Performance Results



# Multimodal Indexable Encryption Update Performance Results

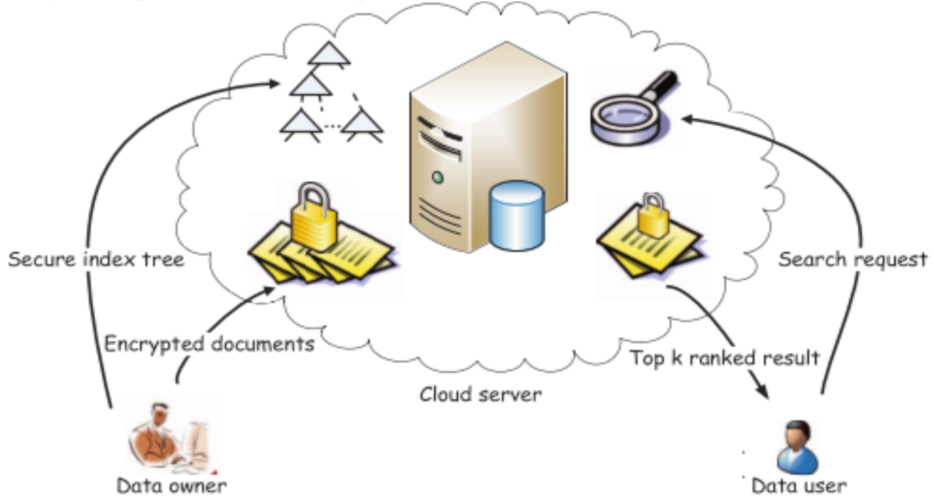


# NOVA LINCS Research on SSE

- First Research Vector
  - Improve usability and performance, preserve security guarantees
- Second Research Vector
  - Achieve high security, usability and performance
- Future Research Vectors…

# Second Research Vector- Insight

- SSE schemes perform critical cryptographic computations in the cloud
  - Performing at the client increases network overhead
  - However this outsourcing leads to severe security issues
    - Recent works explore constrained cryptographic primitives
    - But the fundamental issue remains: outsourcing critical cryptographic computations (even if constrained)



