Security and Privacy Implications of URL Shortening Services

Alexander Neumann, Johannes Barnickel, Ulrike Meyer

IT Security Research Group
RWTH Aachen University
Germany

May 26th, 2011
Introduction

• Alexander Neumann
  ▶ Computer science student, RWTH Aachen University, Germany
  ▶ Working as Penetration Tester for RedTeam Pentesting GmbH

• Johannes Barnickel
  ▶ Research Advisor
  ▶ Ph.D. Student, IT Security Group, RWTH Aachen University
Definitions

- **Short URL** does not exceed 35 characters
- **Shortened URL** belongs to a shortening service
- **Corner cases:**
  - Shortened URL but not short:
    http://urlshorteningservicefortwitter.com/4czyx
  - Short URL but not shortened:
    http://www.google.de/search?q=IEEE
The Problem

- Long URLs tend to wrap (especially in mail)
- Twitter is restricted to 140 characters
- URLs in books should be short
- User tracking and statistics

“Solution”: URL shortening services (USS)
Technical Description

1. User posts long URL to service
   http://itsec.rwth-aachen.de/research
2. Service generates short URL
   http://nvg8.it/8b896c
3. Other user requests short URL
4. Server responds HTTP 301 or 302
5. Other user is redirected to long URL
Downsides

- Link destination is not transparent
- Reachability of the service is not guaranteed
- USS can silently stop working
  ⇒ All shortened URLs defunct
- Request is delayed when connection to USS has high latency
- Service might be hacked and redirect to malware
- Service might be hostile and serve malware to vulnerable browsers
- Secret URLs submitted to USS cannot be deleted
- ...
Facts many users do not know

- URL service accumulates click statistics and creates profiles of users
- Short links might vanish after a few years
- When the USS dies, all shortened URLs are dysfunctional
- When the USS is hacked, links might point to other websites (malware, porn, violence, …)
- Short URLs can be enumerated
  \[\Rightarrow\] All submitted URLs are public(!)
Research Objectives

• Analyze risks for clients, servers, and privacy of users
• Empirical studies:
  1. Determine popular USS on Twitter
  2. Analyze the use of USS in spam
  3. Test for malicious services
  4. Analyze user tracking abilities
  5. Enumerate shortened URLs
  6. Submit honeypot URLs to USS
  7. Test latency and availability of popular USS
Preliminaries: List of USS

Sources:

- Firefox add-on *ShortenURL*
- List of USS on longurl.org (URL expansion service)
- Lists of USS on several blogs
- Hostnames of URLs in Spam E-Mails

Leads to list of 610 USS (distinct host names), includes 527 general purpose USS.
Results
Determine Popular USS on Twitter

- **Base:** Two samples of Twitter messages (24 hours each, max 10%)
- **Results:**
  - 7.5 million / 8.7 million messages
  - 1.2 million / 1.1 million URLs
  - 553,320 / 431,636 shortened URLs
- **Top ten services (cover 96% of all USS on Twitter):**
  1. bit.ly / j.mp
  2. t.co
  3. tinyurl.com
  4. goo.gl
  5. ow.ly
  6. dlvr.it
  7. is.gd
  8. migre.me
  9. dld.bz
  10. lnk.ms
Top ten services

Number of shortened URLs

Rest
Top 9
bit.ly

Proportion of shortened URLs

Sample 1
Sample 2

Alexander Neumann, Johannes Barnickel, Ulrike Meyer

RWTH Aachen University
USS in Spam

- Spam e-mails from SCHNUCKI project
- 7.9 million e-mails collected since 2003
- 12.8 million URLs, 0.3% shortened URLs
- Query shortened URLs and analyze response code
- Calculate spam detection rate for relevant services
Spam detection rate of USS

Alexander Neumann, Johannes Barnickel, Ulrike Meyer

RWTH Aachen University
Malicious USS

Attack scenario: Setup a fast and attractive USS, serve all requests normally, but after a while start sending vulnerable browsers to malware sites.

