CIS 700/003: Distributed Systems meet Social Networks

Deanonymization

April 1, 2010
What does 'zero-day' mean?

- Lots of different definitions

- Even security experts do not appear to use it consistently

Typical comment:

- "You would expect that such a well-used expression have a clear definition, but this is sadly not the case. Even in Secode there is some disagreement in what the expression means."

  -- Security Threats and Trends July 2007, Secode
What do the experts say?

- "A zero-day vulnerability is one that appears to have been exploited in the wild prior to being publicly known."

- "Zero day threat: A hazard so new that no viable protection against it exists."
  -- zerodaythreat.com

- "Your team must be prepared for a zero-day exploit of a vulnerability - one for which a security update does not exist."
  -- Microsoft SDL 4.1a

- "Zero-day exploits: Computer code that exploits a vulnerability for which a patch is not yet available."
  -- McAfee Virtual Criminology Report 2007

- "A zero-day exploit is one that takes advantage of a security vulnerability previously unknown to the general public."
Building a definition

- Does it have to have been exploited?
  - Symantec: Yes (in the wild)
  - Microsoft, McAfee, US-CERT: No (but code must exist?)
  - zerodaythreat.com: No

- Does it have to be unknown to the public?
  - Symantec, US-CERT: Yes
  - Microsoft, McAfee: No

- What has to be true about the protection?
  - Symantec: Irrelevant; need only appear to be exploited
  - Microsoft, McAfee: Must be a security update / patch
  - zerodaythreat.com: Must be viable
My attempt at a definition:

- A k-day vulnerability is one that cannot be exploited if system administrators install all patches within (k-ε) days
- In other words, the 'k' is the response time that is available to the system administrators

So what is a zero-day vulnerability?

- One that is exploited before a patch is available
- Even the quickest sysadmin cannot avoid being vulnerable
And now for something completely different...
Where are we?

- Measurement and analysis of MAD systems
- Building MAD systems
- Faults and Misbehavior
- Internet crime
- Privacy and confidentiality
  - Deanonymization
  - Confidentiality
  - Long-term privacy
  - Differential privacy
- Novel opportunities
- Experience

© 2010 Andreas Haeberlen
What percentage of the U.S. population is uniquely identifiable by these combinations?

<table>
<thead>
<tr>
<th>Source Details</th>
<th>5-digit ZIP code</th>
<th>County</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of birth</td>
<td>0.2%</td>
<td>0%</td>
</tr>
<tr>
<td>Year and month of birth</td>
<td>4.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Year, month, and day of birth</td>
<td><strong>63.3%</strong></td>
<td><strong>14.8%</strong></td>
</tr>
</tbody>
</table>

Table 1 from: P. Golle, "Revisiting the Uniqueness of Simple Demographics in the U.S. Population", WPES 2006

Results based on the 2000 census
Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan     Vitaly Shmatikov

S&P 2008
Discussion points

- What exactly is privacy?
- How can we model the attacker?
  - What kind of auxiliary information does he have? What is he looking for? How certain does he need to be?
- How serious is this attack?
  - For Netflix users? For Netflix? For the research community?
- What could/should Netflix have done instead?
- Can we give privacy guarantees?
  - If so, how? Against what kinds of attackers?
- What kind of privacy assurances would you want?
- Does this mean that we can't release any more data now, ever?
Some of your discussion points

- Are real datasets really sparse?
- Would people really care if their movie ratings were revealed?
- Are other public datasets sanitized better?
- How to verify the results?
- Don't many people have very similar tastes in movies?
- What happens if we don't have column identifiers?
- How hard would it be to repeat this for a different dataset?
- Shouldn't we try this on datasets that are released regularly?
- What do they mean by 'robustness'? Is this really robust?
De-anonymizing Social Networks

Arvind Narayanan    Vitaly Shmatikov

S&P 2009
Motivation

- Social network information is often shared
  - Academic and government data mining
  - Research (sociology, epidemiology, ...)
  - Advertising
  - OSN apps
  - Novel applications (F2F, aggregation sites, ...)

- Data is often anonymized by removing PII

- But is this enough?
  - No! You can use an approach similar to the one in the S&P 2008 paper.
Deanonymizing social networks

- Assumption: Attacker knows two social nets whose membership partially overlaps
  - Additionally, a few 'seed' nodes that are in both networks
- Idea: Attacker iteratively establishes correspondences between nodes
  - Snowball effect: Each step increases auxiliary information
How many seeds do you need?

- Too few: No snowball effect
- Enough: Re-identification rate stabilizes
- In between: Phase transition
De-anonymizing Twitter

- **Experiment**: De-anonymize Twitter using Flickr
  - **Goal**: Given just the bare Twitter graph (no names or usernames), identify as many individual users as possible
  - **Auxiliary information**: Flickr graph (with names)
  - **Seeds**: 150 nodes present in both graphs

- **How do they know the 'ground truth'?**
  - Computed using completely different information: By matching names and usernames from Flickr and Twitter
  - Results in 27,000 mappings
  - Errors in 'ground truth' can only degrade reported performance
Results

- Of the 27,000 'ground truth' mappings...
  - 30.8% were re-identified correctly
  - 12.1% were identified incorrectly
  - 57% were not identified

- Of the incorrectly identified mappings...
  - 41% were very close to the true mapping (distance 1)
  - 55% were mapped to the same geographic location (attacker can still learn something)
  - 27% are completely erroneous
    - Note: The first two categories partially overlap
Take-away points

- Anonymity is required for privacy, but it is not sufficient
  - Structure of the data may allow an attacker with outside information to undo some or all of the anonymization
  - Attacker may be able to learn sensitive information even if he cannot identify the exact individual

- In many cases it takes surprisingly little information to identify an individual
  - 63% of U.S. citizens uniquely identifiable by DoB+zip code
  - 68% of Netflix users: Two movie ratings + approx. dates
  - 84% of Netflix users: 6 out of 8 'rare' movies
Where are we?

- Measurement and analysis of MAD systems
- Building MAD systems
- Faults and Misbehavior
- Internet crime
- Privacy and confidentiality
  - Deanonymization
  - Confidentiality
  - Long-term privacy
  - Differential privacy
- Novel opportunities
- Experience
Stay tuned!

Next week you will learn:
How social networks could protect your privacy better