Pixek

Seny Kamara, Tarik Moataz, Martin Zhu
9,198,580,293*

4%

* since 2013
Why so Few?

“…because it would have hurt Yahoo’s ability to index and search message data…”

— J. Bonforte in NY Times

Incompetence?  Lazyness?  Cost?
Q: can we search on encrypted data?
Encrypted Search (Building Blocks)

- Property-Preserving Encryption (PPE)
- Fully-Homomorphic Encryption (FHE)
- Functional Encryption
- Oblivious RAM (ORAM)
- Structured Encryption (STE)
Evolution from 2001-2018

Property-Preserving Encryption (PPE)
- ’06 DET
- ’09 OPE
- ’11 OPE proofs
- ’12 CryptDB
- ’15 MS Always Enc
- ’15 NKW attacks
- ’16 Snapshot PPE

Oblivious RAM (ORAM)
- ’96 ORAM
- ’12 Tree-based ORAM
- ’13 Path ORAM; ObliviStore
- ’16 Obliv P2P; TaoStore
- ’16 KKNO attacks

Structured Encryption (STE)
- ’01 SSE
- ’06 Efficient SSE
- ’10 STE
- ’12 IKK attacks
- ’12 CS2
- ’13 Boolean SSE
- ’14 OSPIR; BlindSeer
- ’16 Clusion; OpenSSE
- ’17 SQL
Structured Encryption

tk

utk
Would Encryption Even Prevent Breaches?
Q: can encrypted search be deployed?
Why Isn't Encrypted Search Deployed?
Tarik

Martin
End-to-End Encryption
Digital Photos - 1.2 Trillion (2017)

85%

4.7%

10.3%
Photo Collections

Large

Sentimental value

Private

Cloud

Encryption
Celebgate (2014)

- Edward Majerczyk
  - hacked 30 Gmail & iCloud accounts
  - 500 private photos leaked including of many celebrities
AM I NEXT?
Pixek

End-to-end encrypted camera app
Building Blocks

**Clusion**
open source (GPLv3) encrypted search library from Brown ESL
pibase, pidyn, 2Lev, ZMF, IEX-2Lev, IEX-ZMF
coming: DLS, SPX, REX, PBS

**TensorFlow Mobile**
open source machine learning from Google
pre-trained model

**Geomobile**
open source geolocation
Lamp/Bear
23’x21’x24’
Pixek Client

Bear → tk

EC2+S3
What I Didn’t Cover

- Caching
- Crash recovery
- Password recovery
- Multi-device
- Local mode
Pixek v0.1.0 (Current)

• Tags & photos are streamed
  • Encrypted structure needs forward-privacy
• Published state-of-the-art
  • Sophos [Bost16]
  • Diana [Bost-Minaud-Ohrimenko17]
• New scheme
  • pidyn [Cash-Jaeger-Jarecki-Jutla-Krawczyk-Rosu-Steiner14]
    • no public-key operations
    • no constrained PRFs
Background: Data Structures

- DXs map labels to values
- Get: $DX[w_3]$ returns $id_2$

- MMs map labels to tuples
- Get: $MM[w_3]$ returns $(id_2, id_4)$
$$\pi_{\text{dyn}} \quad \text{[CJJJKRS'14]}$$

[Diagram showing the process of EMM.Setup]
\[ \pi_{\text{dyn}} \] [CJJJKRS’14]

\[ \text{EMM.Setup}^k \]

\[ \text{Setup} \]

* PRF and Enc keys are different but derived from \( w_i \)
\[ \pi_{\text{dyn}} \quad \text{[CJJJKRS'14]} \]

Get

\[ \text{EMM.Get} \left[ \begin{array}{c} \text{EMM} \\ \text{Kw1} \end{array} \right] \]

1. DX.Get

2. DX.Get

3. DX.Get

4. DX.Get

\[ \text{DX} \quad F_{\text{Kw1}(1)} \quad \text{id}_1 \]

\[ \text{DX} \quad F_{\text{Kw1}(2)} \quad \text{id}_2 \]

\[ \text{DX} \quad F_{\text{Kw1}(3)} \quad \text{id}_3 \]

\[ \text{DX} \quad F_{\text{Kw1}(4)} \quad \text{id}_4 \]

\[ \perp_{33} \]
\[ \pi_{\text{dyn}} \quad [\text{CJJJKRS'14}] \]

\[ \text{EMM.Get} \quad \text{Kernel} \quad \text{Kw1} \]

\[
\begin{align*}
\text{DX.Get} \quad \text{DX} \quad \text{F}_{\text{Kw1}}(1) \quad \rightarrow \quad \text{id}_1 \\
\text{DX.Get} \quad \text{DX} \quad \text{F}_{\text{Kw1}}(2) \quad \rightarrow \quad \text{id}_3 \\
\text{DX.Get} \quad \text{DX} \quad \text{F}_{\text{Kw1}}(3) \quad \rightarrow \quad \text{id}_4 \\
\text{DX.Get} \quad \text{DX} \quad \text{F}_{\text{Kw1}}(4) \quad \rightarrow \quad \bot
\end{align*}
\]
$\pi_{\text{dyn}}$ [CJJJKRS'14]

$F_{Kw1}(4)$  $id_9$

EMM.Edit$^+$

1. DX.Put

DX
\[ \pi_{\text{dyn}} \] [CJJJKRS'14]

EMM>Edit^+
Forward-Private $\pi_{\text{dyn}}$

- Why is $\pi_{\text{dyn}}$ not forward-private?
  - new pairs encrypted under same key used for search,
    - $K_{wi} := F_K(w_i||1)$
    - so previously searched $w$'s can be linked to new pairs
- Making $\pi_{\text{dyn}}$ forward-private
  - use keys with version number that rotates at each update
    - $K_{wi} := F_K(w_i||\text{version}||1)$
  - To search send keys for all versions
    - $F_K(w_i||\text{version1}||1), \ldots, F_K(w_i||\text{version8}||1)$
Forward-Private $\pi_{\text{dyn}}$

- Search complexity
  - optimal $O(#\text{MM}[w])$
- Token size
  - non-optimal $O(#\text{MM}[w])$
  - new technique makes it $O(1)$ (not implemented yet)
Leakage

• Search pattern
  • *we see if a query is repeated*
  • *ex:* if you search for “bear” 3x, we see you searched for ? 3x

• Access pattern
  • *we see which encrypted photo matched your query*
  • *ex:* if you search for “bear”, we see which encrypted photos match query

• What are the consequences of this leakage?
  • To see your photos we have to break AES
  • To learn about your queries we have to know/guess > 90% of your tags & know the occurrence of each tag
HELP
Testers & Feedback

• Only available on Android
• Let us know @pixekapp if you want access