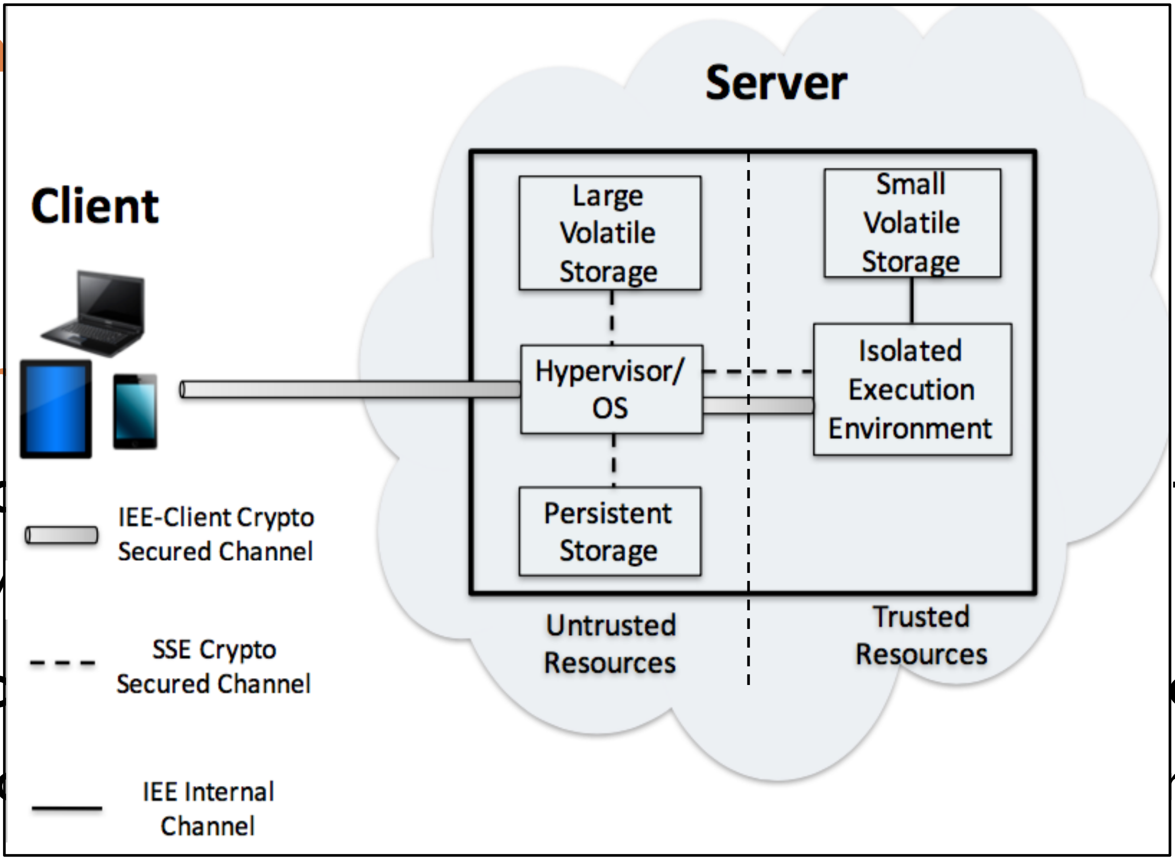
# Second Research Vector- Approach

- Perform critical computations in isolation
  - Through modern attestation-based trusted hardware
- Challenges
  - Minimize assumptions on trusted hardware
  - Avoid trusted hardware vendor-locking
- Solution - Isolated Execution Environments (IEEs)[1]
  - Abstraction for attestation-based trusted hardware
  - Extend formalization to support lightweight IEEs with small trusted resources
    - Expand through standard crypto. over untrusted resources

[1] Barbosa et al. "Foundations of hardware-based attested computation and application to SGX," *EURO S&P'16*.

# Second Research Vector- Approach

- Perform critical computations in isolation
  - Through modern attestation-based trusted hardware
- Challen
  - Minim
  - Avoid
- Solutior EEs)
  - Abstr ted hardw
  - Supp ources
    - Exp resources



# Second Research Vector- Approach

- This approach minimizes information leakage
  - Protects forward and backward privacy
  - Reveals only data accesses
- Optimizes performance
  - Computation, storage, and network overheads
- And opens the way for improved query expressiveness
  - Without sacrificing neither security nor performance