Analysis:
- Query shortened URLs seen on Twitter for 187 different USS with 83 different User-Agent strings
- Analyze HTTP response Location header
- Result: No malicious behaviour found
- But: One service handles browsers different
User-Tracking USS

- Data from previous experiment
- Analyze HTTP response Set-Cookie header
- Results:
  - 65 USS set cookies
  - 38 USS set persistent cookies
  - 28 USS set persistent cookies with validity period 6 months or more
- Define $Q$ as quotient: $\frac{\text{#all unique values}}{\text{#all values}}$ received for the cookie
Validity period of cookies

<table>
<thead>
<tr>
<th>Validity Period (in hours)</th>
<th>one day</th>
<th>one week</th>
<th>one month</th>
<th>half a year</th>
<th>one year</th>
<th>ten years</th>
</tr>
</thead>
<tbody>
<tr>
<td>bit.ly</td>
<td>dlvr.it</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>url.lotpatrol.com</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b23.ru</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>url.lotpatrol.com</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bit.ly</td>
<td>dlvr.it</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Enumerating shortened URLs

For the top ten USS:

- Analyze structure of shortened URLs in both Twitter samples ⇒ character frequency analysis using heatmaps
- Select range, ca. 230k URLs per USS
- Enumerate all URLs in range
- Inspect results by hand, search for secret URLs

- Observation: Only goo.gl imposed restrictions
Character frequency analysis: bit.ly

character frequency of 348,968 bit.ly URLs

character frequency of 259,646 bit.ly URLs

Alexander Neumann, Johannes Barnickel, Ulrike Meyer  
RWTH Aachen University
Character frequency analysis: tinyurl.com

character frequency of 32,831 tinyurl.com URLs

character frequency of 29,541 tinyurl.com URLs

Alexander Neumann, Johannes Barnickel, Ulrike Meyer

RWTH Aachen University
Character frequency analysis: goo.gl

character frequency of 31,065 goo.gl URLs

character frequency of 24,918 goo.gl URLs

Alexander Neumann, Johannes Barnickel, Ulrike Meyer

RWTH Aachen University
Secret URLs found by enumerating

- Archives of private photos
- Several CVs
- Treasurer’s report for a company
- List of names and numbers of a kindergarten in Lindlar
- ...
Submitting secret URLs to USS

- Question: Are secret URLs submitted to USS leaked?
- Set up honeypot web server
- Generate unique URLs for each service
- Suspicious and harmless URLs
- Examples:
  - http://fd0.me/secret/a0df29ac/bb42ce8b
  - http://www.fd0.me/blog/archive/2011/01/14/index.php?article=69e325eb#a5a6c61c
- Submit to 255 USS
- Watch for requests
Results (after four weeks)

- Honeypot is found by Google, Yahoo and Baidu
- 15 URLs requested by Google
- 13 URLs requested by Yahoo
- 2 URLs requested by Baidu
- 13 URLs manually checked by USS administrators
  (9 transmitted the admin URL in the HTTP referrer)
- Four administrators contacted us, are interested in the research
- ⇒ Never submit private URLs to shortening services!
Latency and availability measurements

- Latency and availability measured with Smokeping
- From two different servers (in Germany)
- ICMP and HTTP latency measured for ten services
- Results:
  - Most services have good average HTTP latency
  - Some services have a very bad worst-case HTTP latency
  - goo.gl USS wins
Average ICMP/HTTP Latency
Maximum ICMP/HTTP Latency

Latency (ms)

- ICMP latency (host 1)
- ICMP latency (host 2)
- HTTP latency (host 1)
- HTTP latency (host 2)

Alexander Neumann, Johannes Barnickel, Ulrike Meyer
RWTH Aachen University
Conclusion

- USS have risks
- These risks are very real
- USS leak URLs to search engines
- Do not submit private URLs to USS
- USS are used in spam e-mails
- Several services set long-running cookies and can track the user
- Shortened URLs are not completely random
- goo.gl USS dominates all others in every discipline
Questions?

Thank you for your attention.