Ashley Madison Breach Could Expose Privates Of 37 Million Cheaters
Hacker Obtained Children's Headshots and Chatlogs From Toymaker VTech

WRITTEN BY LORENZO FRANCESCHI-BICCHIERAI

November 30, 2015 // 01:57 PM EST
World’s Biggest Data Breaches
Selected losses greater than 30,000 records
(updated 2nd October 2015)
Q: Why is this happening?
Big Data

- Industry and Governments want more data
  - National security
  - Machine learning
  - Business analytics
  - NLP
  - Location-based services
  - ...
Big Data

- More intrusive & sensitive
  - Photos, medical records
  - Location data, email,
  - browsing history, voicemails
- Greater need for security

- Harder to secure
  - NSA Bluffdale holds 2EBs! (2K PBs)
  - Facebook holds 300PBs of photos/videos
  - Vs. nation states, intelligence agencies, organized crime, insiders, …
Big Data

- End-to-end (e2e) encryption!
  - Reduces attack surface
  - Secure small key instead of Big Data

- Impossible to work with
  - Lose search, DBs, IR
  - Find your photo among 300PBs?
  - Rank results?
Q: Can we search on encrypted data?
An Interesting Question
A Lucrative Question

- Major Corporations
  - Microsoft, IBM,
  - Google, Yahoo
  - Hitachi, Fujitsu

- Funding agencies
  - IARPA
  - DARPA
  - NSF

- Startups
  - CipherCloud ($30M+$50M)
  - Navajo (Salesforce)
  - SkyHigh, Vaultive, Inpher
  - Bitglass, Private Machines, ...
“There are a lot of advancements in things like encrypted search...but in general it is a difficult problem”

-- Edward Snowden @ SXSW‘14
Encrypted Search Solutions
Usage

DB — EDB — tk — Devil

14
Desiderata

- Size of EDB
- tk
- EDB
- Search time
- Storage leakage
- Query leakage
- Size of tk
Many Approaches

- Stream ciphers [SWP01]
- Bucketing [HILM02]
- Structured and searchable encryption (StE/SSE) [SWP01, CGKO06, CK10]
- Oblivious RAM (ORAM) [GO96]
- Functional encryption (e.g., PEKS) [BCOP06]
- Multi-party computation (MPC) [Yao82, GMW87]
- Property-preserving encryption (PPE) [AKSX04, BBO06, BCLO09]
- Fully-homomorphic encryption [G09]
Tradeoffs: Efficiency vs. Security

- Efficiency
  - STE/SSE-based
  - PPE-based
  - skFE-based
  - pkFE-based
  - ORAM-based
  - FHE-based

- Leakage
Tradeoffs: Functionality vs. Efficiency
Leakage

► Theoretical Cryptography [Goldwasser-Micali82,...]
  ► A great success story
  ► Helps us reason about confidentiality, integrity, ...
  ► Focused on leakage-free cryptography

► Real-world systems security relies on tradeoffs
  ► No cryptographic foundations for tradeoffs
  ► Can we leak X but not Y?
  ► How do we model leakage?
Leakage


- **Leakage analysis**: what is being leaked?
- **Proof**: prove that solution leaks no more
- **Cryptanalysis**: can we exploit the leakage?
Applications
Encrypted Search Engines

- Desktop search
  - Windows search, Apple Spotlight
- Personal cloud storage
  - Dropbox, OneDrive, iCloud, ...
- Webmail
  - Gmail, Yahoo! Mail, Outlook.com,...
Encrypted DBs

- Standard DBs
  - DB encrypted in memory
- Cloud DBs
  - DB encrypted in cloud
Encrypted NSA Metadata Program [K.14]

1. To & from numbers, time of call, duration for all US-to-US, US-to-Foreign and Foreign-to-US calls
2. NSA DB can only be queried by individual phone number (seed)
3. Analyst queries must be approved by small number of NSA officials
Systems (Provably Secure)
Systems

- CS2 (C++)
  - Microsoft Research, 2012
  - Queries: single keyword search
  - 16MB email collection in 53ms

- BlindSeer (C++) [IARPA]
  - Columbia & Bell Labs, 2014
  - Queries: boolean
  - Synthetic dataset
  - Search time
    - For ($w_1$ and $w_2$): 250ms
    - $w_1$ in 1 docs
    - $w_2$ in 10K docs
Systems

- IBM-UCI (C++) [IARPA]
  - IBM Research & UC Irvine, 2013
  - Queries: conjunctive
  - 1.3GB email collection
  - Search time
    - For \((w_1 \text{ and } w_2)\): 5ms
    - \(w_1\) in 15 docs
    - \(w_2\) in 1M docs

- Clusion (Java)
  - Brown & Colorado St., 2016
  - Queries: Boolean
  - 1.3GB email collection
  - Search time
    - For \((w_1 \text{ or } w_2) \text{ and } (w_3 \text{ or } w_4)\) in 1.5ms
    - \((w_1 \text{ or } w_2)\) in 10 docs
    - \((w_3 \text{ or } w_4)\) in 1M docs
Systems

► GRECS


► Queries: (approximate) shortest distance on graphs

► 1.6M nodes & 11M edges

► Query time: 10ms
Conclusions

- Exciting and active area of research
- Big potential impact in practice
- Lots of new research directions in theory and systems
- Potential for collaboration between many areas of CS
  - Algorithms and data structures
  - Databases
  - Information retrieval
  - Combinatorial optimization
  - Statistics
Thank You

Different translations of "Thank You" in various languages.