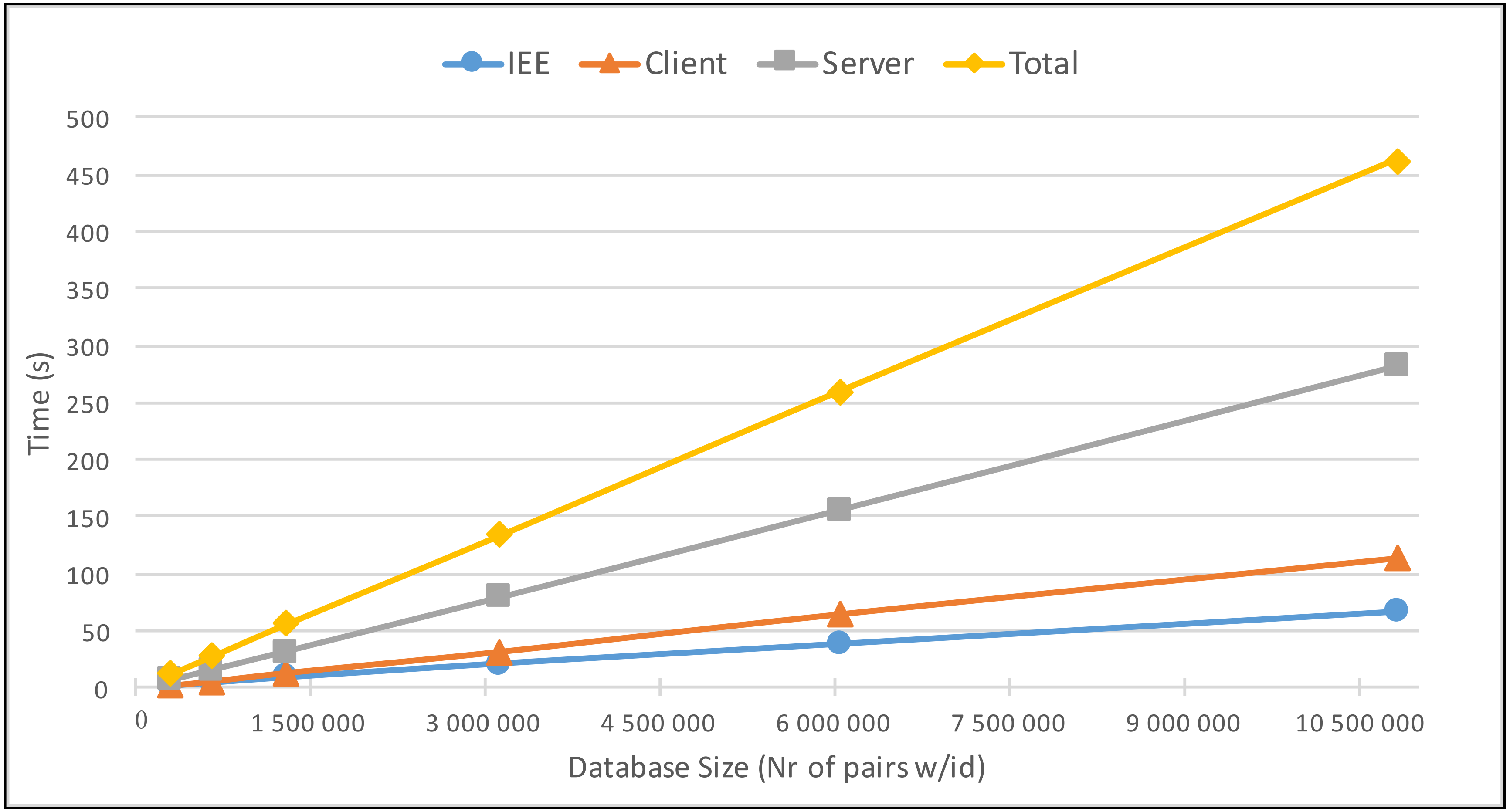
- If IEEs not available in the cloud server…
  - Perform isolated computations in client or trusted proxy

# BISEN – The scheme

- BISEN: Boolean Isolated Searchable Encryption
  - Leverage approach to build a Boolean SSE scheme
    - Boolean SSE literature still very limited in security and performance
    - Efficiently support Boolean queries with arbitrarily complex combinations of conjunctions (ANDs), disjunctions (ORs), and negations
  - Add verifiability for fully malicious adversaries
    - Verify search results and data integrity
  - Open-source implementation based on Intel SGX
    - Available soon

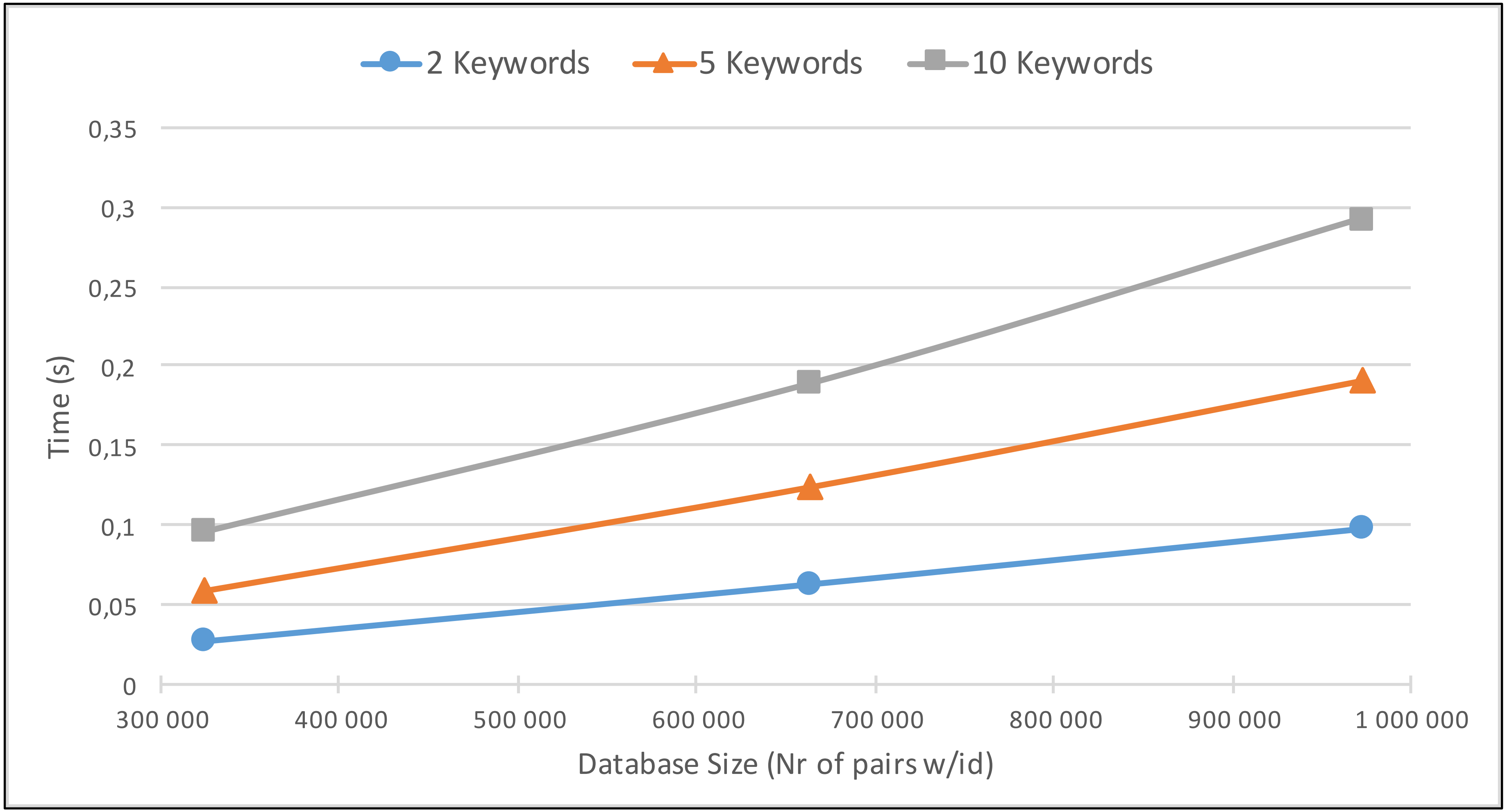
# BISEN – Experimental Results

## □ Update Performance



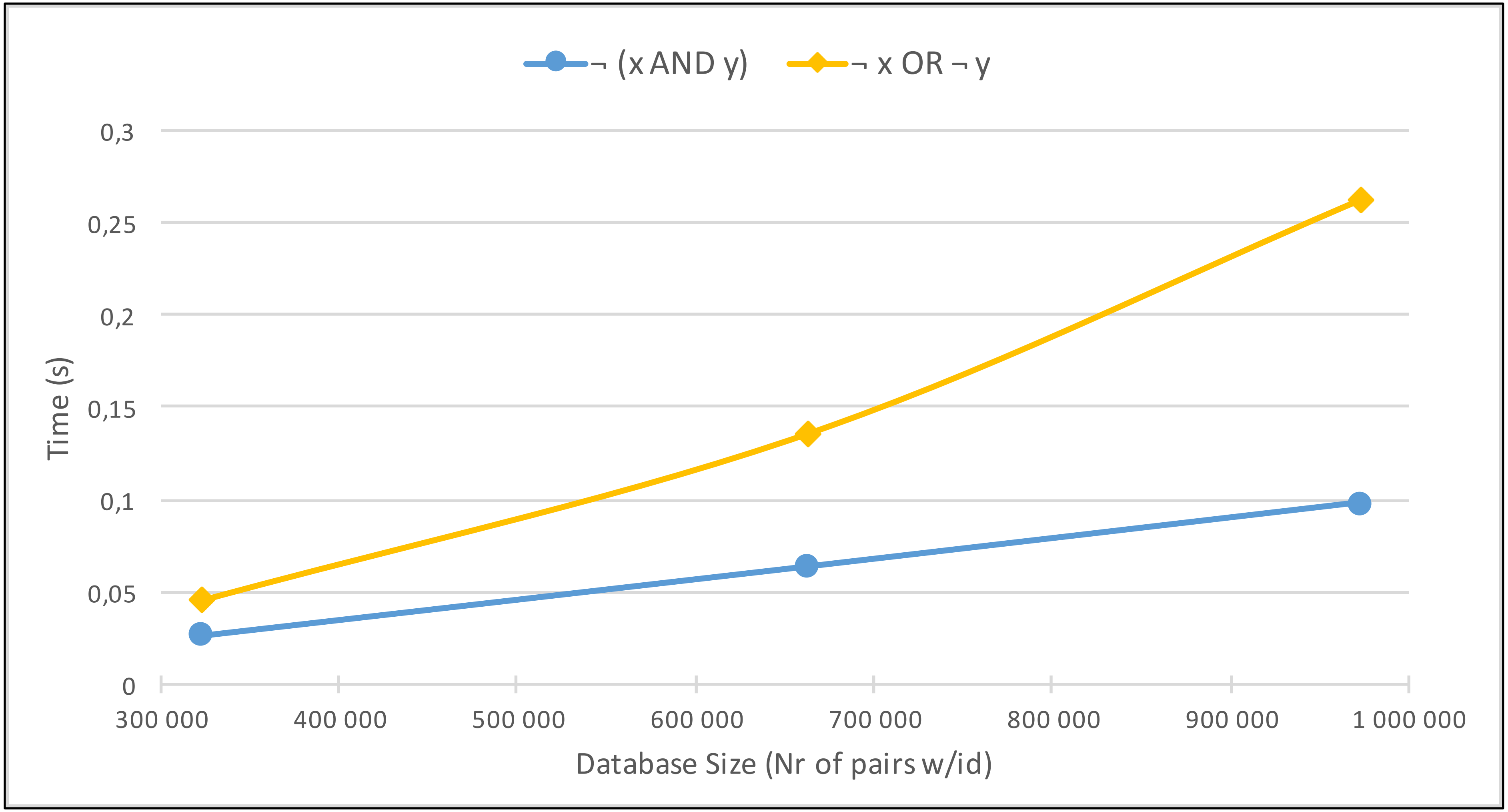
# BISEN – Experimental Results

- Search Performance – Conjunctive Queries (AND)



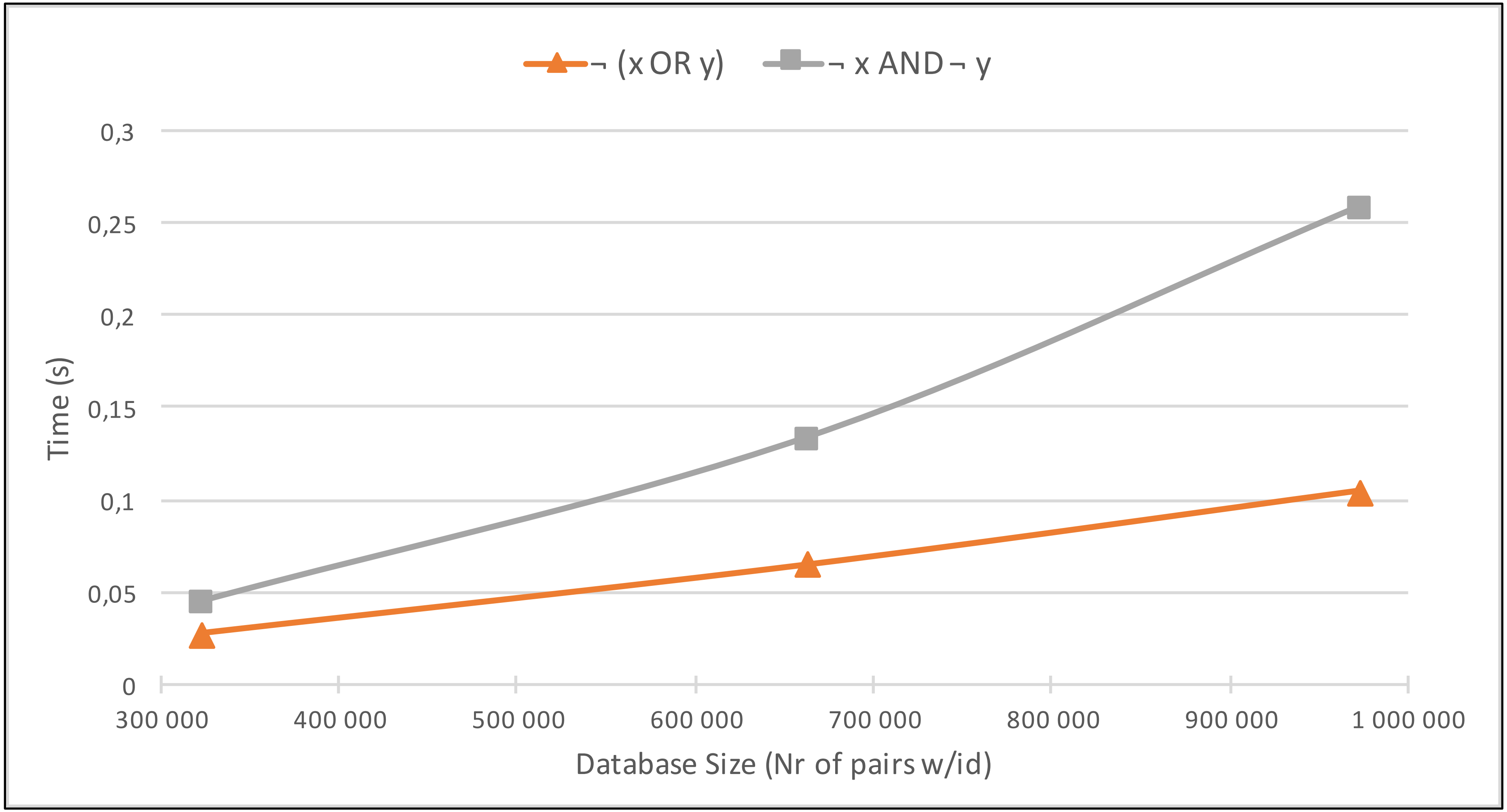
# BISEN – Experimental Results

- Negations



# BISEN – Experimental Results

- Negations



# NOVA LINCS Research on SSE

- First Research Vector
  - Improve usability and performance, preserve security guarantees
    - How? Property-preserving schemes with controlled leakage
- Second Research Vector
  - Achieve high security, usability and performance
    - How? Software-hardware hybrids and Isolated Execution Environments
- Future Research Vectors…

# Future Research Vectors

- Privacy is just part of Cloud Security Issues
  - Explore cloud-of-clouds replication for availability and redundancy
    - How to preserve SSE security guarantees in such scenarios?
- Encrypted Data Structure is an interesting primitive
  - What other use cases can benefit from its properties?
- BISEN laid some foundational work on lightweight IEEs backed by crypto-secured external resources
  - Explore other critical applications that can leverage from this work

# The End

- B. Ferreira, B. Portela, T. Oliveira, G. Borges, H. Domingos, J. Leitão, BISEN: Efficient Boolean Searchable Symmetric Encryption with Minimal Leakage, Technical Report, 2017

- B. Ferreira, J. Leitão, and H. Domingos, Multimodal Indexable Encryption for Mobile Cloud-based Applications, in *DSN'17*, 2017

- B. Ferreira, J. Rodrigues, J. Leitão, and H. Domingos, Practical Privacy-Preserving Content-Based Retrieval in Cloud Image Repositories, *IEEE Transactions on Cloud Computing*, 2017

- B. Ferreira, J. Rodrigues, J. Leitão, and H. Domingos, Privacy-Preserving Content-Based Image Retrieval in the Cloud, in *SRDS'15*, 2015

- B. Ferreira and H. Domingos, Searching private data in a cloud encrypted domain, in OAIR'13, 2013